GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

Institut für Physikalische und Theoretische Chemie

Exploring the application of deep learning in super-resolution microscopy

DISSERTATION

zur Erlangung des Doktorgrades der Naturwissenschaften

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von

Johanna Viola Rahm

aus Kirchheim unter Teck

Frankfurt am Main, 2024
(D30)

Vom Fachbereich Biochemie, Chemie und Pharmazie der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:	Prof. Dr. Clemens Glaubitz
Gutachter:	Prof. Dr. Mike Heilemann
	Prof. Dr. Achilleas Frangakis

Datum der Disputation:

1 Zusammenfassung

Die Fluoreszenzmikroskopie hat unser Verständnis der Zellbiologie erheblich beeinflusst. Die Erweiterung der beugungsunbegrenzten, superauflösenden Mikroskopie eröffnete ein Beobachtungsfenster, das die Untersuchung der zellulären Organisation auf molekularer Ebene ermöglicht. Die nicht-invasive Natur des sichtbaren Lichts macht die Methode geeignet für Beobachtungen in lebenden Zellen und Organismen (Lakowicz, 2006).
Aufbauend auf diesen Fortschritten wird eine vielversprechende Synergie zwischen superauflösender Fluoreszenzmikroskopie und *deep learning* deutlich, die die Möglichkeiten der Bildgebungsmethoden erweitert. Aufgaben wie Transformation von Bildmodalitäten, Verbesserung des Signal-zu-Rausch-Verhältnisses, Einzelmoleküllokalisation, virtuelle Markierung, spektrale Separation und die Molekülquantifizierung werden mit hoher Präzision ermöglicht (Belthangady et al., 2019).
Die in dieser Arbeit untersuchten Techniken befassen sich mit drei kritischen Aspekten der fortgeschrittenen Mikroskopie: der Verringerung der Bildaufnahmezeit, der Einsparung des Photonenbudgets während der Messung und der Erhöhung der *Multiplexing*-Fähigkeit.

In einem ersten Projekt wurde die Aufnahmezeit von superauflösenden *DNA points accumulation for imaging in nanoscale topography* (DNA-PAINT) (Jungmann et al., 2010) Messungen mithilfe eines neuronalen Netzwerks stark reduziert, was eine schnelle Bildgebung mit hohem Durchsatz ermöglichte. DNA-PAINT gehört zu den Techniken der Einzelmolekül-Lokalisierungsmikroskopie, die auf der Lokalisierung isolierter Emitterpositionen über die Zeit basieren. Diese lokalisierten Emitter werden zusammengefasst, um ein superaufgelöstes Bild zu erzeugen. Da eine große Anzahl von Bildern erforderlich ist, um die Struktur vollständig zu erfassen, ist der Durchsatz gering. Eine Erhöhung der Emitterdichte pro Bild würde die Bildgebung beschleunigen, da die zugrunde liegende Struktur mit weniger Bildern erfasst werden kann. Dies erfordert jedoch robuste Lokalisierungsmethoden, um das Signal von überlappenden Punktspreizungsfunktionen (PSFs) in beugungsbegrenzten Bildern zu entkoppeln.
In dieser Arbeit wird *DeepSTORM* (Nehme et al., 2018) zur Rekonstruktion superaufgelöster Bilder aus DNA-PAINT Messungen verwendet (Narayanasamy et al., 2022). Bei *DeepSTORM* handelt es sich ein neuronalen Netzwerk, das sich für die Generierung superaufgelöster Bilder aus *stochastic optical reconstruction microscopy* (STORM) (Rust et al., 2006) Messungen unter der schwierigen Bedingung einer hohen Emitterdichte bewährt hat. Im Vergleich zu STORM, bei dem die Emitterdichte während der Messung exponentiell abnimmt, haben DNA-PAINT Daten den Vorteil, dass die Emitterdichte aufgrund des kontinuierlichen Austauschs von Fluorophoren konstant bleibt. Die resultierenden Emitterdichten von DNA-PAINT passen in das optimale Leistungsfenster eines trainierten neuronalen Netzes. In einem ersten Schritt wurde die Fähigkeit der Software *DeepSTORM*

erprobt, superaufgelöste Bilder aus DNA-PAINT-Messungen mit konventioneller Dichte zu rekonstruieren. Es wurde ein Arbeitsablauf zur Erzeugung von DNA-PAINT-Trainings- und Testdaten mit hoher Dichte aus experimentellen Messungen etabliert, was eine Alternative zum simulationsbasierten Training darstellt. Für den Trainingsdatensatz wurden Filme mit geringer Emitterdichte aufgenommen, die auf herkömmliche Weise lokalisiert wurden, und dann künstlich Daten mit hoher Dichte erzeugt, indem Bildausschnitte mit beugungsbegrenzten Emittern zusammen mit ihren Lokalisierungslisten summiert wurden. Für den Testdatensatz wurden Referenzdaten durch Aufnahmen mit geringer Emitterdichte gewonnen und ein superauflösendes Bild durch Einzelmolekül-Lokalisierungen erzeugt. Im gleichen Bildausschnitt wurde anschließend mit hoher Emitterdichte gemessen und die Bildern mit *DeepSTORM* verarbeitet. Sowohl Messungen mit niedriger als auch mit hoher Dichte ergaben hochaufgelöste Bilder derselben Struktur. Emitterlokalisierungen wurden durch einen Nachbearbeitungsschritt in *DeepSTORM* extrahiert, für den mehrere Parameter optimiert wurden, um eine ähnliche Anzahl von Lokalisierungen zu erhalten wie bei den Referenzdaten. Die Ergebnisse wurden mit verschiedenen Metriken zur Qualitätskontrolle überprüft, darunter MS-SSIM (Wang et al., 2003), SQUIRREL (Culley et al., 2018) und HAWKMAN (Marsh et al., 2021). Der Einsatz von *DeepSTORM* bei DNA-PAINT Messungen mit hoher Dichte ermöglichte die Bildgebung von α-Tubulin und TOM20 im Gehirngewebe von Mäusen innerhalb einer Minute pro Zielprotein anstelle von 25 Minuten. Obwohl bei den vorhergesagten Bildern eine leichte Verringerung der Auflösung zu verzeichnen war (44 nm Vorhersage gegenüber 32 nm Referenz), blieb die Auflösung im akzeptablen Bereich. Das trainierte Modell ist strukturunabhängig, was bedeutet, dass es auf verschiedene Ziele anwendbar ist, da es sich auf lokale Bildausschnitte der Einzelmolekülmessung stützt, die beugungsbegrenzte Emitter ohne offensichtliche Struktur enthalten. Darüber hinaus ermöglicht DNA-PAINT die Verwendung desselben Fluorophors für mehrere Zielproteine, was bedeutet, dass nur ein einziges trainiertes neuronales Netzwerk als digitale Erweiterung für ein Mikroskop erforderlich ist. Einzelmolekül-Lokalisationsmikroskopie in lebenden Zellen profitiert ebenfalls von der hohen Emitterdichte, da somit weniger Bilder nötig sind, um ein hochaufgelöstes Bild zu erzeugen, was die zeitliche Auflösung verbessert. Durch die Möglichkeit des Trainings neuronaler Netze anhand experimenteller Daten mit geringem Erfassungsaufwand, einer benutzerfreundlichen Pipeline zur Erzeugung von Trainingsdaten mit hoher Dichte und einem etablierten Arbeitsablauf für die Qualitätskontrolle wird DNA-PAINT mit *deep learning* ein leistungsstarkes Werkzeug für die biologische Forschung sein.

In einem zweiten Projekt wurde ein neuronales Netz zur Entrauschung von *stimulated emission depletion* (STED) (Hell et al., 1994) Bildern verwendet, um die kontinuierliche Aufnahme einzelner lebender Zellen über Stunden mit einer zeitlichen Auflösung von Sekunden zu ermöglichen. Die STED-Mikroskopie ist eine weitere leistungsstarke superauflösende Technik, die mit hohen Laserleistungen arbeitet, die Photobleaching und Phototoxizität verursachen, was die Bildgebung von lebenden Zellen und deren Dynamik einschränkt.

Mit Ansätzen wie stabilen Fluorophoren oder transienten Markierungsstrategien können sequenzielle STED-Bilder bis zu einigen Minuten lang aufgenommen werden (Glogger et al., 2022; Liu et al., 2022b; Spahn et al., 2019). In diesem Projekt wurden STED-Bilder des endoplasmatischen Retikulums (ER) mit einem UNet-RCAN-Netzwerk (Ebrahimi et al., 2023) entrauscht, was bedeutet, dass Bilder mit geringer Intensität aufgenommen und ihr Signal-Rausch-Verhältnis wiederhergestellt wurden (Rahm et al., 2024). Das UNet-RCAN-Netzwerk (Ebrahimi et al., 2023) ist darauf ausgelegt, Hochfrequenzmerkmale in Bildern zu erhalten. Ein Aspekt des Netzes sind sogenannte *channel attention layers* (Zhang et al., 2018b), die die Wichtigkeit von *feature maps* adaptiv kalibrieren. Dadurch werden Hochfrequenzinformationen durchgelassen, die für die Rekonstruktion der feinen Details eines STED-Bildes erforderlich sind. Automatisches Screening von Hyperparameterkombinationen für das Netzwerktraining wurde ermöglicht, um eine optimale Konfiguration zu ermitteln (Rahm et al., 2024). Die Leistung des Netzwerks wurde durch den Vergleich der vorhergesagten Bilder mit den Referenzbildern anhand von Bildähnlichkeitsmetriken bewertet. Das endgültige Netzwerk war in der Lage, Bilder wiederherzustellen, die mit einer Pixelverweildauer von 0,5 µs aufgenommen wurden unter Beibehaltung einer räumlichen Auflösung von 90 nm.
Die Bedeutung der Hyperparameter des Netzwerks wurde mit Hilfe eines Random-Forest-Regressors untersucht, wobei die Wichtigkeit einer geeigneten Lernrate und Eingabegröße für eine hohe Präzision in der Vorhersage hervorgehoben wurde. Es wurden Strategien entwickelt, um strukturelle Signale von Halluzinationen zu unterscheiden, wobei die Ähnlichkeit benachbarter Bilder genutzt wurde, um zwischen rauschbedingten Fluktuationen und strukturellen Veränderungen zu unterscheiden. Aufgrund der Zufälligkeit des Rauschens kam es zu einer geringen Ähnlichkeit zwischen den Bildern. Dies könnte ein Hinweis darauf sein, dass die Fluorophore ausgeblichen sind oder die Struktur sich aus dem Sichtfeld bewegt hat, ein Mögliches Signal an das Mikroskop, die Messung zu beenden. Darüber hinaus ermöglichte die Einteilung von Pixeln basierend auf ihren lokalen Intensitäten eine Kategorisierung der Pixel in Hintergrund-, unsichere und Signalpixel, wodurch die Aufmerksamkeit auf Regionen mit erhöhter Unsicherheit gelenkt werden kann.
Das Modell konnte auf 3D-Bilder angewendet werden, die mit einer *top-hat depletion* Laser PSF aufgenommen wurden, was die Fähigkeit des Netzwerks zur Entrauschung von schnell aufgenommenen 3D-Bildern beweist. Darüber hinaus ließ sich das Modell auf Bilder mit veränderter Morphologie aufgrund einer Behandlung mit Torin1 und Bafilomycin A1 anwenden. Torin1 induziert Stress und Autophagie (Kim et al., 2015), während Bafilomycin A1 den Autophagieprozess bei der Autophagosomenbildung stoppt (Mauvezin et al., 2015). Zur Visualisierung der ER-Struktur wurde Calreticulin-KDEL mehrere Stunden lang abgebildet und mit *ERnet* (Lu et al., 2023) segmentiert. *ERnet* ist ein auf einem neuronalen Netz basierendes Segmentierungswerkzeug, das die Pixel eines Bildes als Hintergrund, planare Strukturen, Röhren und Röhren in planaren Strukturen klassifiziert. Darüber hinaus bietet es

eine Beschreibung des Röhrennetzes auf Basis der Graphentheorie. *ERnet* wurde mit SIM-Bildern demonstriert und konnte in diesem Projekt direkt auf die STED-Bilder angewendet werden. Es wurden weitere Deskriptoren entwickelt, um die strukturellen Veränderungen zu beschreiben. Strukturelle Veränderungen, einschließlich der Anhäufung von planaren Strukturen und einer allgemeinen Abnahme der Strukturdichte, waren offensichtlich, insbesondere innerhalb der ersten Stunde nach der Behandlung mit Torin1 und Bafilomycin A1.

In einem dritten Projekt wurden markierungsfreie Bilder virtuell gefärbt, um die Fluoreszenz des Zellkerns und der Membran zu generieren. Dies ermöglichte die gleichzeitige Visualisierung mehrerer zellulärer Komponenten und hält die zeitliche und spektrale Bandbreite für weitere Fluorophore und ihre Zielproteine frei. Es wurde ein Modell trainiert, das unter verschiedenen Versuchsbedingungen robuste Fähigkeiten aufweist (Liu et al., 2023). Der Schlüsselaspekt dieser Robustheit war eine Datenerweiterung basierend auf Bildtransformationen, die den Probenraum abdeckte. Es wurden verschiedene Eingabe- und Zielmodalitäten und deren Einfluss auf die Modellleistung verglichen. Alle Eingabe- und Zielkombinationen führten zu erfolgreichen Fluoreszenzbildvorhersagen. Entfaltete Fluoreszenzkanäle verbesserte die nachgelagerte Analyseaufgabe der Segmentierung des Zellkerns erheblich. Die Zellkerne wurden mit *cellpose* (Stringer et al., 2021) segmentiert. Da Segmentierungsnetzwerke oft nicht mit quantitativen Phasenbildern kompatibel sind, wird durch den indirekten Weg der Übersetzung des Kanals in Fluoreszenz und der anschließenden Segmentierung des Bildes die Notwendigkeit vermieden, ein Segmentierungsmodell von Grund auf zu trainieren, für das große Mengen von handannotierten Bildern erforderlich gewesen wären, um eine Instanzsegmentierung zu erreichen. Die Segmentierung des Zellkerns und der Membran ermöglicht die automatische Identifizierung von Zellinstanzen und die Quantifizierung ihrer Eigenschaften in Hochdurchsatz-Experimenten.

In einem vierten Projekt wurde eine Software entwickelt, um die Effizienz der Analyse von *single-particle tracking* (SPT) zu verbessern, und eine neue Analysemethode eingeführt. SPT ist ein leistungsfähiges Verfahren zur Untersuchung der Mobilität von Biomolekülen in lebenden Zellen mit einer räumlichen Auflösung im Nanometer- und einer zeitlichen Auflösung im Millisekundenbereich. Mit der Analyse der mittleren quadratischen Verschiebung wurden die Diffusionskoeffizienten der einzelnen Moleküle bestimmt und zwischen immobilen, räumlich beschränkten und frei diffundierenden Molekülen unterschieden. Das Auftreten von Übergängen zwischen verschiedenen Diffusionszuständen innerhalb einzelner Trajektorien wurde mit einer neuartigen Analyse, dem sogenannten *transition counting*, charakterisiert und mit einem *hidden* MARKOV Modell weiter analysiert. Die Methodik wurde auf Einzelmolekül-Trajektorien der Rezeptortyrosinkinasen MET (Rahm et al., 2021) und HER2 (Catapano et al., 2023) angewandt, um deren Reaktionen auf Liganden in HeLa-Zellen zu untersuchen.

Die aus der mittleren quadratischen Verschiebung abgeleitete Analyse ergab, dass ligandenaktivierte MET-Rezeptoren, die durch InlB induziert wurden, im Vergleich zu ruhenden MET-Rezeptoren eine langsamere und eingeschränktere Bewegung aufwießen, wobei sich ein größerer Anteil der Rezeptoren in einem immobilen Zustand befand. Die Übergänge zwischen den Diffusionszuständen innerhalb einzelner Trajektorien wurden durch Zählung der Übergänge quantifiziert, wobei die Wahrscheinlichkeit, dass Partikel den räumlich ein/-ge/-schränk/-ten Zustand verließen, höher war als bei frei diffundierenden und immobilen Partikeln. Nach einer Ligandenbehandlung mit InlB nahm die allgemeine Tendenz zum Übergang in den immobilen Zustand zu. Das gleiche Muster spiegelte sich in dem *hidden* MARKOV Modell wider, bei dem die Daten durch ein Drei-Zustands-Modell mit einem immobilen, einem langsamen und einem schnellen Diffusionszustand charakterisiert wurden. Ähnlich wie der räumlich eingeschränkte Zustand aus der Analyse der mittleren quadratischen Verschiebung wies der mittlere, langsame Zustand höhere Übergangswahrscheinlichkeiten in Richtung immobiler oder schneller Diffusion auf. Dies charakterisierte die räumlich eingeschränkte Diffusion als einen Zwischenzustand. Es ist wahrscheinlich, dass Rezeptoren, die in den intermediären Diffusionszustand übergehen, dazu neigen, entweder immobilisiert zu werden, was eine Vorstufe zur Endozytose darstellt, oder in einen mobilen Zustand zurückzukehren, was möglicherweise die Interaktion mit anderen Proteinen erleichtert. Die Auswirkungen verschiedener Liganden aus der ErbB-Unterfamilie wurden an HER2 untersucht, einem Rezeptor, für den kein eigener Ligand bekannt ist. Seine Aktivierung erfolgt durch Stimulation anderer Rezeptoren und die Bildung von Heterodimeren. Von den getesteten Liganden wurde der Rezeptor am stärksten durch die Behandlung mit EGF aktiviert, was sich in einer erhöhten immobilen Diffusion mit einem Hochpunkt nach 5 Minuten nach Ligandenbehandlung zeigte.

In einem fünften Projekt wurde die räumliche Organisation von Membranrezeptoren untersucht und eine Datenpipeline eingerichtet, die die Extraktion von Rezeptordichten und -verteilungen in Einzelmolekül-Lokalisationsmikroskopie Bildern mit mehreren Zielproteinen ermöglichte. Verschiedene abstandsbasierte Analysen und Vergleiche mit zufällig verteilten Emittern wurden leicht anwendbar gemacht. Die räumliche Anordnung des FGFR-Netzwerks auf der Zellmembran unter dem Einfluss von FGF1 und Heparin wurde untersucht (Schröder et al., 2021). Nach 5 Minuten Ligandenbehandlung nahm die Dichte der FGFR1-4-Rezeptoren ab und ihre Abstände zueinander zu. Trotz der geringeren Anzahl von Rezeptoren auf der Zelloberfläche, die möglicherweise auf eine schnelle Internalisierung der aktivierten Rezeptoren zurückzuführen ist, blieben die Abstände zwischen FGFR1 und FGFR1 und FGFR3 im Vergleich zum Ruhezustand ähnlich. Mit der *k-nearest neighbor* Analyse wurden die Rezeptorverteilungen vom Mikro- bis zum Makrometerberich analysiert. Diese Analyse ergab, dass sowohl in ruhenden als auch in FGF1- und Heparin-stimulierten Zellen Rezeptoranhäufungen vorwiegend in einem Bereich von 200 nm auftreten. Dies könnte auf die potenzielle Akkumulierung von Rezeptoren entlang zellulärer Strukturen wie

Heparansulfat-Proteoglykanen, die als Kofaktor für die Bildung von Rezeptorkomplexen dienen (Ori et al., 2008; Sarrazin et al., 2011), oder mit Clathrin beschichteten Pits, die die Endozytose vermitteln (Li et al., 2016), zurückgeführt werden, da die räumlichen Abmessungen mit den bekannten Abmessungen dieser Strukturen übereinstimmen.

Zusammenfassend berichtet diese Arbeit über die Anwendung von *deep learning* zur Beschleunigung von Einzelmolekül-Lokalisationsexperimenten, zur Ermöglichung STED-Messungen mit hoher Zeitauflösung über mehrere Stunden, zur Verbesserung der *Multiplexing*-Fähigkeit der Fluoreszenzmikroskopie durch virtuelles Färben und über die Beschreibung der Rezeptordynamik und -organisation. Die Synergie, die sich aus der superauflösenden Mikroskopie und neuronalen Netzen ergibt, wird neue Potenziale der superauflösenden Mikroskopie erschließen, indem sie unter anderem den Durchsatz beschleunigt und Zugang zu neuen Kombinationen von Messparametern bietet, die ein tieferes Verständnis der Zellbiologie ermöglichen.

2 Summary

Fluorescence microscopy has significantly impacted our understanding of cell biology. The extension of diffraction-unlimited super-resolution microscopy opened an observation window that allows for the scrutiny of cellular organization at a molecular level. The non-invasive nature of visible light in super-resolution microscopy methods renders them suitable for observations in living cells and organisms (Lakowicz, 2006).
Building upon these advancements, a promising synergy between super-resolution fluorescence microscopy and deep learning becomes evident, extending the capabilities of the imaging methods. Tasks such as image modality translation, restoration, single-molecule fitting, virtual labeling, spectral demixing, and molecular counting, are enabled with high precision (Belthangady et al., 2019).
The techniques explored in this thesis address three critical facets in advanced microscopy, namely the reduction in image acquisition time, saving photon budget during measurement, and increasing the multiplexing capability.

In a first project, the acquisition time in super-resolution DNA points accumulation for imaging in nanoscale topography (DNA-PAINT) (Jungmann et al., 2010) was strongly reduced using a neural network, enabling fast and high-throughput imaging. DNA-PAINT belongs to the single-molecule localization microscopy (SMLM) techniques, that are based on the localization of isolated emitter positions over time. These localized emitters are merged to generate a super-resolved image. The inherent requirement for a large number of imaged frames to completely capture the structure results in low throughput. Increasing the emitter density per frame would expedite imaging, as the underlying structure is captured with less frames. However, this necessitates robust localization methods to disentangle the signal from overlapping point spread functions (PSFs) in diffraction-limited images.
In this work, *DeepSTORM* (Nehme et al., 2018), a neural network based approach that has been demonstrated to obtain super-resolved images from stochastic optical reconstruction microscopy (STORM) (Rust et al., 2006) measurements under the challenging condition of high emitter density, is used to reconstruct super-resolved images from DNA-PAINT measurements (Narayanasamy et al., 2022). Compared to STORM with an exponential decrease in emitter density during measurement, DNA-PAINT data has the advantage of constant emitter densities due to the continuous exchange of labels. The resulting emitter densities of DNA-PAINT fit to the optimal performance window of a trained neural network. In a first step, the capability of *DeepSTORM* was probed to reconstruct super-resolved images from DNA-PAINT measurements with conventional density. A workflow for generating high density DNA-PAINT training and test data from experimental measurements was established, providing an alternative route to simulation-based training. The training dataset involved acquiring movies with low imager strand concentration, local-

ized conventionally, and then artificially creating high density data by summing patches of diffraction-limited emitters along with their localization lists. For the test dataset, ground truth data was obtained by imaging with low emitter densities, and a super-resolution image was generated through single-molecule localizations. The imager strand concentration was increased in the same field of view (FOV), and high density images were processed using *DeepSTORM*. Both low and high density measurements yielded super-resolved images of the same structure. Emitter localizations were extracted through a post-processing step, for which multiple parameters were optimized to retrieve a similar number of localizations compared to the ground truth data. The results were scrutinized with various quality control metrics, including MS-SSIM (Wang et al., 2003), SQUIRREL (Culley et al., 2018), and HAWKMAN (Marsh et al., 2021). Leveraging *DeepSTORM* in high density DNA-PAINT measurements enabled imaging of α-tubulin and TOM20 in mouse brain tissue within one minute per target protein instead of 25 minutes. Although a slight reduction in resolution was noted in the predicted images (44 nm prediction versus 32 nm ground truth), the resolution remained in acceptable range. The trained model is structure-independent, meaning that it is applicable to various targets, given its reliance on single patches from images of SMLM measurement, that contain diffraction-limited emitters without apparent structural positions. Furthermore, DNA-PAINT allows the usage of the same fluorophore for multiple targets, meaning that only one trained neural network is required as a digital add-on to a microscope. Live-cell SMLM also benefits from the high emitter density, as fewer imaged frames are required to generate a super-resolved image, enhancing the temporal resolution. Enabling neural network training using experimental data with low barrier acquisition efforts, a user-friendly pipeline to generate high density training data, and an established workflow for quality control, deep learning enabled DNA-PAINT will be a powerful tool for biological discovery.

In a second project, neural network assisted denoising was implemented into stimulated emission depletion (STED) microscopy (Hell et al., 1994) to facilitate continuous imaging of single live cells for hours with a temporal resolution of seconds. STED microscopy is another powerful super-resolution technique, which operates with high laser powers that cause photobleaching and phototoxicity, constraining the imaging of living cells and their dynamics. With approaches such as stable fluorophores or transient labeling strategies, sequential STED images can be acquired for up to a couple of minutes (Glogger et al., 2022; Liu et al., 2022b; Spahn et al., 2019). Here, STED images of the endoplasmic reticulum (ER) were denoised with a UNet-RCAN network (Ebrahimi et al., 2023), meaning that low-intensity images were captured and their signal-to-noise ratio (SNR) restored (Rahm et al., 2024). The UNet-RCAN network (Ebrahimi et al., 2023) is tailored to preserve high frequency features of the images. One aspect of the network are channel attention layers (Zhang et al., 2018b), designed to adaptively recalibrate the importance of the feature maps. This allows high frequency information to pass, necessary to reconstruct the fine

details of an STED image. A workflow was established for the automated screening of hyperparameter combinations for network training to discern an optimal configuration (Rahm et al., 2024). The performance of the network was assessed by comparing the predicted images to ground truth images with image similarity metrics. The final network was able to restore images acquired with 0.5 µs pixel dwell time, maintaining a spatial resolution of 90 nm.
The importance of network hyperparameters was scrutinized using a random forest regressor, underscoring the role of an appropriate learning rate and input size in achieving high predictive performance. Strategies were devised to discern structural signal from hallucination, leveraging adjacent frame similarity to distinguish between noise-induced fluctuations and structural changes. Due to the randomness of noise, low similarity occurred between the frames. This could potentially serve as an indication that the fluorophores are bleached or the structure moved out of the field of view, triggering a response to the microscope to stop measuring. Furthermore, thresholding the local intensities of the low intensity movie allowed to categorize the pixels into background, uncertain, and signal pixels, gearing the attention to regions with heightened uncertainty.
The model was able to generalize to 3D images acquired with a top-hat depletion laser PSF, demonstrating the ability of the network to denoise rapidly acquired 3D stacks. Furthermore, the model generalized to images with changed morphology due to drug treatment with torin1 and bafilomycin A1. Torin1 induces stress and autophagy (Kim et al., 2015), while bafilomycin A1 halts the autophagy process at autophagosome formation (Mauvezin et al., 2015). To visualize the ER structure, calreticulin-KDEL was imaged for several hours and segmented using *ERnet* (Lu et al., 2023). *ERnet* is a neural network based segmentation tool, classifying the pixels of an image as background, sheets, tubes, and sheet-based tubes. It furthermore offers a graph based description of the tube network. *ERnet* has been demonstrated with SIM images and could be directly applied to the STED images in this project. Further descriptors were developed to describe the structural changes. Structural changes, including sheet accumulation and an overall decrease in structure density, were evident, particularly within the first hour post-drug treatment.

In a third project, label-free images were virtually stained to predict the nucleus and membrane fluorescence. This allowed the visualization of multiple cellular landmarks at once, keeping temporal and spectral bandwidth free for more fluorophores and their targets. A model was trained with robust capabilities across various experimental conditions (Liu et al., 2023). The key aspect of this generalizability was data augmentation that covered the sample space. Different input and target modalities and their influence on model performance were compared. All input and target combinations led to successful fluorescence image predictions. Targeting deconvolved fluorescence channels significantly improved the downstream analysis task of segmenting the nucleus. The nuclei were instance segmented

with *cellpose* (Stringer et al., 2021). As segmentation networks are often not compatible with quantitative phase images, taking an indirect route of translating the channel to fluorescence and then segmenting the image prevents the necessity of training a segmentation model from scratch, where vast amounts of hand-labeled annotations would have been required to achieve instance segmentation. Segmenting the nucleus and membrane will enable the automated identification of cell instances and the quantification of their properties in high throughput experiments.

In a fourth project, a comprehensive framework was developed to enhance the efficiency of analyzing single-particle tracking (SPT) data and a new method of analysis was introduced. SPT is powerful for investigating biomolecule mobility in living cells, with nanometer spatial and millisecond temporal resolution. With mean-squared displacement (MSD) analysis, diffusion coefficients of the single molecules were determined and immobile, confined, and freely diffusing molecules were distinguished. Occurrences of transitions between different diffusion states within single trajectories were characterized with a novel analysis termed transition counting and further analyzed with hidden MARKOV modeling. The methodology was applied to single-molecule trajectories of the receptor tyrosine kinases MET (Rahm et al., 2021) and HER2 (Catapano et al., 2023) to study their responses to ligands in HeLa cells.
Analysis derived from mean-squared displacement revealed that ligand-activated MET receptors, induced by InlB, exhibited a slower and more confined motion with an increased amount of receptors being in an immobile state, compared to resting MET receptors. Diffusion state transitions within single trajectories were quantified by transition counting, highlighting a higher probability of particles leaving the confined state compared to freely diffusing and immobile particles. Upon ligand treatment with InlB, the overall tendency to transfer to the immobile state increased. The same pattern was reflected in hidden MARKOV modeling, where the data was characterized with a three state model with an immobile, slow, and fast diffusion state. Similarly to the confined state from MSD analysis, the intermediate, slow state had higher transition probabilities towards immobile or fast diffusion. This characterized confined diffusion as an intermediate state. It is likely that receptors transitioning to the intermediate diffusion state are prone to either undergoing immobilization as a preliminary phase preceding endocytosis or returning to a mobile state, possibly facilitating interactions with other proteins. The effects of different ligands from the ErbB sub-family were studied on HER2, which is an orphan receptor with no known ligand. Its activation occurs through stimulation of other receptors and the formation of heterodimers. The receptor exhibited strongest activation with EGF treatment among the tested ligands, marked in enhanced immobile diffusion, with a peak after 5 minutes post ligand treatment.

In a fifth project, the spatial organization of membrane receptors were investigated and a data pipeline was established, facilitating the extraction of receptor densities and distributions in multi-target SMLM images. Various distance based analyses and comparisons to randomly distributed emitters were made readily applicable. The spatial arrangement of the FGFR network on the cell membrane under influence of the FGF1 and heparin was studied (Schröder et al., 2021). After 5 minutes of ligand treatment, the density of FGFR1-4 receptors decreased, and their distances from each other increased. Despite less receptors on the cell surface, potentially due to rapid internalization of the activated receptors, the distances between FGFR1 to FGFR1 and FGFR3 remained similar compared to the resting condition. With k-nearest neighbor analysis, receptor distributions were analyzed from a micro to macro level. This analysis unveiled that both in resting as well as in FGF1 and heparin stimulated cells, receptor accumulations appear predominantly within a range of 200 nm. This could be attributed to the potential clustering of receptors along cellular structures like heparan sulfate proteoglycans, which serves as a co-factor for receptor complex formation (Ori et al., 2008; Sarrazin et al., 2011), or clathrin-coated pits, which mediate endocytosis (Li et al., 2016), as the spatial dimensions align with the know dimensions of these structures.

In summary, this thesis reports on the application of deep learning to accelerate SMLM experiments, enable fast and long-time STED imaging, enhancing the multiplexing capability of fluorescence microscopy through virtual staining, and frameworks for describing receptor dynamics and organization. In particular, the synergy that emerges from super-resolution imaging and neural networks will open up new exciting potentials of super-resolution microscopy by, among other things, accelerating throughput and providing access to new combinations of measurement parameters that enable a deeper understanding of cell biology.

Contents

1 Zusammenfassung **i**

2 Summary **vii**

3 Introduction **1**
3.1 A concise history: From perceptrons to convolutional neural networks . . . 2
3.2 Drawbacks and challenges of neural networks . . . 3
3.3 Applications of neural networks in super-resolution microscopy . . . 3
3.3.1 Fast single-molecule localization microscopy . . . 4
3.3.2 Denoising of fluorescence microscopy images . . . 5
3.3.3 Translation between label-free and fluorescence images . . . 7
3.4 Aims of this thesis . . . 8

4 Theory **11**
4.1 Absorption and fluorescence . . . 11
4.2 Fluorophore labeling . . . 14
4.2.1 Fluorescent labels . . . 14
4.2.2 Labeling strategies . . . 15
4.3 Illumination schemes . . . 17
4.4 Diffraction limit . . . 18
4.5 Super-resolution microscopy . . . 19
4.5.1 Stimulated emission depletion microscopy . . . 20
4.5.2 Single-molecule localization microscopy . . . 22
4.5.2.1 DNA-PAINT . . . 24
4.5.2.2 Analysis of SMLM data . . . 26
4.6 Single-particle tracking . . . 28
4.6.1 Diffusion . . . 28
4.6.2 Labeling and imaging strategies for single-particle tracking . . . 29
4.6.3 Localization and tracking . . . 30
4.6.4 Single-particle tracking data analysis . . . 31
4.6.4.1 Mean squared displacement analysis . . . 31
4.6.4.2 Hidden MARKOV modeling . . . 32
4.7 Quantitative phase microscopy . . . 34
4.8 Basic concept of neural networks . . . 37
4.8.1 Multi-layer perceptron . . . 37
4.8.2 Convolutional neural networks . . . 38
4.8.3 Training process . . . 41
4.8.3.1 Loss functions . . . 41
4.8.3.2 Train-validation-test split . . . 42
4.8.3.3 Gradient descent . . . 42
4.8.3.4 Optimizers . . . 43
4.8.3.5 Variance and bias . . . 45
4.9 Network architectures . . . 48
4.9.1 U-Net architecture . . . 48

4.9.2 Residual channel attention network . . . 49
4.10 Quality control metrics . . . 51
4.10.1 Image similarity metrics . . . 51
4.10.1.1 Pixel-independent metrics . . . 51
4.10.1.2 Pixel-dependent metrics . . . 52
4.10.2 Super-resolution microscopy specific metrics . . . 53
4.10.3 Image resolution . . . 55
4.10.4 Metrics for classification tasks . . . 55
4.10.4.1 Basic metrics . . . 56
4.10.4.2 Performance metrics . . . 56
4.10.4.3 Spatial metrics . . . 57

5 Accelerated DNA-PAINT imaging with a neural network 59
5.1 Results and discussion . . . 60
5.1.1 Training and testing of *DeepSTORM* with DNA-PAINT imaging . . . 60
5.1.2 Single-emitter fitting with *DeepSTORM* . . . 62
5.1.3 *DeepSTORM* reconstruction of DNA-PAINT movies in high density tissue . . . 63
5.1.4 Retrieving emitter positions in postprocessing . . . 66
5.2 Methods . . . 69
5.2.1 Sample preparation . . . 69
5.2.2 Labeling . . . 69
5.2.3 Setup . . . 70
5.2.4 DNA-PAINT imaging . . . 70
5.2.5 Generating training patches . . . 71
5.2.6 DeepSTORM training and prediction . . . 71
5.2.7 Image analysis . . . 72

6 Fast and long-time STED imaging of ER nano-structural dynamics in living cells 73
6.1 Results and discussion . . . 75
6.1.1 Denoising pipeline for live cell ER STED images . . . 75
6.1.2 Quality control of STED image restoration . . . 77
6.1.3 Effects of morphological changes on model robustness . . . 78
6.1.4 Strategies to discern structural signals from hallucinations . . . 80
6.1.5 Denoising and segmentation to describe fast and long-term ER dynamics . . . 80
6.2 Methods . . . 86
6.2.1 Sample preparation . . . 86
6.2.2 Image acquisition . . . 86
6.2.3 Image processing and dataset generation . . . 86
6.2.4 Network training . . . 87
6.2.5 Quality control metrics . . . 87
6.2.6 Hyperparameter importance . . . 88
6.2.7 *ERnet* . . . 88
6.2.8 Statistics . . . 88

7 Virtual staining of cellular landmarks 89
7.1 Robust virtual staining for cell instance segmentation . . . 90
7.1.1 Results and discussion . . . 90
7.1.1.1 Virtual staining and nucleus segmentation workflow . . . 90
7.1.1.2 Generalizability towards different experimental conditions 92
7.1.1.3 Changed input and output modalities . . . 95
7.1.2 Methods . . . 97
7.1.2.1 Sample preparation . . . 97
7.1.2.2 Setup and data acquisition . . . 97
7.1.2.3 Data preprocessing . . . 98
7.1.2.4 Neural network architecture . . . 99
7.1.2.5 Model training . . . 100
7.1.2.6 Augmentation . . . 100
7.1.2.7 Segmentation with cellpose . . . 101
7.1.2.8 Evaluation metrics . . . 101
7.1.2.9 Statistics . . . 101
7.2 Virtual staining and denoising of the endoplasmic reticulum . . . 102
7.2.1 Results and discussion . . . 102
7.2.2 Methods . . . 104
7.2.2.1 Cell preparation and imaging . . . 104
7.2.2.2 Data preprocessing . . . 104
7.2.2.3 Neural network architecture and model training . . . 104
7.2.2.4 Evaluation . . . 104

8 A pipeline for the analysis of single-particle tracking data 105
8.1 Results and discussion . . . 106
8.1.1 Analysis pipeline for single-particle tracking data . . . 106
8.1.2 MSD-based analysis of diffusion states from trajectory segments of the MET receptor . . . 108
8.1.3 Transition counts reveal differences in diffusion state transitions for activated MET receptors . . . 110
8.1.4 Hidden MARKOV modeling of single-particle trajectories of MET receptors . . . 112
8.1.5 Temporal response of HER2 in ligand treated cells . . . 113
8.2 Methods . . . 116
8.2.1 A pipeline for the analysis of single-particle tracking data . . . 116
8.2.1.1 Data acquisition . . . 116
8.2.1.2 Single-molecule localization . . . 116
8.2.1.3 Single-particle tracking . . . 116
8.2.1.4 Diffusion coefficient and diffusion state determination . . . 117
8.2.1.5 Transition counts . . . 118
8.2.1.6 Hidden Markov modeling . . . 118
8.2.1.7 Statistical analysis . . . 119
8.2.2 Temporal response of HER2 in ligand treated cells . . . 119
8.2.2.1 Data acquisition . . . 119
8.2.2.2 Time-series analysis . . . 119
8.2.2.3 Statistical analysis . . . 120

9 Spatial organization of membrane receptors 121
9.1 Results and discussion . . . 122
9.1.1 Analysis pipeline to unravel the spatial organization of proteins . . 122
9.1.2 Spatial organization of the FGFR network . . . 123
9.2 Methods . . . 127
9.2.1 Cell preparation and labeling . . . 127
9.2.2 DNA-PAINT data acquisition . . . 128
9.2.3 Localization and clustering . . . 128
9.2.4 Distance analysis . . . 129
9.2.5 Statistical analysis . . . 129

10 Conclusion and outlook 131
10.1 Accelerated DNA-PAINT imaging with a neural network . . . 131
10.2 Fast and long-time STED imaging of ER nano-structural dynamics in living cells . . . 132
10.3 Virtual staining of cellular landmarks . . . 133
10.4 A pipeline for the analysis of single-particle tracking data . . . 133
10.5 Spatial organization of membrane receptors . . . 134
10.6 Deep learning for super-resolution microscopy in the future . . . 134

11 Author contributions 137

12 References 139

13 Appendix 159
13.1 Fast and long-time STED imaging of ER nano-structural dynamics in living cells . . . 159
13.2 Robust virtual staining for cell instance segmentation . . . 164
13.2.1 Data preprocessing . . . 164
13.2.1.1 Phase reconstruction and fluorescence deconvolution . . . 164
13.2.1.2 Registration . . . 165
13.2.1.3 Tiling . . . 166
13.3 A pipeline for the analysis of single-particle tracking data . . . 171
13.4 Spatial organization of membrane receptors . . . 172

14 Acronyms 173

15 List of Figures 177

16 Acknowledgements 181

17 Publications 183

18 Curriculum Vitae 185

3 Introduction

Fluorescence microscopy revolutionized our understanding of cell biology. The extension of diffraction-unlimited super-resolution microscopy opened an observation window that allows to visualize organization in cells on a molecular level. Moreover, the non-invasiveness of visible light makes super-resolution microscopy methods suitable for measurements in living cells and organisms (Lakowicz, 2006).
A clever experimental strategy followed by image reconstruction and analysis enables the relation of observed patterns to function and provides a mechanistic understanding of the biological sample (Belthangady et al., 2019). Data extraction and evaluation from microscopy images is strongly computer-aided due to its quantity and complexity. Recent advances show great potential of applying deep learning techniques in this field to perform tasks such as image modality translation, restoration, single-molecule fitting, artificial labeling, spectral demixing, and molecular counting, with high precision (fig. 1).

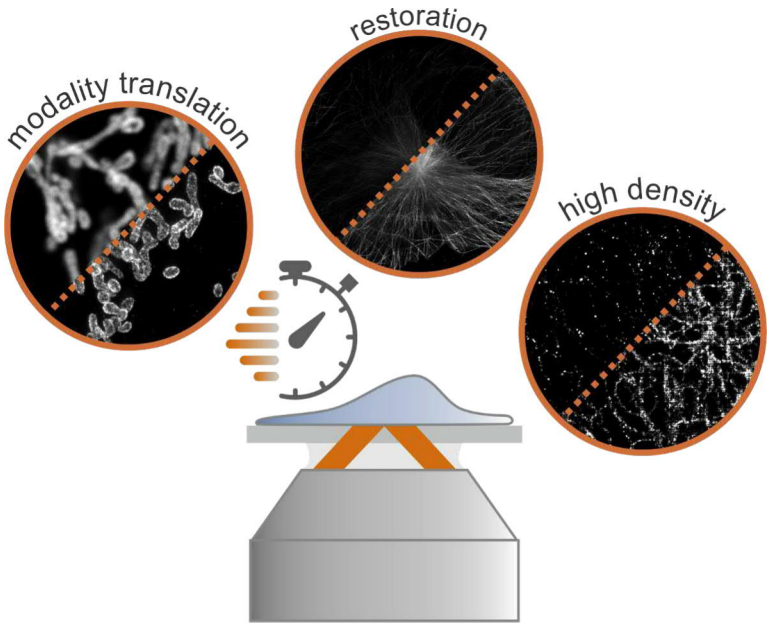

Figure 1: Examples of deep learning enabled tasks for super-resolution microscopy. Modality translation transfers one microscopy modality to another, for example confocal to STED. Image restoration recovers the signal of low signal-to-noise ratio images and generates high signal-to-noise ratio images. Deep learning can be used in single-molecule localization microscopy to fit the point spread functions in high emitter density images. All examples accelerate the imaging process.

Deep learning methods have demonstrated superior performance compared to state-of-the-art classical and machine learning techniques in several fields, with computer vision (the task of interpreting visual data), standing out prominently (Voulodimos et al., 2018). Deep learning allows computational models to learn and represent data with multiple levels of abstraction, resembling the perceptual mechanisms of the brain. Situated in the realm

of machine learning methods, deep learning involves the training of computational models on datasets to autonomously discern complex representations and patterns, enabling the model to make predictions or classifications without the need for explicit programming. The terms “learning” and “training” underscore the capability of acquiring knowledge and improving performance over time through the exposure to data. A model architecture can be employed across various types of data and used for different tasks, due to its inherent flexibility.

3.1 A concise history: From perceptrons to convolutional neural networks

One of the first works of artificial neural networks was from the neurophysiologist Warren McCulluch and mathematician Walter Pitts (McCulloch et al., 1943). They introduced networks that utilize basic elements resembling binary devices with predetermined thresholds. This logic function was characterized by an "all-or-none" nature similar to neuron activity. The perceptron is a more complex neural network model capable of learning from data through a process known as supervised learning (Rosenblatt, 1957). In a perceptron, multiple binary inputs are processed and weights applied to produce a single binary output. However, the perceptrons has limitations, particularly in handling non-linearly separable problems, leading to a period of reduced interest in neural networks. The backpropagation algorithm, a method for training multi-layer neural networks, sparked the interest again (Rumelhart et al., 1986). In this algorithm, the weights in a network are adjusted by propagating errors backward from the output layer to the input layer. With this approach, deep neural networks can effectively learn complex and non-linear relationships.
The name convolutional neural network, a design specifically developed for spatial data, originated with the LeNet neural network, tasked to recognize handwritten digits from images provided by the U.S. Postal Service (LeCun et al., 1998). Convolutional neural network gained further popularity through the ImageNet large scale visual recognition challenge (ILSVRC), where approaches for image classification are benchmarked on a common dataset. In 2012, the CNN network AlexNet reduced the error rate by a margin. In 2015, He et al., 2015 introduced the idea of residual connections and the error rate dropped below human performance.
The popularity of deep learning can be furthermore attributed to the broad availability in data and GPU computing, which enabled a significant acceleration in the training. Frameworks like TensorFlow (Abadi et al., 2016) and PyTorch (Paszke et al., 2019) made deep learning widely applicable.

3.2 Drawbacks and challenges of neural networks

While deep learning methods have shown remarkable success in various domains, they also come with certain drawbacks and challenges when compared to classical approaches.
A considerable quantity of data is necessary for effectively training deep learning models (Belthangady et al., 2019). Large amounts of (labeled) data can be expensive and time-consuming to collect and annotate. Overfitting occurs when a model memorizes the training data but fails to generalize to new data, often due to insufficient data or an excessive number of trainable weights. Despite efforts like regularization and dropout, overfitting may remain a challenge, particularly with large models. Moreover, the performance of deep learning models heavily depends on the quality and representativeness of the training data. While models excel at learning patterns from the data they are trained on, they may struggle to generalize to new, unseen data, for instance different imaging conditions and microscopes. Furthermore, deep learning models have a "black box" nature. Unlike classical algorithms that are based on explicit rules, the complexity of the deep learning models make it challenging to interpret how they arrive at specific decisions or predictions. Training deep learning models is time- and energy-consuming and requires specific computational resources like graphics processing units (GPUs). However, once a model has been trained, inference is usually executed quickly.
In conclusion, while deep learning has revolutionized various fields, its drawbacks should be carefully considered. Balancing these drawbacks with the remarkable capabilities of deep learning models is crucial for making informed decisions about whether and when to apply this technique.

3.3 Applications of neural networks in super-resolution microscopy

The scope of successful applications of neural networks in super-resolution microscopy is extensive. Neural networks have for instance been successfully applied for spectral demixing. Hershko et al., 2019 were able to separate the point spread functions (PSFs) from two quantum dots of different wavelengths. With wavelength-dependent aberration, four colors could be distinguished. Additional advancements include the demixing of signals from two dyes in STORM imaging (Kim et al., 2019b) and the demixing of four dyes without the necessity for PSF engineering (Jiang et al., 2023).
Long short-term memory (LSTM) networks, specialized for time-series data (Hochreiter et al., 1997), have proven instrumental in deciphering the stoichiometry of fluorescently labeled protein complexes (Xu et al., 2019). Bleaching curves of protein complexes are analyzed to assign stoichiometries, even under challenging conditions such as low signal-to-noise ratios, substantial sample variability, and brief bleaching intervals (Wills et al., 2023). Temporal resolution enhancement is achieved through various approaches, for instance

with content-aware frame interpolation (CAFI), which interpolates frames in a time-course dataset (Priessner et al., 2024). The network ANNA-PALM reconstructs super-resolution views from sparse, rapidly acquired localization images and/or widefield images (Ouyang et al., 2018). Super-resolved videos with continuous structure are reconstructed from single-molecule localization images with *DBlink* (Saguy et al., 2023).
Cross-modality image translation uses one microscopy technique as an input and another microscopy technique as an output. Structured illumination microscopy (SIM) images can be predicted from wide-field fluorescence microscopy (Zhuge et al., 2021) or STED images from confocal images (Bouchard et al., 2023; Zhang et al., 2018b). Switching between different stains is also possible, for example from actin to nucleus staining (Chamier et al., 2021).

As this work concentrates on fast single-molecule localization microscopy with a neural network based fitting approach, denoising of STED images, and virtual staining of quantitative phase images (QPI), these topics are further elaborated in subsequent sections.

3.3.1 Fast single-molecule localization microscopy

Single-molecule localization microscopy (SMLM) is a super-resolution microscopy method enabling the visualization of structure with nanometer accuracy. In this method, fluorescent labels are attached to the target structure. Large number of images are acquired, each containing a fraction of labels that emit fluorescent signals, so that the emitted signals do not overlap. This allows precise localization of the emitters per imaged frame. Merging the localizations of all imaged frames generates a super-resolved image of the biological structure the labels are attached to.
The need for large amounts of images in a SMLM measurement results in low throughput and poor time resolution, especially critical when imaging live cells. The process can be sped up by activating more emitters per image. A higher density of emitters can be processed per image and the whole target is captured quicker. This requires sophisticated localization algorithms termed "multi-emitter" fitting algorithms, which are able to entangle overlapping PSF signals (Gazagnes et al., 2017; Huang et al., 2011; Ovesny et al., 2014). Maximum likelihood estimation (MLE) is a classical approach that assigns the intensity in a region to n emitters with close spatial proximity (Huang et al., 2011).

Deep learning approaches typically use the diffraction limited images of the SMLM measurement as input, which is mapped to a super-resolved image of the emitters. With a post-processing step, localizations are retrieved from the super-resolved image. This approach outperforms classical methods by a margin in terms of accuracy. Once a model is established, localization is additionally less time consuming, as classical approaches are computationally demanding.

Boyd et al., 2018 demonstrated fitting in 2D and 3D on datasets of the SMLM challenge (Sage et al., 2019), a collection of SMLM datasets of different imaging conditions, to probe and compare localization algorithms, and was orders of magnitudes faster than classical approaches with comparable accuracy.
Kim et al., 2019b demonstrated the potential of extracting the axial position of emitters in STORM images without PSF engineering, solely based on the increased width of a PSF when being out of focus.
Nehme et al., 2018 focused on 2D images and outperformed classical approaches with *DeepSTORM* in high density conditions. A encoder-decoder architecture maps diffraction limited images of emitters to super-resolved images. This approach was extended to dense 3D data, where the PSF design to capture 3D information was optimized along with the localization model (Nehme et al., 2020).
The deep context dependent (DECODE) method (Speiser et al., 2021) is a 3D localization tool that uses a calibration measurement, for instance with an astigmatism lens, along with a model of the microscope to simulate emitters, that serve as training data. Instead of using one image as input, a sequence of multiple frames is used to enhance the performance.

3.3.2 Denoising of fluorescence microscopy images

In fluorescence microscopy, trade-offs must be made between spatial resolution, light exposure, and imaging speed (Weigert et al., 2018). Live cell imaging is especially tricky, as the sample is dynamic and requires appropriate temporal resolution and suffers from the phototoxicity of light. This compromises photon collection, crucial for spatial resolution, leading to increased image noise and a discrepancy between the imaged signal and the true signal (Laine et al., 2021).
Image noise in fluorescence microscopy is mainly caused by shot noise and detector noise. The shot noise arises from the random nature of photon arrival, which follows a Poisson distribution. The distribution describes the probability of a given number of photon arrivals occurring in a fixed time interval. This causes fluctuations in the number of photons detected within a certain time frame. When the average number of photons is large, shot noise becomes less significant compared to the recorded signal. However, at low photon counts, shot noise introduces strong variability into the measured signal. The detector noise typically follows a Gaussian distribution independent of the underlying signal. Depending on the detector, signal-dependent noise plays a role and more complex noise models are required.
Noise has two interesting properties. Noise occurs independently for each pixel. Additionally, the noise is centered around the true signal, sometimes yielding higher and sometimes lower values. This means that the result converges towards the true signal by averaging many acquisitions.

The task of denoising is to take the noisy input and return an approximation of the true signal. This enables imaging with conditions that generate noisy images, which undergo reconstruction to attain a high signal-to-noise ratio. Different denoising strategies have been developed from classical denoising algorithms to neural network based approaches.
Classical approaches include GAUSSIAN filtering, where a GAUSSIAN kernel is convolved with an image. High-frequency features are removed, but at the same time this causes the loss of fine-details and the image appears blurry. Median filters replaces the value of each pixel with the median value of its neighborhood. This method is effective for removing salt-and-pepper noise, which appears as isolated bright and dark pixels. Methods such as non-local means (NLM) (Boulanger et al., 2010; Buades et al., 2005) and block matching and 3D (BM3D) filtering (Hasan et al., 2018) have enhanced structural preservation capabilities. They identify similar patches in an image and group them together. In NLM, the patches are averaged to reduce noise. BM3D filtering applies a row of transformations, thresholding, and filters to remove the noise.
Deep learning has shown great potential for denoising fluorescence images (Chamier et al., 2021; Weigert et al., 2018; Zhang et al., 2018b). These methods learn to discern how an image is expected to appear within the context of the particular dataset used for training (Weigert et al., 2018). As a result, these methods tend to yield optimal outcomes for datasets that closely resemble the training data. Deep learning denoising methods can be divided into supervised and self-supervised techniques. Supervised techniques require matching noisy and clear image pairs. These can either be experimentally recorded or simulated, if the simulation resembles the experiments well enough. Self-supervised techniques only require the noisy input. This has the advantage that no clear image is required, which can be challenging and time consuming to obtain. However, supervised methods outperform self-supervised methods, as they have more information available (Chamier et al., 2021).
Noise2Noise (N2N) is a self-supervised denoiser that requires the training on image pairs with independent noise (Lehtinen et al., 2018). Two versions of noisy images, which contain same signal but different versions of noise, are mapped with a neural network. As the network can not learn to perfectly predict the noise due to its randomness, it converges to a denoised image. Noise to void (N2V) eliminates the need of requiring image pairs and uses a masking approach to predict the masked pixels in patches of the image based on local information of the neighboring pixels (Krull et al., 2018). N2V has been used to denoise confocal images of F-actin in ovarian carcinoma cells (Chamier et al., 2021). Furthermore, it has been applied to denoise STED images of the nucleus (Hajiabadi et al., 2022). There are also approaches focusing on reducing the amount of training data and only requiring one noisy image, called blind zero-shot denoisers (Lequyer et al., 2022). If the results from a self-supervised method are satisfying, they shine due to not requiring ground truth data for training and the reduced amount of training instances required.

Weigert et al., 2018 demonstrated the potential of supervised denoising on different microscopy techniques including laser scanning microscopy, light sheet microscopy, and super-resolution radial fluctuations (SRRF). The method, termed content aware image restoration (CARE), uses a U-Net architecture (Ronneberger et al., 2015) and outperformed classical approaches for denoising fluorescence images enabling 60x reduced light exposure. SIM images have been successfully denoised using various supervised architectures (Jin et al., 2020; Shah et al., 2021; Zhang et al., 2018b). An RCAN network (Zhang et al., 2018b) is utilized in Chen et al., 2021 for mapping mitochondria SIM images acquired at low illumination (4.2 W cm^{-2}) to high illuminated images (457 W cm^{-2}). The network has also been used to translate the microscopy modality from confocal to STED. Ebrahimi et al., 2023 combined the RCAN architecture with a U-Net to denoise STED images while maintaining high frequency information. Tubulin and mitochondria were imaged in HeLa cells for multiple minutes with a pixel dwell time of 1 µs instead of 90 µs. Furthermore, they demonstrated that their approach outperforms cross-modality image translation (confocal to STED) and deconvolution.

3.3.3 Translation between label-free and fluorescence images

The process of generating fluorescence images from label-free images is termed "virtual staining" (Chamier et al., 2021). Label-free methods, such as bright-field, phase, and differential interference contrast (DIC) imaging, contain information about cellular organization, but lack the clear contrast and specificity of fluorescence microscopy. However, label-free imaging is less phototoxic and photobleaching is not an issue. Generating fluorescence images from label-free images allows to detect fluorescence signal over long periods of time. Furthermore, stains for different organelles can be predicted at once, increasing the multiplexing capability.
Studies have used the bright-field channel or qualitative phase information with phase contrast or DIC to predict fluorescence (Chamier et al., 2021; Christiansen et al., 2018; Ounkomol et al., 2018). For instance, Ounkomol et al., 2018 trained several models to predict the fluorescence targets from DIC images of more than 10 organelles, including nucleus, cell membrane, mitochondria, golgi apparatus, and the endoplasmic reticulum. The level of detail captured in the predictions varied based on the extent to which information about the organelles was encoded in the label-free image. Jo et al., 2021 showed that three-dimensional refractive index tomograms acquired with a holographic microscope significantly improved the predictions for multiple organelles and cell types compared to bright-field and qualitative phase imaging, rendering quantitative phase imaging as a more informative input channel for cross-modality inference.

3.4 Aims of this thesis

This thesis focuses on applying deep learning techniques to enable accelerated SMLM experiments, fast and long-time STED imaging, and automated fluorescence staining of wide-field images with ensured image quality. Furthermore, the study extends the analytical toolbox for characterizing mobility patterns and the spatial organizations of membrane receptors. Each objective is addressed in a dedicated chapter, with the overarching goals outlined in the subsequent paragraphs.

In single-molecule localization microscopy, a movie of stochastically blinking emitters is recorded and the emitters localized to generate a super-resolved image. This method is time consuming, as the emitters must be sparse per acquired image. An increased emitter density per imaged frame requires advanced methods of entangling the overlapping PSF signals. *DeepSTORM*, a deep learning based "multi-emitter" fitting approach, has been demonstrated to be a powerful localizer on STORM data. Here, the tool was applied to DNA-PAINT data of tubulin and mitochondria in mouse brain tissue in chapter 5. Acquisition time was reduced from 1 hour to 1 minute while maintaining comparable image quality. A comprehensive workflow of acquiring the training data based on experimental measurements and the test data, which contains both the high emitter density images and the ground truth, is introduced. This concept furthermore enabled live-cell SMLM of structural dynamics.

Fast and long-time STED imaging of living cells was enabled with a deep learning based denoising approach in chapter 6. The focus was on the endoplasmic reticulum dynamics in the context of induced autophagy. Torin1, a drug that induces stress and autophagy in mammalian cells was simultaneously added with bafilomycin A1, a substance that inhibits the lysosome machinery responsible for breaking down cellular components. This allowed to activate the ER-phagy process and the observation of autophagosomes as they increased in size over time. This process was imaged for multiple hours with high spatiotemporal resolution. The sheets and tubes of the ER were segmented and morphological changes quantified over time.

Furthermore, a pipeline for robust virtual staining of the membrane and nucleus in high-throughput experiments on a wide-field microscope with a neural network is introduced in chapter 7. The model robustness was tested across varying experimental conditions. The predicted fluorescence images were segmented, which is a step towards single-cell characterization in high-throughput experiments.

The intricate and time-consuming analysis of single-particle tracking data of receptor movements on cell membranes was simplified by introducing a pipeline that encompasses mean-squared displacement analysis and diffusion state classification in chapter 8. A novel analysis termed "transition counting" was developed that focused on transitions between different diffusion states of receptors. Hidden MARKOV modeling, which also focuses on state transitions, was seamlessly integrated into this framework. The pipeline was applied to single-particle tracking (SPT) data of the hepatocyte growth factor receptor (MET) and human epidermal growth factor receptor 2 (HER2) in living HeLa cells, to reveal information on receptor functional states.

Chapter 9 introduces methodologies for describing the spatial relationships of multiple membrane receptors in SMLM experiments. The pipeline was employed to study the fibroblast growth factor receptor (FGFR) network, which is involved in numerous intracellular pathways, including cell growth, differentiation, and survival. Effects of FGF1 ligand treatment on the spatial arrangement were investigated.

4 Theory

This chapter gives a comprehensive overview of the fundamental concept of fluorescence and its utilization in the field of microscopy. The constraint of conventional light microscopy regarding its limit of spatial resolution is explained, followed by super-resolution microscopy methods that overcome this limitation. Methods used within the scope of this study are highlighted and techniques for data analysis explained. Furthermore, the chapter introduces the foundational principles of neural networks, placing a particular emphasis on convolutional neural networks and diverse network architectures. Lastly, principles of network performance assessments are explained.

4.1 Absorption and fluorescence

Light can be described as an electromagnetic wave that interacts with matter in different ways such as reflection, scattering, refraction, and absorption (Lakowicz, 2006). Absorption is a process in which the photon energy of the electromagnetic wave is taken in by matter and is converted into internal energy. A relationship between the energy E of a photon and its frequency ν or wavelength λ is provided by PLANCK's constant h (eq. 1) (Planck, 1901). The prerequisite for for the photon to be absorbed is that the energy of the photon must fit the energy difference ΔE between an electronic ground state and an excited state of the absorber and must correspond to a multiple of h multiplied by the frequency of the wave.

$$\Delta E = E_1 - E_0 = h \cdot \nu = \frac{h \cdot c}{\lambda} \tag{1}$$

The effectiveness of absorption is given by the LAMBERT-BEER law that describes a linear dependence of absorption with the molar extinction coefficient ϵ, the concentration c, and the path length d of a sample (eq. 2) (Beer, 1852; Lambert, 1852). In absorption spectroscopy, the sample is irradiated with light of intensity I_0. The intensity of the light I after passing through the sample is measured to determine the attenuation of intensity, which is termed extinction E. Absorption, scattering, diffraction, and reflection effects contribute to the extinction. Scattering and diffraction are often negligible and reference measurements allow the determination of reflection contributions. This provides a way to access information about the absorption properties of a sample.

$$E = lg\left(\frac{I_0}{I}\right) = \epsilon \cdot c \cdot d \tag{2}$$

The possible relaxation paths of absorbed energy are shown in a JABLONSKI diagram with corresponding time scales (fig. 2) (Lakowicz, 2006). The diagram showcases various electronic states of a molecule, including the electronic ground state S_0, excited state S_1,

and a triplet state T_1, each comprising multiple vibrational levels denoted as v_i. Upon absorption, the electron undergoes a transition to an excited state and can relax to the ground state via several pathways. Relaxation to the lowest vibrational level within an electronic state is called vibrational relaxation. The transition to lower electronic states is termed internal conversion. Relaxation from the lowest vibrational state of S_1 back to the electronic ground state can be accompanied by emission of a photon and is called fluorescence. Emission of fluorescence requires vibrational relaxation and internal conversion to the lowest vibrational level of S_1. An alternative pathway back to the electronic ground state is the transition to a triplet state T_i, called intersystem crossing. Relaxation to the ground state with radiation is called phosphorescence. Since the transition between a triplet and singlet state is spin-forbidden, the conversion exhibits smaller transition rates. Alternatively, an electron can return from the S_1 or T_1 state to the ground state by non-radiative internal conversion or intersystem crossing.

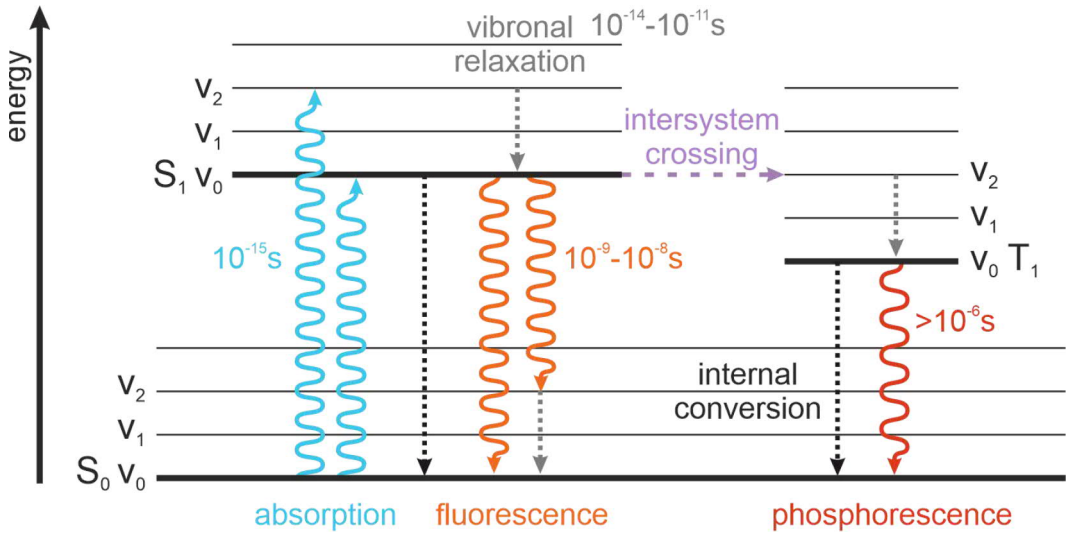

Figure 2: Jablonski diagram. A molecule is excited from the electronic ground state S_0 to higher electronic states S_i by the absorption of a photon. Within an electronic state the molecule can relax to lower vibrational levels via energy exchange with the surroundings called vibrational relaxation. It can relax to a lower electronic state via internal conversion. The transition is called intersystem crossing if the states differ in multiplicity ($S_i \rightarrow T_i$, $T_i \rightarrow S_i$). Radiative relaxation from T_1 to S_0 is termed phosphorescence and from S_1 to S_0 fluorescence.

The occupation of the states is temperature-dependent and given by the BOLTZMANN distribution (eq. 3), which is an exponential function dependent on the energy difference between two states ΔE, the BOLTZMANN constant k_B, and the temperature T (Banwell, 1983). N is the number of particles in a state. The indices j and i specify the state, whereas the state j is energetically higher. Higher energy differences between the states or small temperatures render lower occupancy of the upper state.

$$\frac{N_j}{N_i} = exp\left(\frac{-\Delta E}{k_B T}\right) \tag{3}$$

The probability of transitions between different vibrational states is further described by the FRANCK-CONDON principle (Condon, 1926; Franck et al., 1926). A vertical transition is more likely if the wave functions of two vibrational states overlap more significantly. The intensity of transition is proportional to the square of the overlap integral I between two vibrational wavefunctions $\Psi^{v'}$ and $\Psi^{v''}$ (eq. 4) (Dunbrack, 1986). Fig. 3 A depicts the FRANCK-CONDON principle in a molecule with MORSE-like potential energy functions of an electronic ground state S_0 and excited state S_1. The transition probabilities and occupations according to BOLTZMANN render the shape of the absorption and fluorescence spectrum (fig. 3 B).

$$I^{v'v''} \propto \left\{ \int \Psi^{v'} \cdot \Psi^{v''} dr \right\}^2 \tag{4}$$

According to KASHA's rule, photon emission occurs from the lowest vibrational state of S_1 (Kasha, 1950). This induces a bathochromic shift of the emitted light relative to the absorbed light, called STOKES shift (Stokes, 1852). The similarity of vibrational energy functions between the electronic states induces symmetry in the absorption and emission spectra (Lakowicz, 2006).

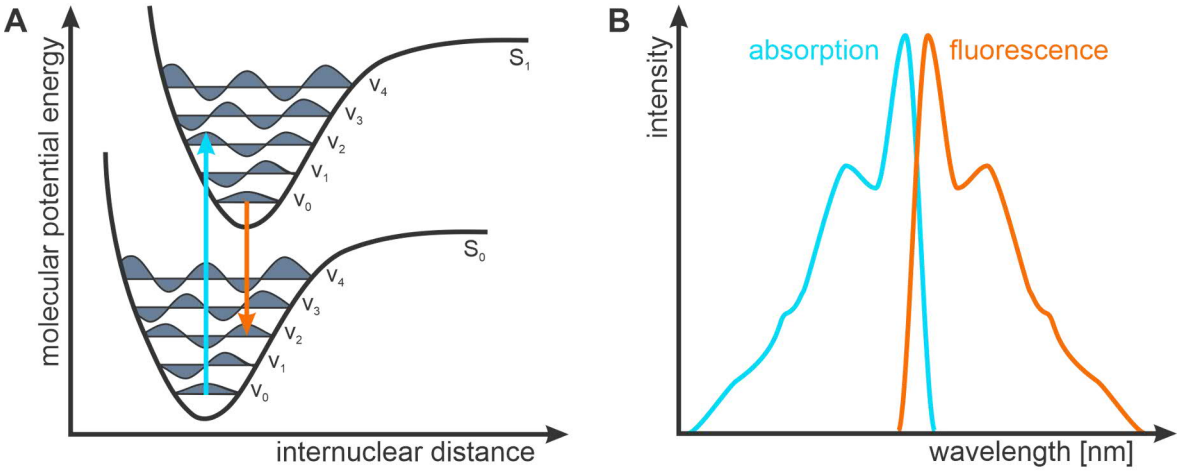

Figure 3: Franck-Condon principle. A The intensity of state transition between the electronic ground state S_0 and excited state S_1 depend on the overlap of the vibrational wave functions. Radiative relaxation occurs from the lowest vibrational state of S_1 to the electronic ground state. **B** The relaxation to the lowest vibrational state of S_1 induces a red shift in emission and the similarity of vibrational energy functions of the electronic states symmetry in the absorption and emission spectra.

4.2 Fluorophore labeling

Fluorescence microscopy is a powerful tool for studying the properties of biomolecules, as the non-invasiveness of light allows the examination of living samples. Most biomolecules contain intrinsic fluorescence in form of aromatic amino acids. However, to create a high contrast image of a specific target, a labeling strategy is needed. This requires binding of a fluorescent probe to a target molecule. Desirable properties of a fluorescent probe are specificity, brightness, photostability, a small size, non-toxicity to the cell and live-cell compatibility. A broad toolbox of labels and ways to attach them to the target structure exist; both must be selected appropriately for the use case.

4.2.1 Fluorescent labels

Organic fluorophores absorb and emit light due to their conjugated π-electron system. This can be approximated with the quantum mechanics model of a particle in a box (Kuhn, 1949). The energy levels are described with eq. 5, with n being the number of energy levels, h the Planck constant, m the mass of the electron, and L the length of the box. Upon absorbing light, a molecule transitions from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO). The required energy difference ΔE is described in eq. 6, with the number of π electrons N in the box. The energy can be expressed as a wavelength (eq. 1). The wavelength of the absorbed and emitted light depends on the length of the conjugated system.

$$E_n = \frac{n^2 h^2}{8mL^2} \tag{5}$$

$$\Delta E = \frac{h^2}{8mL^2}(N+1) \tag{6}$$

Organic fluorophores exhibit high brightness and photostability, resulting in strong and long-lasting fluorescence signal (Requejo-Isidro, 2013; Resch-Genger et al., 2008). With ∼1 nm in size they are comparably small labels. Near-infrared-emitting fluorophores are particularly biocompatible due to their lower phototoxicity (Wäldchen et al., 2015). Commonly used organic fluorophore classes are oxazines, carbocyanines, carbopyronines, and rhodamines (fig. 4) (Requejo-Isidro, 2013). Other labels are quantum dots (Resch-Genger et al., 2008) and fluorescent proteins (FP) (Odell et al., 2013).

The first discovered FP is the green fluorescent protein (GFP), which was originally isolated from the jellyfish *Aequorea victoria* (Shimomura et al., 1962) (fig. 5). GFP has a chromophore composed of three amino acids, which is encased within a β-barrel structure. A variety of FPs are available covering the range of visible light (Matz et al., 1999; Shaner et al., 2004). Compared to organic dyes they are less photostable and bright, but allow for

stoichiometric labeling and are live-cell compatible (Toseland, 2013).

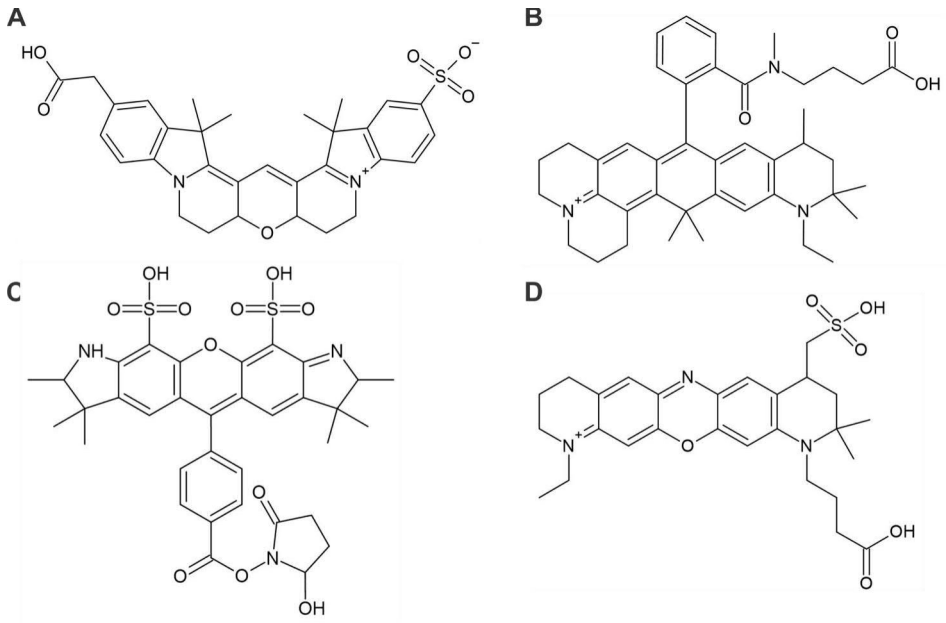

Figure 4: Exemplary structures of organic fluorophores. **A** Cy3B is a yellow-absorbing dimethine cyanine. **B** ATTO 647N is a red-absorbing carbopyronine. **C** Alexa Fluor 532 is a green-absoring rhodamine. **D** ATTO 655 is a red-absoring oxazines.

Figure 5: Crystal structure of GFP from the jellyfish *Aequorea victoria*. The β-barrel of the protein shields the chromophore, which is built by the three amino acids S65, Y66, and G67 (from the protein databank, PDB: 4KW4) (Barnard et al., 2014).

4.2.2 Labeling strategies

Genetic engineering allows the fusion of a fluorescent protein to a target biomolecule. Typically to the N- or C-terminus of a target protein the fluorescent protein is fused. This facilitates stoichiometric labeling and is live-cell compatible (fig. 6 A). The introduction of fusion proteins into cells is achieved through plasmid transfection, but often results in variable expression levels. This variability represents a significant drawback, given that overexpression has been demonstrated to impact both protein organization and cellular

function (Gibson et al., 2013). Genomic engineering techniques like CRISPR/Cas allow for direct insertion of fluorescent protein DNA into the genomic DNA, maintaining natural and consistent protein expression levels controlled by native promoters (Ran et al., 2013).

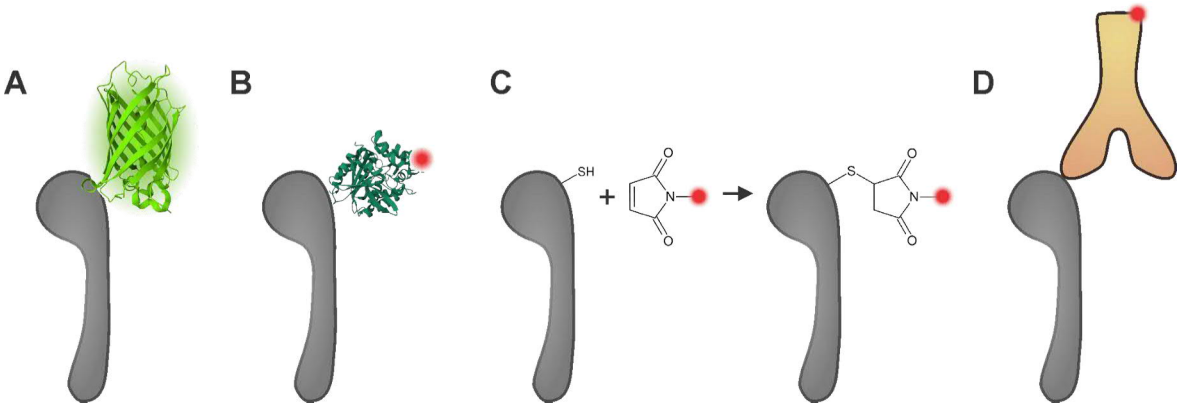

Figure 6: Fluorescence labeling strategies. **A** A FP is fused to the protein of interest. **B** The protein of interest is fused to a HaloTag, which covalently binds an organic flurophore. **C** The cystein residue of the protein of interest reacts with a maleimide-coupled organic dye. **D** An AB carries an organic dye and is attaches to the protein of interest.

Labeling with organic fluorophores is possible with a range of strategies, including protein tags, site-specific labeling through chemical reactions, and immunofluorescence.
Protein tags, such as SNAP (Keppler et al., 2004), HaloTag (Los et al., 2008), and CLIP (Gautier et al., 2008), are genetically encoded with the protein of interest. They typically form a covalent connection between the protein of interest and an organic dye (fig. 6 B). This allows for stoichiometric labeling. Furthermore, these tags have orthogonal substrate specificity and are compatible with live-cell imaging. The HaloTag is derived from a bacterial hydrolase enzyme and with 33 kDa in size it is comparable to the size of a fluorescent protein (Los et al., 2008). Fluorescent dyes that are bound to a chloroalkane linker can be fused to the tag as it carries an activate site that reacts with the linker. The reaction is fast and irreversible under physiological conditions. Transiently binding HaloTags have been recently developed and their suitability for single-molecule localization microscopy is discussed in chapter 4.5.2 (Kompa et al., 2023). A method for directly labeling proteins with fluorescent organic dyes uses N-hydroxysuccinimide (NHS) ester-modified dyes (Zhang et al., 2018c). These NHS esters form covalent bonds with primary amines, such as those found in lysine residues or N-terminal amino groups of proteins. As primary amines are numerous, multiple fluorophores might be attached to the protein of interest, which increases the brightness and therefore the signal-to-noise ratio. At the same time this causes lack in specificity and results in non-stoichiometric and site-unspecific labeling (Dempsey et al., 2018). An alternative approach is the reaction of cysteine residues with maleimide-coupled fluorophores (Zhang et al., 2018c) (fig. 6 C). Cysteine residues are comparatively less abundant within proteins and their presence or absence in the amino acid sequence of the protein can be manipulated to provide site-specific labeling. Furthermore, an unnatural

amino acid can be introduced in the protein of interest and connected to the organic fluorophore via click-chemistry, for example with azid-alkin cycloaddition (Kolb et al., 2001). Immunofluorescence uses antibodies (AB) to attach organic dyes to a target molecule (Coons et al., 1941) (fig. 6 D). In indirect immunofluorescence, a primary antibody binds to the target and a secondary antibody, which carries a fluorophore, binds to the primary antibody. This provides versatility and cost-effectiveness due to the diverse array of secondary antibodies that can be combined (Odell et al., 2013). As antibodies are substantially larger in size (10-15 nm), they cause an offset between the target of interest and the fluorophore (Klein et al., 2009). This can be reduced with direct immunofluorescence, which uses only a primary antibody modified with a fluorophore (Odell et al., 2013). The labeling offset can be further reduced by using antibody fragments such as the Fab fragment with a size of 7 nm (Klein et al., 2009) or nanobodies with a size of around 4 nm (Chames et al., 2020).

4.3 Illumination schemes

Different illumination schemes are common for fluorescence microscopy (Sanderson et al., 2014). With widefield illumination, the laser light is focused on the back of the focal plane of the objective to illuminate the entire FOV (fig. 7 A). In an epifluorescence microscope, the light emitted by the sample is collected by the same objective for illumination. The fluorescence light is split from the excitation light with a dichroic beamsplitter and directed to an electron-multiplying charge-coupled device (EMCCD) camera. Since the entire sample is illuminated, fluorescence signal is collected from multiple planes, which reduces the signal-to-noise-ratio due to signal from out-of-focus planes. The background can be reduced by axially limiting the illumination using total internal reflection fluorescence (TIRF) (fig. 7 C) (Ambrose, 1956) or a highly inclined laminated optical sheet (HILO) (fig. 7 B) (Tokunaga et al., 2008).

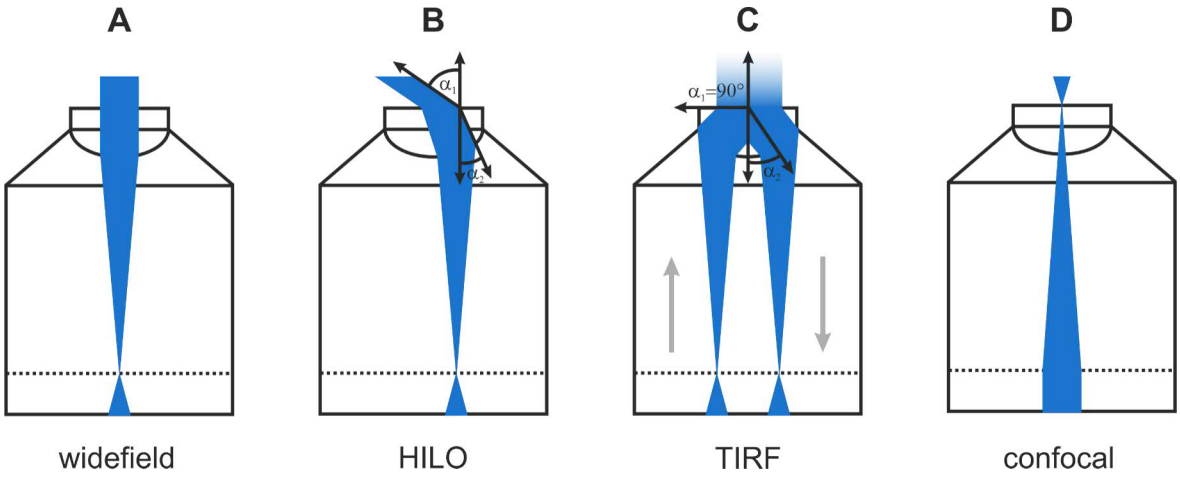

Figure 7: Illumination schemes. **A** Widefield, **B** HILO, **C** TIRF, and **D** confocal illumination.

To achieve total internal reflection for TIRF microscopy, the incident beam is shifted even further to the periphery of the objective. When a critical angle θ_{crit} at the interface between coverglass and sample is surpassed, total internal reflection occurs, with n_1 being the refractive index of the coverslip and n_2 of the sample (eq. 7). Approximately ∼200 nm from the interface between the coverglass and the sample are illuminated (Ambrose, 1956). This effectively diminishes fluorescence signals originating above the excited region. The illuminated area is characterized by an exponential intensity decay, commonly referred to as the evanescent field (Axelrod, 1981) (eq. 8). The intensity of light along the z-axis $I(z)$ is dependent on the intensity of the incident beam $I(0)$ at the glass-water surface. The penetration depth d is the point in z direction where the initial intensity decreased to $1/e$. TIRF microscopy is well suited to, for example, study the cell membrane but is limited to the study of samples at their surfaces (Axelrod, 1981; Tonzani, 2009).

$$\theta_{crit} = sin^{-1}\left(\frac{n_2}{n_1}\right) \tag{7}$$

$$I(z) = I(0) \cdot e^{-z/d} \tag{8}$$

To achieve HILO illumination, the incident beam is focused on the back focal plane of the objective, but spatially shifted to the periphery of the objective (Tokunaga et al., 2008). A thin optical sheet is illuminated through the sample. The thickness of the sheet dz depends on the illumination angle s and illumination diameter R (eq. 9).

$$dz = \frac{R}{tan(s)} \tag{9}$$

With confocal illumination, a laser beam excites only a small, diffraction-limited volume of the specimen and the image is generated by scanning the spot over the FOV (fig. 7 D) (Heilker et al., 2005; Sanderson et al., 2014). Only fluorophores within a tightly confined volume of ∼one femtoliter are excited, effectively reducing background signal. Emitted light from the sample is directed through a pinhole, which diminishes signal from out of focus planes, thereby improving the signal-to-noise ratio. The emitted photons are detected by a photomultiplier. This approach attains a lateral resolution of ∼200 nm and an axial resolution of ∼500 nm. The temporal resolution of confocal laser scanning microscopy is inherently constrained by the scanning process.

4.4 Diffraction limit

The spatial resolution of an optical microscope is limited by diffraction. Abbe postulated that the minimal resolvable distance between two structures d_{min} depends on the wavelength of light λ traveling trough a medium with refractive index n and the half-angle of the

objective α (eq. 10) (Abbe, 1873). The term $n \cdot sin\alpha$ can be summarized to the numerical aperture NA. A resolution up to 200 nm can be reached with a classical optical microscope.

$$d_{min} = \frac{\lambda}{2n \cdot sin\alpha} = \frac{\lambda}{2NA} \tag{10}$$

When light is emitted from a point source, it passes through the optics of the microscope. Due to diffraction, AIRY discs are created with a central intensity peak surrounded by rings of decreasing intensity (Airy, 1895). The image of the point source, when observed through the microscope, is characterized by a point spread function (PSF). The PSF describes how a point source of light transforms into a finite-sized spot or pattern. RAYLEIGH proposed a resolution criterion based on the separation between two point sources. The point sources are resolved if the maximum of one PSF falls on the first minimum of the other PSF, described by minimal distance d_{min} in eq. 11 (Rayleigh, 1896).

$$d_{min} = \frac{0.61\lambda}{NA} \tag{11}$$

4.5 Super-resolution microscopy

Super-resolution microscopy (SRM) techniques improve the resolution by more than an order of magnitude compared to diffraction-limited fluorescence microscopy and close the gap of resolution towards electron microscopy (Liu et al., 2022a). Because biological structures are less than 200 nm in size at the molecular level, SRM is widely used in biomedical research. Recognized with the NOBEL prize in 2014, numerous SRM methods have evolved, with their respective advantages and hindrances (Möckl et al., 2014).

In general, SRM methods can be categorized as deterministic and stochastic approaches. Among the deterministic approaches is stimulated emission depletion (STED), which depletes fluorophores with stimulated emission around an excitation beam in a confocal laser scanning microscopy (Hell et al., 1994). Another approach is structured illumination microscopy (SIM), in which the sample is illuminated with a periodic pattern (Gustafsson, 2000). From images recorded by translating and rotating the pattern on the sample, a super-resolved image is reconstructed.
Stochastic super-resolution approaches are based on the idea of exciting fluorophores to emit light at separate time and spatial coordinates. This enables the discrimination of dense emitters over multiple camera frames, resulting in a super-resolved image. Super-resolution optical fluctuation imaging (SOFI) utilizes the analysis of fluorescent intensity fluctuations and their temporal correlations to achieve higher resolution (Dertinger et al., 2009). Single-molecule localization microscopy (SMLM) detects the positions of sparse emitters across many camera images to generate a super-resolved image (Heilemann, 2010;

Liu et al., 2022a). Exemplary methods are photoactivated localization microscopy (PALM) (Betzig et al., 2006), stochastic optical reconstruction microscopy (STORM) (Rust et al., 2006), direct STORM (*d*STORM) (Heilemann et al., 2008), and DNA-points accumulation for imaging in nanoscale topography (DNA-PAINT) (Jungmann et al., 2010).

Recent advances achieve resolution even below a nanometer. Maximally informative luminescence excitation (MINFLUX) combines the idea of single-molecule localization and a donut-shaped illumination scheme to improve resolution (Balzarotti et al., 2016). Resolution enhancement by sequential imaging (RESI), which is a concept based on Exchange-PAINT (Jungmann et al., 2014), repeatedly images the same target with multiple orthogonal labeling sequences and achieves resolutions in the Ångström scale (Reinhardt et al., 2023). The remainder of this chapter focuses on STED microscopy and SMLM techniques used for this work.

4.5.1 Stimulated emission depletion microscopy

Stimulated emission depletion (STED) microscopy is a confocal technology that operates on a dual-beam principle, where a STED beam deactivates fluorophores in a donut-shaped ring that is overlayed on an GAUSSIAN-shaped excitation beam (Hell et al., 1994) (fig. 8). The excitation beam elevates fluorophores to higher-energy states, while the STED beam depletes fluorophores through stimulated emission. Only fluorophores close to the center of the excitation beam, where the STED beam has zero intensity, remain in the excited state and fluoresce. This creates a small effective PSF, allowing features below the diffraction limit to be visualized. The coaligned beams are scanned across the sample to generate a super-resolved image.

The resolution Δr of a STED image can be approximated with eq. 12, where λ is the wavelength of the excitation laser, NA the numerical aperture of the objective, I_{STED} the intensity of the STED beam and I_S the saturation intensity (Hell, 2007). The saturation intensity is the intensity of the STED beam where 50% of the molecules are depleted (Harke et al., 2013). This depends mainly on the cross-section for stimulated emission at a specific wavelength and is fluorophore-characteristic. To improve the resolution, the intensity of the STED beam can be increased, or the selection of dye and STED beam wavelength optimized. Theoretically, the attainable spatial resolution is unlimited (Hell, 2007; Tortarolo et al., 2018).

$$\Delta r = \frac{\lambda}{2NA\sqrt{1 + I_{STED}/I_S}} \tag{12}$$

In practice, the resolution is limited by the signal-to-noise ratio, meaning the highest frequency of the sample that emerges from the noise dictates the resolution (Tortarolo et al.,

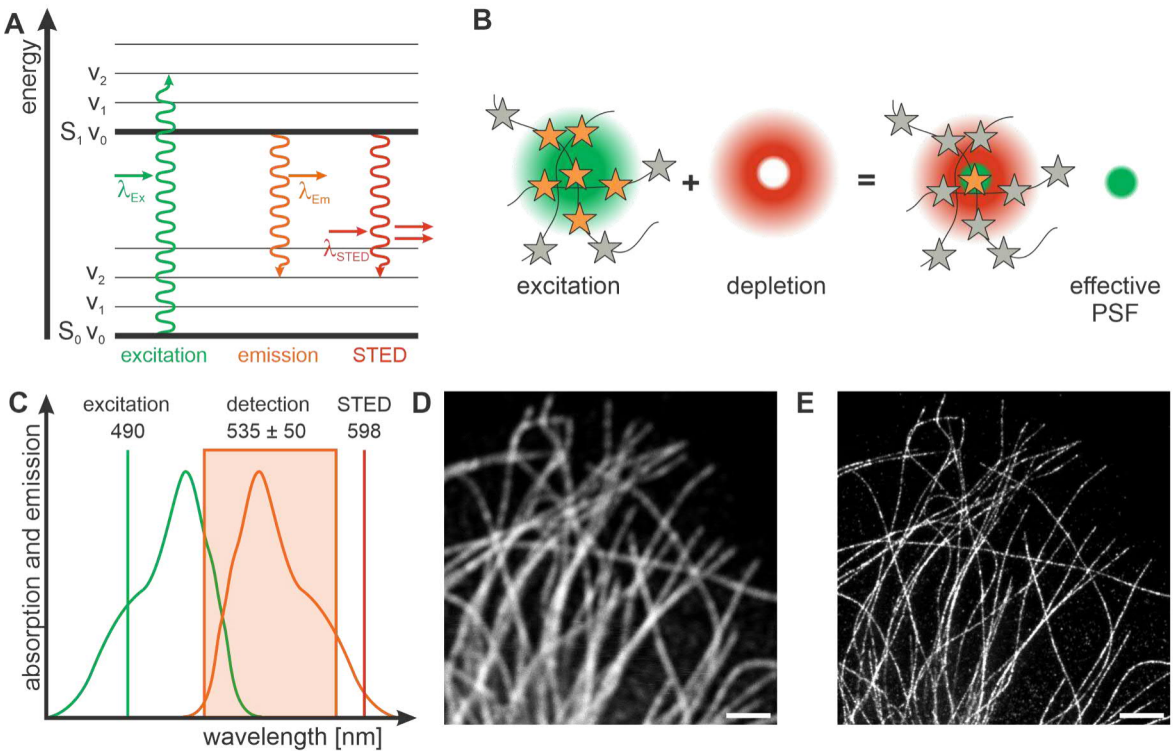

Figure 8: Stimulated emission depletion microscopy concept. **A** A fluorophore is excited to the first excited state S_1. By relaxation into the ground state S_0, fluorescence light is emitted. Alternatively, the STED beam depopulates the excited state by inducing stimulated emission. **B** Fluorophores are excited within the diffraction-limited excitation beam and their fluorescence is suppressed at the periphery of the excitation beam with the STED beam, resulting in a narrow effective PSF. **C** Exemplary excitation and emission spectrum of a fluorophore, wavelengths of the excitation and STED lasers and the detection window. **D** Confocal and **E** STED image of α-tubulin, scale bars 2 µm.

2018). Therefore, a sufficient number of photons must be recorded while scanning the sample. Additionally, it is crucial to scan the sample in sufficiently small steps in accordance with the NYQUIST-SHANNON sampling criterion (Vicidomini et al., 2018). Furthermore, the excitation and depletion beams must overlap perfectly in space, otherwise the imaging quality will degrade. The STED beam itself might cause unwanted effects that corrupt the quality of the STED image. The wavelength must be chosen to overlap well with the emission spectrum of the fluorophore for optimal depletion. However, if the STED beam is overlapping with the excitation spectrum of the fluorophore, this might lead to an increased probability of exciting the fluorophores with the STED beam. Additionally, the STED beam can excite fluorophores to higher electronic states that shows increased rates of photobleaching (Hotta et al., 2010). It is essential to maximize the depletion of the fluorophore with the STED beam and minimize photobleaching to obtain a good SNR. Examining the interplay between these parameters is essential to optimize STED microscopy.

To deactivate fluorophores with the STED beam, stimulated emission must win the competition of spontaneous emission of the excited flurophore, which typically happens within a few nanoseconds after excitation (Vicidomini et al., 2018). This short temporal window requires a high flux of stimulating photons and therefore, high STED laser powers were required in the earlier days with multiple MW/cm^2 up to GW/cm^2 (Heilemann, 2010; Vicidomini et al., 2018). The development of more stable fluorophores and an improved understanding of the photophysical processes lifted the constraints of STED and made it compatible with live-cell and long-term measurements. Furthermore, the experimental complexity of the setup has been reduced by the use of continuous instead of pulsed lasers (Willig et al., 2007). STED has been successfully applied in living cells (Hein et al., 2008; Liu et al., 2022b) and in *in vivo* imaging (Berning et al., 2012; Steffens et al., 2020). The extension to 3D-STED is possible by using two superimposed, incoherent STED beams that confine the fluorescence both laterally and axially (Klar et al., 2000). Exchangeable labels have been combined with STED that extend the acquisition time and allow multi-color imaging based on the concept of Exchange-PAINT (Glogger et al., 2022; Spahn et al., 2019) (see chapter 4.5.2).

4.5.2 Single-molecule localization microscopy

In SMLM the emitter density is actively controlled utilizing concepts from photochemistry or transiently binding labels (fig. 9 A). The control allows spatiotemporal separation of the PSFs that would otherwise overlap and would not be distinguishable due to the diffraction limit. The positions of the emitters can be determined from the separated PSFs, typically using a GAUSSIAN fit to approximate its shape (fig. 9 B). This yields localizations with an accuracy of a few nanometers (Sauer et al., 2017). By accumulating images of separated PSFs, a super-resolution image can be reconstructed consisting of thousands to millions of individual localizations (fig. 9 C).

The localization uncertainty depends on the number of photons N collected, the background level b, the pixel size of the detector a, and the standard deviation of the GAUSSIAN fit σ. Depending on the regression method, the localization uncertainty is described with eq. 13 and 14, where τ is a normalized dimensionless background parameter, for least squares (LS) or with eq. 14 and 15 (Mortensen et al., 2010; Thompson et al., 2002) for maximum likelihood estimation (MLE) (Mortensen et al., 2010; Stallinga et al., 2012).

$$\delta^2_{loc}|_{LS} = \frac{\sigma^2 + a^2/12}{N}\left(\frac{16}{9} + 4\tau\right) \tag{13}$$

$$\tau = \frac{2\pi b^2(\sigma^2 + a^2/12)}{Na^2} \tag{14}$$

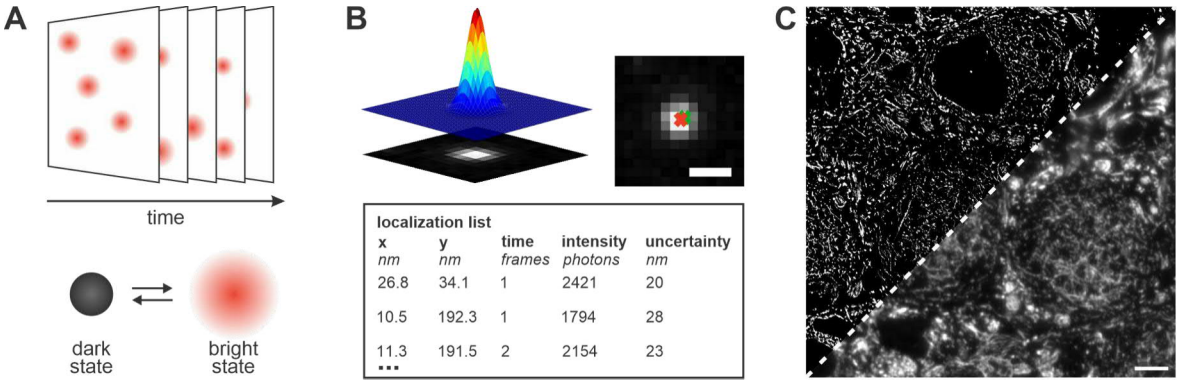

Figure 9: Principle of single-molecule localization microscopy. A The PSFs of isolated emitters are recorded over time. Spatiotemporal separation of the emitters can be provided by reversibly switching the fluorophores between a dark and bright state. **B** The PSFs are fitted with a GAUSSIAN function. The orange cross visualizes the determined position of the emitter by the fit and the green cross depicts the true position of the emitter. A localization list is curated that stores information of all emitters positions, time, intensity, localization uncertainty, etc, scale bar 500 nm. **C** Widefield (bottom right) and super-resolution (top left) image recorded with the DNA-PAINT principle. The image shows alpha-tubulin in a tissue section of the medial nucleus of the trapezoid body (MNTB) of a mouse brain (Narayanasamy et al., 2022). The super-resolution image was rendered based on the localization list. Scale bar 5 µm.

$$\delta^2_{loc}|_{MLE} = \frac{\sigma^2 + a^2/12}{N}\left(1 + 4\tau + \sqrt{\frac{2\tau}{1+4\tau}}\right) \tag{15}$$

Furthermore, the localization uncertainty can be experimentally determined via nearest neighbor analysis (NeNA) (Endesfelder et al., 2014), based on the distance distribution of nearest neighbor positions in adjacent frames. The distribution is fitted with eq. 16 to obtain the experimental localization uncertainty. The first term describes the distances d of localizations from the same molecule and the other terms are correction factors for distances arising from different molecules (false nearest neighbors).

$$p(d) = A_1 \cdot \left(\frac{d}{2\sigma^2} \cdot e^{\frac{-d^2}{4\sigma^2}}\right) + A_2 \cdot \left(\frac{1}{\sqrt{2\pi\omega^2}} \cdot e^{\frac{-(d-d_c)^2}{2\omega^2}}\right) + A_3 \cdot d \tag{16}$$

Prominent examples of SMLM techniques grounded on the concept of photoswitching are STORM (Rust et al., 2006), *d*STORM (Heilemann et al., 2008), and photoactivated localization microscopy (PALM) (Betzig et al., 2006).

In STORM, fluorophores undergo transitions into long-lived dark states, ensuring that only a subset of fluorophores contributes to signal emission (Heilemann et al., 2005; Rust et al., 2006). This is facilitated with buffers containing reducing reagents and an oxygen scavenger system. The *d*STORM approach has simplified this concept by demonstrating photoswitching with a singular dye, eliminating the necessity for an additional activating dye (Heilemann et al., 2008).

PALM utilizes photoactivatable and photoconvertible fluorescent proteins (FPs), which can be stochastically turned on (activated) and off (deactivated) with light (Betzig et al., 2006). Fluorescent proteins can be genetically encoded, allowing them to be fused to specific proteins of interest, which renders PALM to be minimally invasive. This also enables stoichiometric labeling and allows for the quantification of molecules (Fricke et al., 2015).
Alternative to photoswitching, blinking (the appearance and disappearance of emitters) can be generated through transient binding of freely diffusing fluorophores to the target protein. This concept (Sharonov et al., 2006) was combined with oligonucleotides (DNA-PAINT) (Jungmann et al., 2010) and is explained in more detail in the following chapter. Another approach is HaloTag-PAINT (Kompa et al., 2023), where the HaloTag ligand mentioned in chapter 4.2.2 is chemically modified for transient binding to the target. Further analysis of SMLM data is highlighted in chapter 4.5.2.2.

4.5.2.1 DNA-PAINT

Points accumulation for imaging in nanoscale topography (PAINT) utilizes fluorescent probes that repetitively bind and unbind to a specimen (Sharonov et al., 2006). The initial PAINT concept was demonstrated using the lipophilic dye Nile Red that binds to lipid membranes. The PAINT concept was extended by combining it with short, fluorescently labeled oligonucleotides and termed DNA-PAINT (Jungmann et al., 2010).
In DNA-PAINT, single-stranded fluorophore-labeled oligonucleotides (imager strands) of 7-10 nucleotide length transiently bind to their complementary labeled targets (docking strands), effectively splitting the fluorescence signal both spatially and temporally (fig. 10). The binding kinetics are highly tunable by adjusting the DNA sequence (Schueder et al., 2019; Strauss et al., 2020), strand length, concentration, temperature, and buffer conditions (Civitci et al., 2020; Schueder et al., 2019).
Ideally, the number of localization events should be dense but spatially separated to allow single-molecule detection (Schueder et al., 2019). At the same time, the duration of the binding events should match the speed of image acquisition to effectively visualize the complete structure of interest in a short measurement period. Simply increasing the concentration of imager strands and therefore the number of binding events would lead to an increase in background fluorescence and hamper the imaging quality. This is why the number of binding events was optimized through the strand sequence. A combination of A and G or T and C bases within a DNA-strand prevent self-interaction. Multiple overlapping binding sites in the strand sequence increase the rate of binding events (Strauss et al., 2020). Higher binding frequencies are also achieved with the addition of $MgCl_2$ that stabilizes the DNA duplex formed between imaging and docking strand (Schueder et al., 2019). Addition of ethylene carbonate accelerates the dehybridization of the DNA duplex and results in shorter binding times (Civitci et al., 2020).

Compared to other SMLM techniques, DNA-PAINT omits the principle of photoswitching. It allows for continuous imaging without photobleaching, as fluorophores are constantly exchanged. No reducing buffers and near-UV irradiation are required as in STORM and PALM experiments, which makes this method very suitable for biological samples. In combination with TIRF illumination, which reduces background signal, this method is suitable for studying cell membrane receptor distributions and dynamics (Harwardt et al., 2020; Niederauer et al., 2023; Schröder et al., 2021; Stehr et al., 2021).

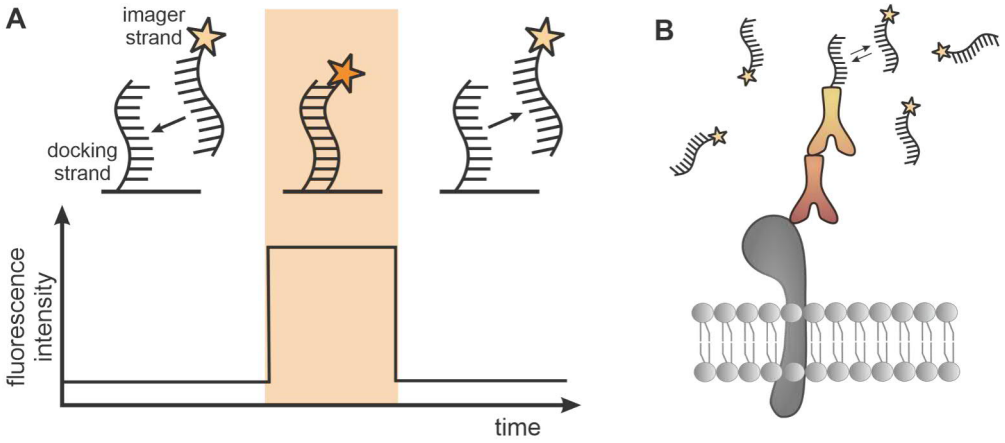

Figure 10: DNA-PAINT concept. **A** Single-stranded oligonucleotides labeled with fluorescent markers (imager strands) are present in solution and are complementary to the docking strand sequence. When the imager strand and docking strand undergo reversible hybridization, a fluorescence signal becomes observable. Subsequent dissociation of the imager strands results in a reduction in fluorescence intensity. **B** The docking strand is attached to a probe, for example an antibody, that binds to the biomolecule of interest. Transient binding of the imager to the docking strand reveals the position of the biomolecule.

Exchange DNA-PAINT extends the DNA-PAINT concept and enables multi-target imaging. Different targets within the sample are imaged sequentially (Jungmann et al., 2014). In each imaging round, a specific imager strand is introduced, binding temporarily to its complementary docking strand (fig. 11). After capturing data for the first target, the sample is washed to remove the initial imager strand, and a new imager strand, specific to the next target, is added. This process is repeated for multiple targets, resulting in individual super-resolution images for each target, which are then combined to create a multi-target super-resolution image.

This approach offers numerous advantages, including the ability to use the same fluorophore for all targets, eliminating chromatic aberrations and channel crosstalk. Moreover, the number of targets is not restricted by the capability of separating fluorescence signal based on the wavelength, as is the case with PALM or STORM. The number of targets is restricted by the number of orthogonal strand pairs. By disentangling the binding kinetics of strand pairs, multiple targets can be imaged at once, as demonstrated in Wade et al., 2019, where the authors recorded a 124-plex super-resolution image.

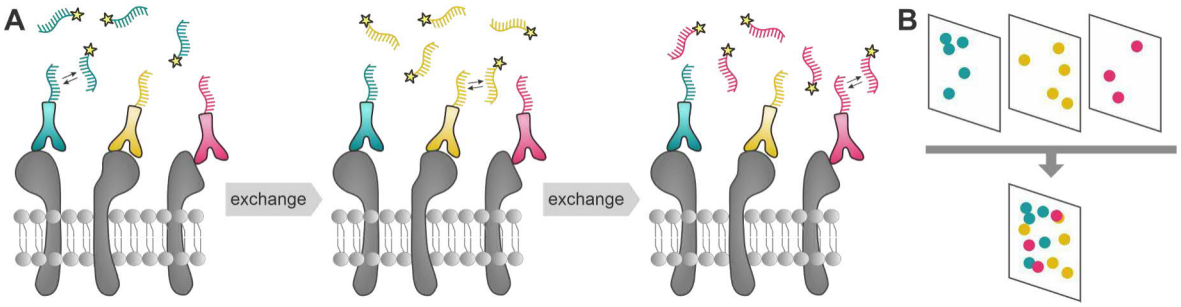

Figure 11: Exchange DNA-PAINT concept. A The scheme exemplary depicts three membrane receptors, which are labeled with primary antibodies that carry different docking strand sequences. In the first round of imaging, a buffer is added that contains matching imager strand to the docking strand of the first receptor and a SMLM-movie is recorded. The imaging buffer is exchanged to an imaging buffer containing the imager strand matching the docking strand of the second receptor. This procedure can be repeated several times to image multiple targets in the same sample with the same dye. **B** Super-resolved images of the individual targets are merged into one multiplex image.

Exchange DNA-PAINT has been demonstrated to visualize different organelles (Agasti et al., 2017), membrane receptors (Schröder et al., 2021), and synaptic proteins (Narayanasamy et al., 2021). The concept was combined with nanobodies to reduce the displacement error between target and fluorescent dye (Agasti et al., 2017; Sograte-Idrissi et al., 2019).

4.5.2.2 Analysis of SMLM data

Protein interactions often lead to the formation of heterogeneous and dynamic multimolecular protein assemblies. Investigating the arrangement and assembly of these protein clusters is vital for unraveling their roles within cellular processes. To gain insights into the spatial distribution of molecules, specialized distribution and clustering algorithms come into play, which will be elucidated in the following sections.

SMLM data are point clouds in 2D or 3D space in which fluorophore positions are stored as a point list $P = P_1, P_2, ...P_N$, where N is the total number of recorded fluorophore positions (Khater et al., 2020). Each event stores information about the coordinates (x_i and y_i for 2D and x_i, y_i, and z_i for 3D) along with other information about the time of detection, localization uncertainty, fluorescence intensity, background, etc. Some fluorophores might be detected multiple times due to the stochastic blinking. In order to group these repeated localizations, cluster analysis algorithms aim to use mathematical descriptions to identify patterns, for example a dense group of molecule localizations. These localizations are pooled into clusters representing individual molecules or protein complexes.
Popular cluster analysis methods for SMLM data are VORONOI-Tesselation (Andronov et al., 2016; Levet et al., 2015; Okabe et al., 2000), and density-based algorithms such as density-based spatial clustering of applications with noise (DBSCAN) (Ester et al., 1996; Harwardt et al., 2020; Mollazade et al., 2017; Zhang et al., 2017). VORONOI diagrams par-

tition the space between localizations and create VORONOI cells, where each cell contains one localization. This allows to describe clustering based on the VORONOI cell volume and shape and highlights clustering at different scales. Furthermore, cells can be combined to segment regions of interest. DBSCAN distinguishes between core, boundary, and isolated points (fig. 12). Core and boundary points are grouped together, while isolated points do not belong to a cluster. This method is particularly effective in identifying clusters of arbitrary shapes and excludes noise from clustering. The determination of critical cluster parameters, namely the minimum neighbor count and the radius, can be achieved through established methodologies (Verzelli et al., 2022).

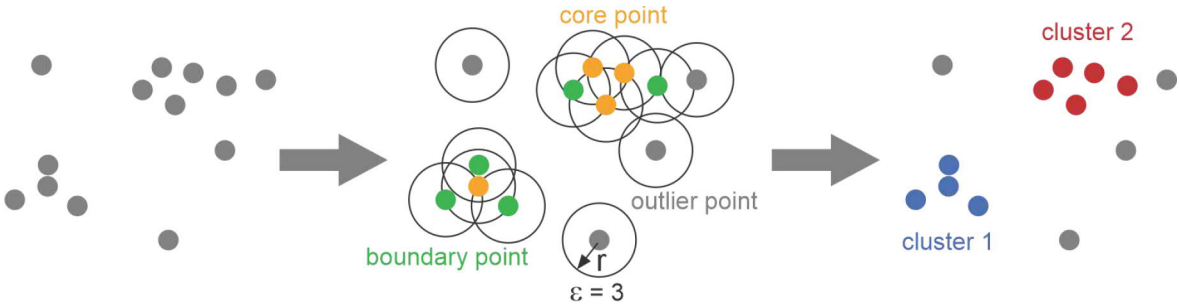

Figure 12: Clustering with DBSCAN. A defined radius r and a minimum number of points ε is used to determine core, boundary and outlier points. Core points have $\geq \varepsilon$ neighbors within their radius. Boundary points are within radius distance to a core point but have $\leq \varepsilon$ neighbors. All core and boundary points build a cluster. The cluster can only grow from core points.

The distribution of localizations and protein clusters can be further analyzed. RIPLEY's functions are descriptive statistics to compare the spatial distribution of a point pattern to a random distribution (Lopes et al., 2017; Pereira et al., 2012; Ripley, 1977). This set of functions examines if a dataset is clustered or dispersed.

The colocalization of two proteins can be characterized with coordinate-based colocalization (CBC), that describes the degree of colocalization between the proteins within a defined radius based on SPEARMAN correlation (Malkusch et al., 2012). The concept was extended to work with a single color (Díez et al., 2014) and combined with DBSCAN to distinguish colocalized and non-colocalized clusters (Pageon et al., 2016). Multicolor VORONOI tessellation allows the study of co-localization based on the ratio of superposed cells in both channels (Andronov et al., 2016).

Nearest neighbor distances describe the distribution of distances between localizations or protein clusters directly (Chapman et al., 2023; Chou et al., 2013; Fischer et al., 2021; Knuth, 1968). For each localization or protein cluster, the nearest neighbor is found and the distance determined. The distances from all molecules form a distance distribution. This concept can be extended to k-nearest neighbors to reveal clustering at different scales (fig. 13) (Barth et al., 2020).

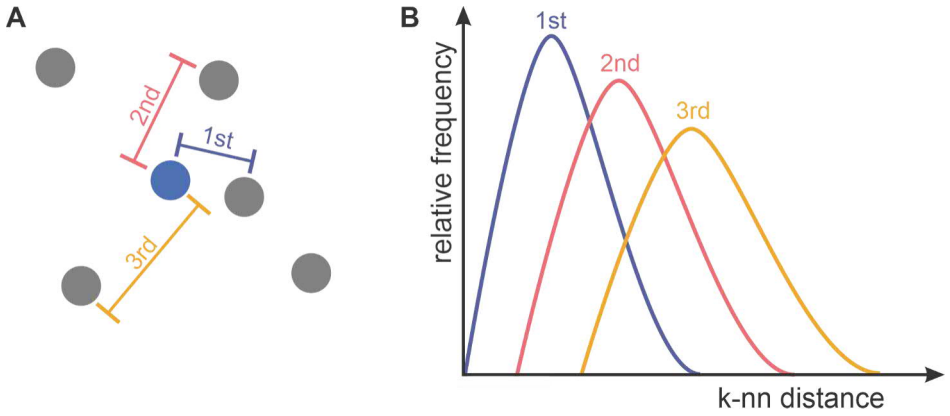

Figure 13: k-nearest neighbor distances. **A** The distances for the first, second, and third nearest neighbors are determined for each localization. **B** The distances from all molecules form a distance frequency distribution.

4.6 Single-particle tracking

Single-particle tracking (SPT) enables the analysis of biomolecule mobilities in living cells with nanometer spatial and millisecond temporal resolution (Dietz et al., 2019a; Stone et al., 2017). It is based on the concept of single-molecule localization microscopy and captures the positions of a target over time. Targets to study are membrane receptors, as they often build the starting point for cell signaling cascades by forming oligomers and interacting with intracellular signaling proteins and their cell environment (Harwardt et al., 2017). This may lead to changes in diffusion behavior that are measurable by SPT and can be linked to biological interactions (Harwardt et al., 2017; Manzo et al., 2015; Shen et al., 2017). SPT resolves the heterogeneous movement of individual fluorophore-labeled proteins over time and space that would otherwise remain hidden in ensemble measurements.
In the following chapter, diffusion with a focus on membrane receptors is explained (chapter 4.6.1). Labeling strategies specifically designed for SPT are highlighted in chapter 4.6.2. The concept of localization and tracking is introduced in chapter 4.6.3. The last chapter focuses on different SPT data analysis strategies, which were used in this work (chapter 4.6.4).

4.6.1 Diffusion

Diffusion refers to the movement of particles along a gradient of decreasing concentration. FICK's law of diffusion describes the diffusion coefficient D as a proportional constant between the particle flux J and the concentration gradient $\frac{\partial c}{\partial x}$ with respect to a length x (eq. 17) (Fick, 1855). The particle flux signifies the quantity of particles passing through a specified area over a given time.

$$J = -D\frac{\partial c}{\partial x} \tag{17}$$

The diffusion coefficient was further defined by EINSTEIN (eq. 18), with k_B being the BOLTZMANN constant, T the temperature, η the viscosity of the medium, and R_H the hydrodynamic radius of the diffusing particle — a concept that highlights that the diffusion of a particle in solution depends on its size including the solvate shell surrounding the particle (Einstein, 1905).

$$D = \frac{k_B T}{6\pi\eta R_H} \tag{18}$$

The cell membrane contains a multitude of biomolecules, including lipids and proteins, that build a complex and dynamic system of interaction (Singer et al., 1972). Biomolecules are heterogeneously distributed on the membrane, yielding local variations of viscosity and therefore diffusion. The diffusion of membrane receptors may be influenced by the lipid composition of the membrane, interactions with other receptors, signaling molecules or organelles such as the actin cytoskeleton or membrane nanodomains (Harwardt et al., 2017; Rossier et al., 2012). Receptor uptake in the cell is typically prefaced by immobilization. This leads to a heterogeneous spectrum of receptor velocities and diffusion patterns. The cell membrane can be approximated as a two-dimensional surface and descriptors of movement such as the diffusion coefficient per detected biomolecule can be computed, for example with the Saffman-Delbrück model (Saffman et al., 1975).

4.6.2 Labeling and imaging strategies for single-particle tracking

SPT requires low labeling density to enable single-molecule detection. For high endogeneous protein densities, this can be achieved by sub-stoichiometric labeling strategies, such as introducing a photoactivatable fluorophore (Manley et al., 2008) or by using transiently binding labels (Giannone et al., 2010; Stehr et al., 2021). Universal PAINT (uPAINT) utilizes probes that weakly interact with membrane molecules and label a constant fraction of them by transient binding of the target protein (Giannone et al., 2010) (fig. 14). The probe carries an organic fluorophore that allows localization of the position of the membrane molecules. The labeling density is influenced by the affinity of the probe to the target and the concentration of the probe. As the labels cannot penetrate the cell membrane, the method typically targets the extracellular domain of membrane proteins. For single-molecule imaging, TIRF or HILO illumination is used to reduce the background signal from unbound markers in solution. DNA-PAINT has also been utilized as a labeling strategy for SPT, in which a fluorescently marked oligonucleotide reversibly binds to a single-stranded DNA attached to a target molecule (Stehr et al., 2021).

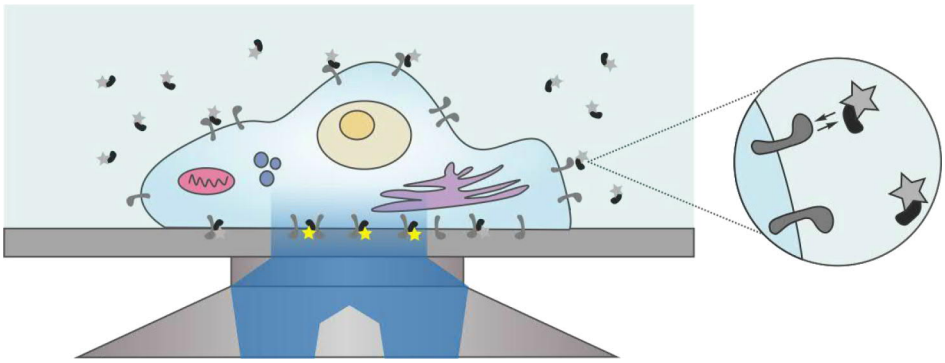

Figure 14: uPAINT concept. Probes that carry a fluorescent dye are added to the imaging solution. The probes have an affinity to transiently bind to the target probe, for example a membrane receptor, and build a complex. A fraction of the membrane receptors is labeled based on binding and unbinding kinetics. The surface of the cell is illuminated with the evanescent field from TIRF microscopy. Only receptors on the cell surface that carry a probe with a dye are detected.

4.6.3 Localization and tracking

Each molecule carrying a fluorescent label creates a diffraction-limited spot in the imaging plane that can be localized. The localizations are linked over time to form single-particle trajectories (fig. 15). Events of particle merging, splitting, death or birth are possible (fig. 16) (Jaqaman et al., 2008; Racine et al., 2006). The events are treated as a global combinatorial optimization problem, with the goal of finding the overall most likely set of trajectories across the movie. Each connection is evaluated with a cost function that determines the likelihood of a connection based on, for example, the distance and intensity of the localizations. Starting with a coarse solution that is fine-tuned (Jaqaman et al., 2008), with for example a simulated annealing approach (Racine et al., 2006), an optimum can be found.

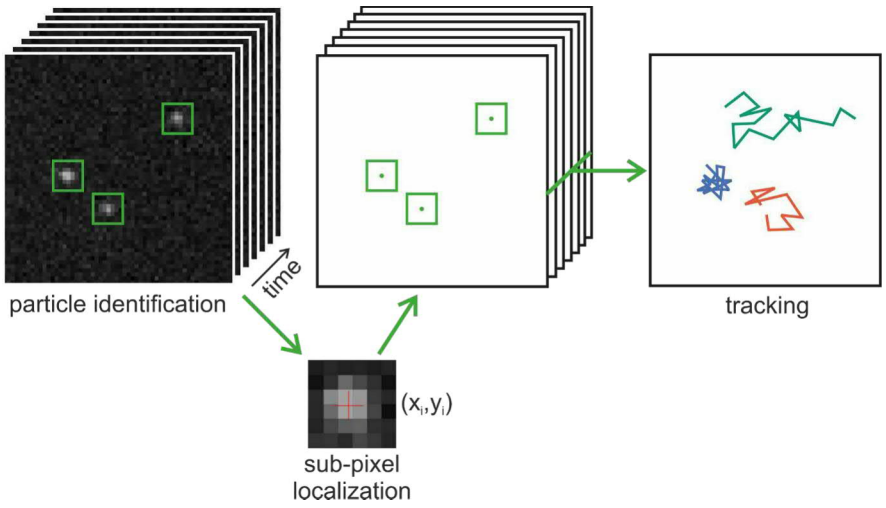

Figure 15: Principle of trajectory computation. Single-molecule signal is localized with sub-pixel accuracy and joined over time to trajectories. Each trajectory has its own properties like a diffusion state.

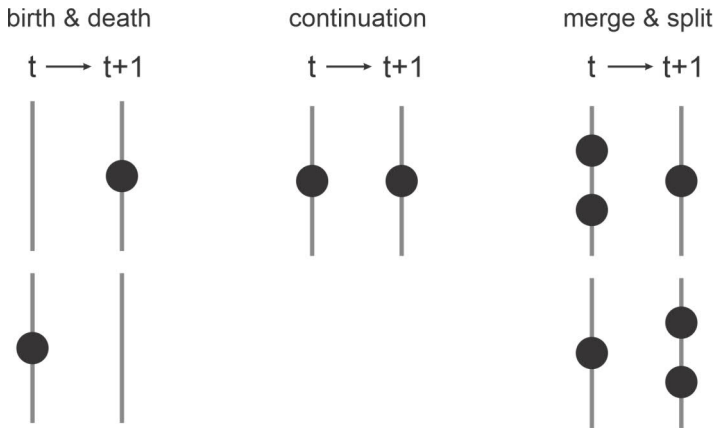

Figure 16: Events occurring in single-particle tracking. Three different event types are possible between two consecutive time steps t and t+1.

4.6.4 Single-particle tracking data analysis

Various models have been developed to describe the motion characteristics of single-particle trajectories. Approaches include the mean squared displacement (MSD), the mean jump distance (MJD), and hidden MARKOV modeling (HMM).

4.6.4.1 Mean squared displacement analysis

The MSD of a trajectory is calculated across different time steps Δt (eq. 19, fig. 17 A), with x and y being the spatial coordinates at the respective time points (Saxton et al., 1997; Sibarita, 2014). By computing the average displacement per time step, a MSD curve is generated, which describes the MSDs against their respective time steps.

$$MSD(\Delta t) = \langle (x_{t+\Delta t} - x_t)^2 + (y_{t+\Delta t} - y_t)^2 \rangle \tag{19}$$

Different motion states can be extracted from the shape of the MSD curve (fig. 17 B). For each time interval, the MSD is computed and plotted against the time interval. One way of motion state characterization is sorting the trajectories into motion models based on their α values derived by eq. 20 (Saxton et al., 1997), which influences the curvature of the MSD plot. The diffusion coefficient D can be calculated with eq. 21 with $dof = 4$ for 2D and $dof = 6$ for 3D BROWNIAN motion.

$$MSD(\Delta t) = 4D\Delta t^{\alpha} \tag{20}$$

$$D = \frac{MSD(\Delta t)}{dof \Delta t} \tag{21}$$

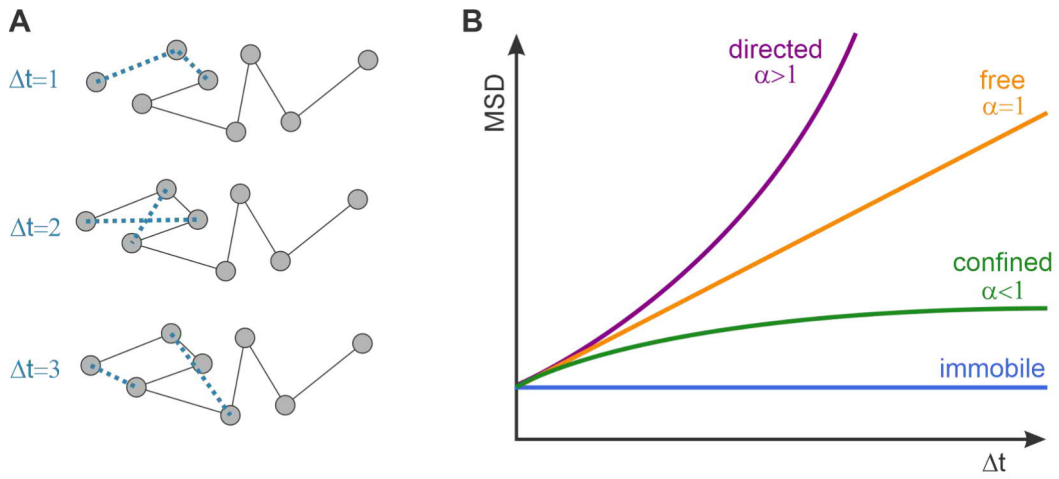

Figure 17: MSD calculation and curves of different diffusion states. **A** MSD values are calculated for different time steps. **B** Different diffusion states show characteristic MSD curves. Freely diffusion particles show a linear dependency, directed motion a rising curve, confined motion a saturated curve, and immobile trajectories a slope equal to zero.

4.6.4.2 Hidden Markov modeling

Receptor dynamics are complex and switching between diffusion states within a single receptor trajectory is possible. Hidden MARKOV modeling allows the analysis of transitions between diffusion states (Chung et al., 2010; Liu et al., 2019; Low-Nam et al., 2011; Ott et al., 2013; Persson et al., 2013; Sungkaworn et al., 2017).

A hidden MARKOV model is generally defined by a number n of hidden states $S = \{S_1, S_2, ... S_n\}$, which are obscured by noise (Yoon, 2009). The model takes a sequence of observations $x = x_1, x_2, ... x_l,\ x_i \in O$, drawn from an observation alphabet $O = \{O_1, O_2, ... O_m\}$, and links them to a sequence of hidden states $y = y_1, y_2, ... y_l,\ y_i \in S$. Each hidden state has an emission probability distribution e_p that describes the probability of an observed value being generated from a hidden state, $p\{x_t = x|y_t = i\} = e_p(x|i)$. Some hidden states might be more likely than others and their probabilities of occurrence is described by the equilibrium matrix π. For example, π for a two-state model equals to $[\pi_A, \pi_B]$ and $\sum_{i=1}^{n} \pi_i = 1$. The transition matrix tp of shape n x n describes the probability of staying or switching between hidden states for an adjacent time step (eq. 22, example for $n = 2$ states). The transition probability is constant over time $p\{y_{t_1+1} = j|y_{t_1} = i\} = p\{y_{t_2+1} = j|y_{t_2} = i\}\ i, j \in S$ and the row entries sum up to 1, $tp_{A\to B} + tp_{A\to A} = 1$. Two simplifying assumptions are made in a hidden MARKOV model. First, the future value of the hidden state depends only on the value of the present state, $p\{y_{t_1+1}|y_{t_1}\}$. Second, the value of the currently observed state depends only on the value of the present hidden state, $p\{x_t = x|y_t = i\} = e_p(x|i)\ i \in S$.

$$\begin{array}{c} \\ A_1 \\ B_1 \end{array} \begin{array}{c} \begin{array}{cc} A_2 & B_2 \end{array} \\ \begin{pmatrix} tp_{A\to A} & tp_{A\to B} \\ tp_{B\to A} & tp_{B\to B} \end{pmatrix} \end{array} \tag{22}$$

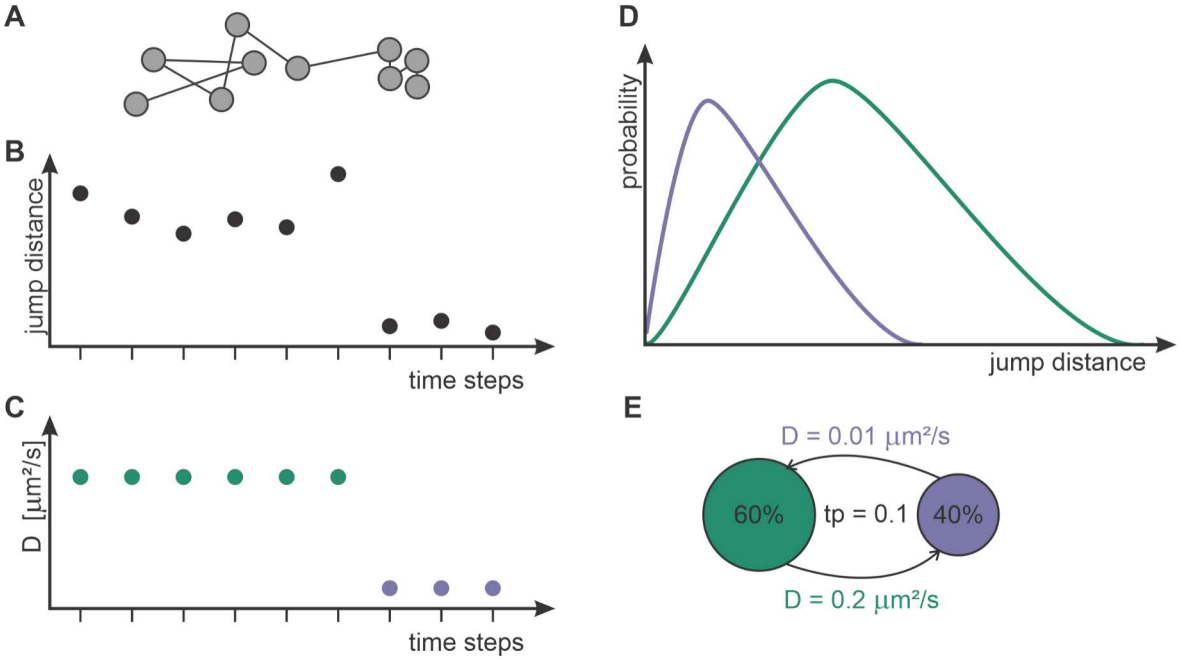

Figure 18: Hidden Markov modeling of single-particle tracking data. **A** Exemplary trajectory with dots marking localizations and lines marking the jump distances between them. **B** Jump distances of the trajectory. **C** A diffusion state is assigned to each jump distance. **D** Emission probabilities of the two diffusion states. **E** State-transition-diagram of an exemplary two state model (states are highlighted with color) with distinct diffusion coefficients (0.2 µm^2 /s and 0.01 µm^2 /s) and populations (60% and 40%). The transition probability tp to switch between the diffusion states is 0.1 for both directions.

In the context of single-particle tracking, a trajectory consists of a sequence of observations, which are the jump distances based on the displacement between two consecutive recorded time points (Ott et al., 2013) (fig. 18 A, B). The hidden states are discrete diffusion coefficients, which are assigned to the jump distances of the trajectory (fig. 18 C). The probability that a jump distance r belongs to a hidden state is described by the emission probability e_p (fig. 18 D), eq. 23). The hidden states are connected with transition probabilities tp, describing how likely it is that diffusion state switching occurs within a trajectory. The optimal number of states, their diffusion coefficients, occupation, and transition probabilities are found (fig. 18 E).

$$e_p(r, t|D) = \frac{2r}{4Dt} \cdot e^{-\frac{r^2}{4Dt}} \tag{23}$$

4.7 Quantitative phase microscopy

This section provides an overview of quantitative phase imaging (QPI), a technique that enables the measurement of optical path length delays and refractive index (RI) changes in transparent samples. The concept of phase and qualitative as well as quantitative phase imaging methods are presented. A comparison to fluorescence microscopy is drawn and some current applications are highlighted.

A light wave is composed of an amplitude A and a phase component θ, which form a complex amplitude U (eq. 24) (Zuo et al., 2020). The squared amplitude is called the intensity and describes the energy of light (eq. 25). The phase describes in which part of an oscillation cycle the light wave is. While human eyes or imaging sensors can detect the intensity, they cannot directly access the phase information.

$$U(x,y) = A(x,y) \cdot e^{j\theta(x,y)} \tag{24}$$

$$I(x,y) = A(x,y)^2 \tag{25}$$

In light microscopy, optically thin samples such as cells absorb and scatter light weakly and therefore appear transparent in the FOV (Popescu, 2011; Zuo et al., 2020). However, they induce phase changes due to different structural regions of nonuniform optical density (fig. 19 A). A phase shift is caused by light passing through material with a different refractive index than water and can be expressed as

$$\theta = \frac{2\pi}{\lambda} \int_{z=0}^{h} n(z) dz \tag{26}$$

where θ is the phase shift of light contributed by all elements in the sample of varying refractive index n through sample height h in the z direction (Nguyen et al., 2022).
The phase can not be measured directly and must be mixed with another reference field to retrieve the inference pattern (Popescu, 2011) (fig. 19 B). Methods such as Zernike's phase contrast microscopy (Zernike, 1955) and differential inference contrast (Nomarski, 1955) convert phase changes into intensity changes that can be detected by a camera (Guo et al., 2020; Zuo et al., 2020). Here, the measured intensity modulations have a non-linear relationship to the phase of the sample and phase changes can only be reported qualitatively.

Quantitative phase imaging reports on refractive index changes of the sample (Guo et al., 2020; Park et al., 2018; Popescu, 2011). A way to achieve this is to use interferometry, where a sample beam that passes the structure of interest and a reference beam that bypasses the structure of interest from a laser source are interfered to access the quantita-

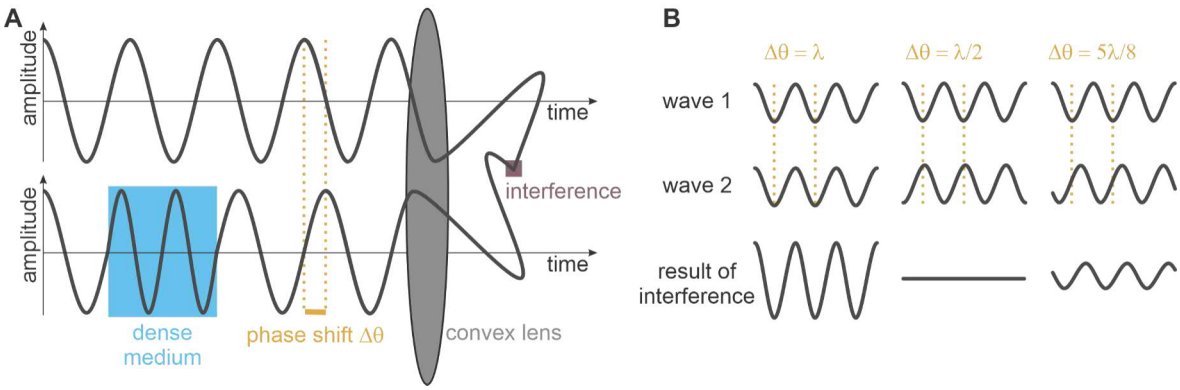

Figure 19: Phase shift and interference. **A** Light passes through a dense medium, which causes a shift in phase compared to an unperturbed reference beam. When the two beams collide, for example through a convex lens, they cause interference. **B** Interference patterns of different phase shifts. The amplitudes are added with a phase shift of zero, canceled out with a phase shift of $\lambda/2$, and diminished with a phase shift of $5\lambda/8$.

tive phase information (Nguyen et al., 2022). The integration into microscopy with the aid of digital methods for image processing is called digital holographic interference microscopy.

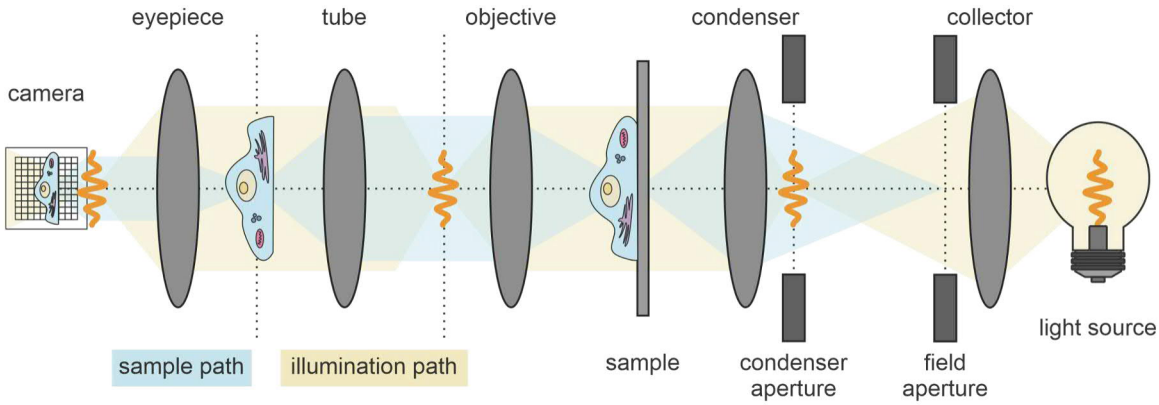

Figure 20: Köhler illumination. The light from a light source with uneven illumination is passed homogeneously through the sample by adjusting the alignment and opening of the condenser aperture, field aperture, and objective lens.

Soto *et al.* introduced a method, where only slight adjustments must be made to a commercial bright-field microscope to record quantitative phase images (Soto et al., 2017). The setup is operated in Köhler illumination (Köhler, 1893) (fig. 20), where the sample is evenly illuminated by adjusting the positions and openings of the condenser aperture, field aperture, and objective lens in the illumination path. Using this method, the 3-dimensional refractive index distribution of so called weak objects (objects with low refractive index variation, such as cells and micro-organisms) can be reconstructed with a stack of bright-field images. This method belong to partially coherent optical diffraction tomography (PC-ODT) as a standard light source, such as a LED or a halogen lamp of a bright-field microscope is sufficient. Assuming a central wavelength of $\lambda_m = 450$ nm and a high numerical aper-

ture of the objective lens ($NA_o = 1.4$) operating with immersion oil ($n_{oil} = 1.518$) and a condenser lenses with $NA_c = 0.95$, results in a lateral resolution of 200 nm and axial resolution of 500 nm (eq. 27 and 28).

$$\Delta x, y = \frac{\lambda_m}{NA_o + NA_c} \approx 200\ nm \tag{27}$$

$$\Delta z = \frac{\lambda_m}{RI_{oil} - (RI_{oil}^2 - NA_o^2)^{\frac{1}{2}}} \approx 500\ nm \tag{28}$$

QPI focuses on the physical properties of the sample such as the refractive index. The measured phase shift is directly proportional to the dry mass of the sample (Barer, 1952; Guo et al., 2020). Compared to fluorescence microscopy, phase shift measurements do not require exogenous labels and do not alter the sample behavior (Guo et al., 2020). They allow for long-term imaging, as photobleaching and -toxicity are not an issue. Fluorescence microscopy on the other hand offers molecular specificity and highlights the targets of interest with fluorescence light.

Some applications of QPI in the field of biomedicine are the determination of cancerous cell mass decrease by cytotoxic T lymphocyte attacks (Zangle et al., 2013), quantification of red blood cell membrane properties and potential influences on oxygen delivery (Park et al., 2010), and mass redistribution rates for multiple cell lines during epithelial-to-mesenchymal cell state transition (Nguyen et al., 2020). With its ability to capture quantitative and dynamic information, QPI continues to contribute to the understanding of biological systems and holds promise for further advancements in biomedical research.

4.8 Basic concept of neural networks

Neural networks are a fundamental concept in the field of machine learning, inspired by the structure and functioning of the human brain (Livingstone, 2010). In the following, the basis of neural networks is explained, starting with the multi-layer perceptron (chapter 4.8.1), one of the simplest network architectures. Convolutional neural networks are introduced, a design especially appropriate for spatial data such as images (chapter 4.8.2). Neural network training is explained in chapter 4.8.3, highlighting different loss functions (chapter 4.8.3.4) and optimization techniques (chapter 4.8.3.1).

4.8.1 Multi-layer perceptron

Neural networks are thought to be universal function approximators that learn a mapping from an input vector to a desired output vector, provided sufficient and suitable data is available and the architecture of the network is matching the task (Hornik et al., 1989). Among the simplest neural network architectures is the multi-layer perceptron (MLP), which utilizes perceptrons as fundamental building units, introduced by Rosenblatt, 1957. These units, often referred to as nodes or neurons, are interconnected with trainable weights (Livingstone, 2010). The MLP consists of an input layer, one or more hidden layers, and an output layer (fig. 21 A).

In each node, information from the previous layer is processed through a series of weighted connections, followed by a non-linear transformation (fig. 21 B). This is expressed in eq. 29, where the weighted inputs $w_i \cdot x_i$ are summed with an optional bias term b. A non-linear activation function $f(z)$ is applied to generate the final output of the node (eq. 30). The weights are adjusted or "learned" during the training process (see chapter 4.8.3 for more details).

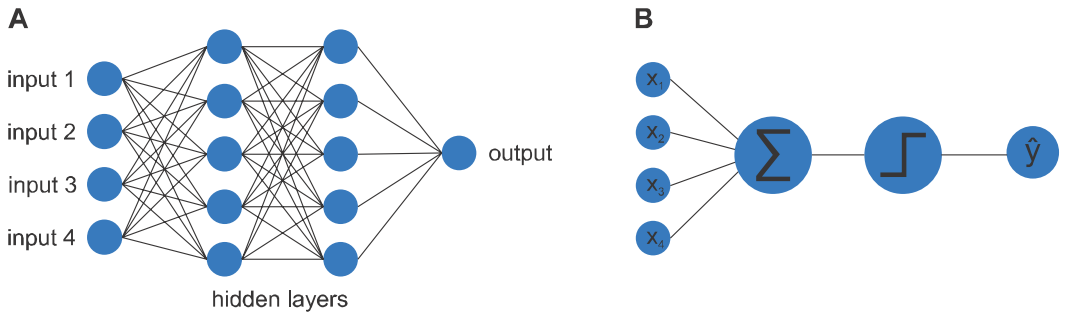

Figure 21: Architecture of a multi-layer perceptron. A The network contains an input layer, one or multiple hidden layers, and an output layer. Each layer consists of nodes, which are connected through weights, depicted as lines. **B** In an individual node, inputs are received from the previous layers, which are processed by calculating a sum of weighted inputs, adding a bias term and applying an activation function to produce the output of the node.

$$z = \sum_{i=1}^{n}(w_i \cdot x_i) + b \tag{29}$$

$$\hat{y} = f(z) = f(\sum_{i=1}^{n}(w_i \cdot x_i) + b) \tag{30}$$

Activation functions introduce non-linearity to the network, which is crucial for learning complex relationships within the data (Dubey et al., 2021). Common activation functions are the sigmoid function, the tanh function, and rectified linear units (ReLU) (Nair et al., 2010) (fig. 22 A). These functions play a pivotal role in enabling the approximation of complex and non-linear problems. A non-linear problem is a relationship between input and output variable that cannot be approximated with linear combinations of the input features. Conversely, for a linear problem, such linear combinations are adequate. For instance, in a linear classification task, a linear decision boundary can separate instances from different classes (fig. 22 B). In a non-linear classification task, a linear decision boundary would not be sufficient to separate instances from different classes (fig. 22 C). Omitting activation functions would render a linear model incapable of addressing non-linear complexities. This also translates to regression problems, where a continuous numerical output is predicted instead of classes.

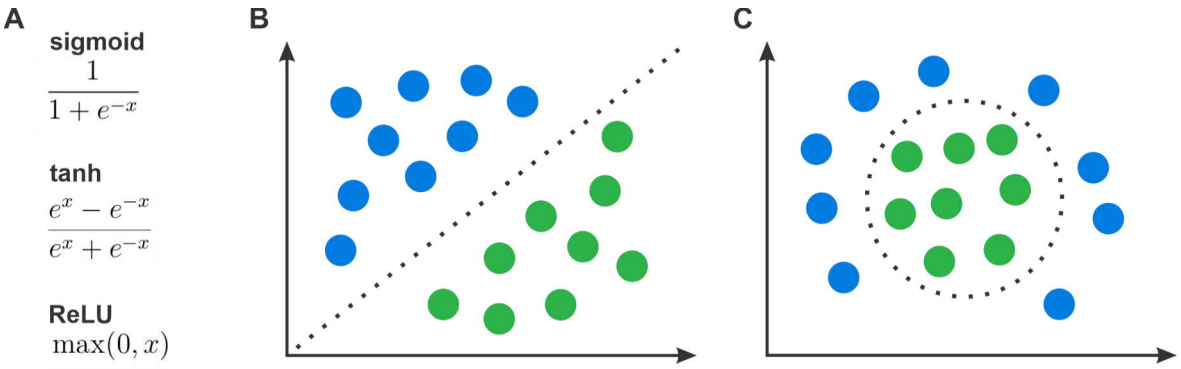

Figure 22: Activation function and a linear vs a non-linear classification problem. **A** Commonly used activation functions. **B** The dots have two features which are plotted on the x- and y-axis and stem from two classes highlighted in color. Since the dots can be separated with a line, this is a linear classification problem. **C** The dots from the two classes cannot be separated by a line, rendering this a non-linear classification problem.

4.8.2 Convolutional neural networks

Convolutional neural networks (CNNs) have emerged as a powerful class of artificial neural networks, particularly suited for spatial data and images. In comparison to traditional multi-layer perceptrons, CNNs exhibit distinct advantages that make them more effective for tasks where spatial relationships and hierarchical features play a crucial role.

CNNs leverage localized receptive fields, allowing them to focus on specific regions of input data (fig. 23 A) (Fukushima, 1980; LeCun et al., 1998). This design captures spatial hierarchies and helps detect features at different scales, which is challenging for MLPs with a fully connected architecture (Dumoulin et al., 2016; LeCun et al., 2015). CNNs use parameter sharing, meaning the same set of weights is applied to different parts of the input. This reduces the number of learnable parameters, making CNNs more effective for learning hierarchical features compared to MLPs. Furthermore, CNNs are translation invariant, meaning that features are recognized independent on their position in the image. This property is especially valuable for tasks such as image recognition, where the position of a feature in the input should not affect its detection. Unlike MLPs, CNNs can handle variable input sizes.

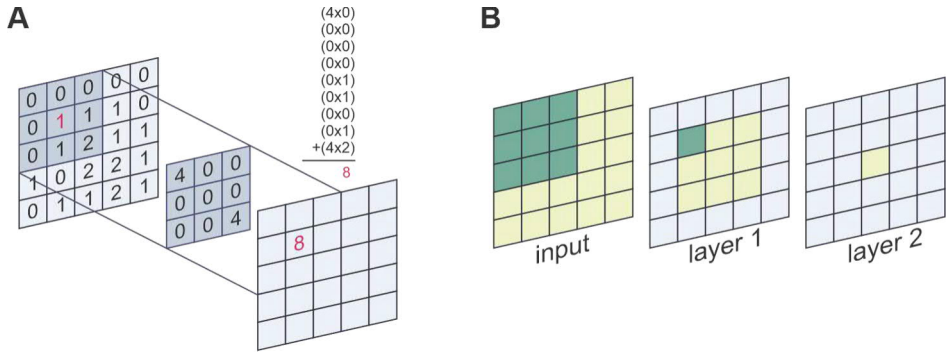

Figure 23: Receptive field and the convolutional operation. A The convolutional operation involves the multiplication of filter weights with the input, followed by summation to yield the output. The filter is slided over the input to create a feature map.**B** The receptive field is the region in the input space of a convolutional layer from which the layer extracts information. The receptive field is marked in green for the first and in yellow for the second convolutional layer.

At the core of a CNN is the convolutional operation (fig. 23 B). The convolution involves sliding a filter (also known as a kernel) over the input, performing element-wise multiplications, summation, and aggregating the results to create a feature map, which is the output of the convolutional operation. Multiple filters are applied in a convolutional layer to capture different features within the images. The filter weights are adapted during training and optimized for the problem at hand. The first convolutional layers in a neural network are known to capture lower-level features, while convolutional layers deeper in the architecture capture more complex features. The size of the feature map O is determined by the input size I, the kernel size K, padding P, and the stride S of the sliding operation (eq. 31). Padding is a technique of adding extra pixels as a border around an image. In the context of CNNs, it is used to control the output size of a convolutional operation and can help to reduce boundary artifacts and preserve the spatial dimension of the image.

$$O = \frac{I - K + 2P}{S} + 1 \tag{31}$$

Other layer types often used in CNNs are pooling layers (LeCun et al., 1998). They reduce the spatial dimensions of the input by downsampling. Common pooling operations include max pooling and average pooling (fig. 24 A). Pooling helps in creating a spatial hierarchy of information and reduces computation while retaining essential information (Yu et al., 2014).
Up-sampling layers increase the spatial dimensions of the input. Techniques like nearest-neighbor interpolation (fig. 24 B) or transposed convolutions (fig. 24 C) (Zeiler et al., 2011) are employed for up-sampling. Up-sampling layers regain resolution after downsampling and are handy for tasks where high-resolution outputs are required.

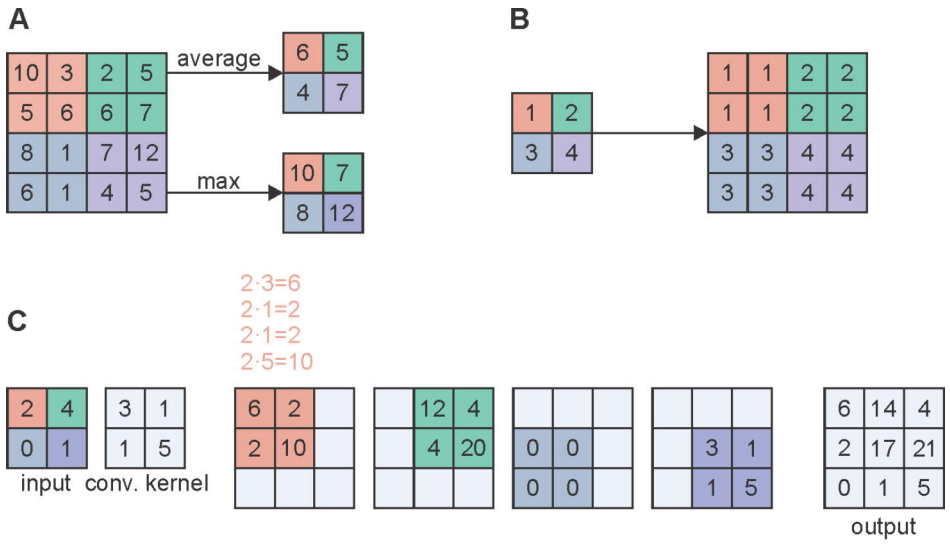

Figure 24: Different types of down- and upsampling layers. A Average and max pooling layers downsample the input by determining the maximum or average value in a defined region. **B** In nearest neighbor upsampling, each pixel in the output image is assigned the value of its nearest neighbor in the input image, replicating pixel values to achieve an enlarged image **D** A transposed convolution also enlarges the input image, but learns a mapping through training weights of a convolutional kernel.

Dense layers, also known as fully connected layers, are typically used in the final layers of a CNN for classification tasks. They take the high-level features extracted by convolutional and pooling layers and map them to the output classes.

A few practical applications are considered in the following. An example of a binary classification task would be to distinguish cells from two cell lines. Each cell has features such as the cell size, nucleus size, and growth rate. The input to the network are the features of the cells. The output is the probability of the cell belonging to cell line A or cell line B. The network is trained on numerous feature-class pairs to learn a mapping that discriminates between these classes. The size of the input vector is equivalent to the number of cell features. The number and size of hidden layers must be optimized. The size of the output vector refers to the number of cell lines. In the special case of two classes, one output node is sufficient with values between 0-0.5 referring to cell line A and values between 0.5-1

referring to cell line B.
Another example is the translation of one image to another, for example a bright-field image to a fluorescence image. The input are the pixels of the bright-field image and the output the pixels of the fluorescence image. Here, a CNN is an appropriate network choice, as it maintains the spatial relation of the pixels.

4.8.3 Training process

In this chapter the concepts of loss functions (chapter 4.8.3.1), train-validation-test splits (chapter 4.8.3.2), gradient descent (chapter 4.8.3.3), optimizers (chapter 4.8.3.4), and (chapter 4.8.3.5) are introduced.

4.8.3.1 Loss functions

The loss function measures the difference between the predicted output and the actual target values. It is minimized during the training process by adjusting the weights of the model. The choice of the loss function depends on the nature of the task.

For regression models, the mean absolute error (MAE) (eq. 32), or mean squared error (MSE) (eq. 33) are common choices as loss functions. The loss is expressed as an average across n data instances, while y refers to the ground truth and $\hat{y}$ to the prediction.

$$\mathcal{L}_{MAE} = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i \tag{32}$$

$$\mathcal{L}_{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{33}$$

The Charbonnier loss is defined as the square root of the MSE loss plus a small constant ϵ, typically set to 0.001 (eq. 34) (Charbonnier et al., 1994). Compared to the MSE loss it is less sensitive to outliers. The edge loss applies a Laplace filter Δ to the images to enhance the focus on edges during loss penalization (eq. 35) (Zamir et al., 2021).

$$\mathcal{L}_{Charbonnier} = \frac{1}{n}\sum_{i=1}^{n} \sqrt{||y_i - \hat{y}_i||^2 + \epsilon^2} \tag{34}$$

$$\mathcal{L}_{edge} = \frac{1}{n}\sum_{i=1}^{n} \sqrt{||\Delta y_i - \Delta \hat{y}_i||^2 + \epsilon^2} \tag{35}$$

LPIPS (Zhang et al., 2018a) and (MS)-SSIM (Wang et al., 2004; 2003), which are introduced as quality control metrics in chapter 4.10.1, are also suitable loss functions. Furthermore, mixing of multiple loss functions is an acknowledged possibility (eq. 36) (Zhao et al., 2015).

$$\mathcal{L}_{mix} = \alpha \cdot L_A + (1 - \alpha) \cdot L_B \tag{36}$$

For a classification model, the cross-entropy loss (eq. 37) is a common option.

$$\mathcal{L}_{cross} = -\frac{1}{n} \sum_{i=1}^{n} y_i log(\hat{y}_i) \tag{37}$$

4.8.3.2 Train-validation-test split

The train-validation-test split partitions the available dataset into three distinct subsets: the training set, the validation set, and the test set (Livingstone, 2010). The training dataset constitutes the largest portion of the dataset. The model learns patterns, relationships, and features within this set and its weights are adjusted accordingly during the training process to minimize the loss function. During the training process, the performance of the model is evaluated on the validation set, typically by calculating the loss on the validation data. This assessment helps in examining the generalization of the model ability to unseen data and identifying potential issues such as overfitting or underfitting (see chapter 4.8.3.5 for more details). The performance of the model is further benchmarked by applying suitable quality control metrics to a separate set of input-target pairs known as the test dataset, which were not involved in the training and validation process. Various quality control metrics tailored to different problem types are described in chapter 4.10.

Additionally to the trainable parameters (weights), there are so called hyperparameters that define aspects of the network architecture and training process, but are fixed and not adjusted during training. They typically include the learning rate, number of nodes, hidden layers, etc. They are preselected and different sets of hyperparameters are tried out to find an optimal setting for the network and data combination.

4.8.3.3 Gradient descent

In neural networks, the adjustment of the model weights w is the fundamental process to minimize the difference between predicted outputs and actual target values (Ruder, 2016). The iterative nature of this adjustment allows the model to adapt to complex patterns within the data. This is guided by an optimization algorithm, which tunes the weights so that the predictions of the model are in concordance to the ground truth. The difference between the predicted outputs and ground truth is described by a loss function and the weights are adapted in a way that minimizes the loss.

One of the basic optimization algorithms is gradient descent, which aims to minimize a loss function $\mathcal{L}(w)$ parametrized by the weights w of the model by iteratively adjusting the weights in the direction opposite to the gradient of loss function $\nabla_w \mathcal{L}(w)$ (Curry, 1944; Robbins et al., 1951; Ruder, 2016; Rumelhart et al., 1986; Werbos, 1974). The update rule for the weight adjustments during one iteration t is given by eq. 38, where α is the

learning rate, determining the step size in the weight space to reach a (local) minimum. In other words, the slope of the surface is followed until a valley is reached.

$$w_{t+1} = w_t - \alpha \cdot \frac{\delta \mathcal{L}}{\delta w_t} \tag{38}$$

Backpropagation is a crucial algorithm for efficiently computing gradients of the loss function with respect to the weights in a neural network (Rumelhart et al., 1986; Werbos, 1974). The algorithm begins with a forward pass, during which input data is fed into the neural network to produce output. The forward pass involves computing the weighted sum of inputs at each node, applying activation functions, and passing the information through the network layer by layer until the final output is obtained. The output of the neural network is compared to the actual target values, and a loss (error) is calculated. The loss function quantifies the difference between the predicted and actual outputs. The next step is the backward pass in which the gradient of the loss with respect to the weights of the network are computed. The chain rule is employed to calculate the gradient of the loss with respect to each weight. This allows the algorithm to calculate how changes in the weights at one layer affect the overall loss. The gradient indicates how much the loss would increase or decrease if the weights were adjusted. For a particular weight w_i in layer l, the gradient is calculated as

$$\frac{\delta \mathcal{L}}{\delta w_i^l} = \frac{\delta \mathcal{L}}{\delta z_i^l} \cdot \frac{\delta z_i^l}{\delta w_i^l}, \tag{39}$$

where z_i^l is the weighted sum at node i in layer l. The term $\frac{\delta \mathcal{L}}{\delta z_j^l}$ is obtained from the subsequent layer in the backward pass.

The gradients calculated in the backward pass are used to update the weights of the network in the direction that minimizes the loss (eq. 38). These steps are repeated multiple times. Each iteration involves a forward pass to generate predictions, a computation of the loss, a backward pass to calculate gradients, and an update of the weights using an optimization algorithm like gradient descent. The process continues until the model converges to a set of weights that minimize the loss function.

4.8.3.4 Optimizers

Optimizers are optimization algorithms that enhance the efficiency and speed of the gradient descent computation and weight updates during training (Ruder, 2016). Several optimizers have been developed to address challenges such as slow convergence and overcoming saddle points.

Gradient descent variants

There are different variants of the gradient descent algorithm, which differ in the specific point at which the weight updates occur during training. In case of batch gradient descent,

the gradient of the loss function is calculated for the entire dataset. This is computationally demanding and can pose challenges if the dataset exceeds memory capacity. Stochastic gradient descent (SGD) executes a weight update after each training instance, introducing high variance that leads to substantial fluctuations in the loss function during training (Robbins et al., 1951; Rosenblatt, 1958). While the volatility of SGD allows for exploration of new and potentially superior local minima, it simultaneously complicates the path to exact convergence. However, with SGD no memory issues occur. Mini-batch gradient descent strikes a balance between both strategies and takes a defined mini-batch of training data to perform the weight updates on (Bilmes et al., 1997; Shalev-Shwartz et al., 2010). This approach diminishes variance in the loss function while preserving computational efficiency, making it the preferred choice in neural network training. Typical batch sizes (the number of training examples utilized in one iteration) range from 32 to 256 and the choice of batch size is a hyperparameter (Ruder, 2016).

Momentum

SGD is known to encounter challenges in effectively navigating and surpassing saddle points during the optimization process (Dauphin et al., 2014; Sutton, 1986). Momentum is a method, in which SGD is accelerated in the direction of interest by introducing a momentum term v, based on a momentum constant γ (usually set to 0.9) and a moving average of past gradients (Qian, 1999) (eq. 40). This approach reduces oscillations in non-optimal directions and accelerates convergence. Nesterov accelerated gradient (NAG) also adds the ability to slow down momentum, preventing overshooting of the minimum (eq. 41) (Nesterov, 1983). The weights are updated for both approaches following eq. 42.

$$v_{t+1} = \gamma v_t + \alpha \nabla_w \mathcal{L}(w) \tag{40}$$

$$v_{t+1} = \gamma v_t + \alpha \nabla_w \mathcal{L}(w - \gamma v_t) \tag{41}$$

$$w_{t+i} = w_t - v_{t+1} \tag{42}$$

Adaptive learning rate

Another effective technique to find an optimum are adaptive learning rates. The adaptive gradient algorithm (AdaGrad) adjusts the learning rate for individual parameters by considering their past gradients (Duchi et al., 2011). The learning rate is tailored in AdaGrad to assign a smaller rate to parameters updated frequently (those with large past gradients) and a larger rate to parameters updated infrequently. This adaptability can be particularly advantageous in scenarios where certain parameters may require more cautious updates, preventing overshooting, while others may benefit from more substantial adjustments to expedite convergence. However, the algorithm suffers from radically diminishing learning

rates (Ruder, 2016). Adadelta builds on this concept, but uses the root mean square (RMS) of a running average of gradients, which alleviate the issue of overly small learning rates for frequently updated parameters (Zeiler, 2012). RMSprop, which was introduced in a lecture by Geoff Hinton[A] and remains unpublished, uses an exponentially decaying average of squared gradients in a sliding window.

Adaptive moment estimation (Adam) combines adaptive learning rates and momentum (Kingma et al., 2014). Adam with decoupled weight decay (AdamW) extends this concept by incorporating the regularization method weight decay in the update rule (Loshchilov et al., 2017). Further optimizers that combine these aspects are AdaMax (Kingma et al., 2014) and Nesterov momentum Adam (Nadam) (Dozat, 2016).

Learning rate schedules

Learning rate schedules influence the model convergence. Instead of using a fixed learning rate throughout training, learning rate schedules adjust the learning rate dynamically at different stages. A learning rate schedule may be defined independently from the optimizer. Commonly, a high learning rate is employed at the beginning to facilitate rapid convergence, followed by gradual reductions to fine-tune the model as it approaches a minimum.
Examples of learning rate schedules include step decay, where the learning rate is decreased by a fixed factor after a predefined number of epochs, and exponential decay, where the learning rate diminishes exponentially over time (You et al., 2019). A cyclic learning rate schedule periodically varies between a minimum and maximum value during training, allowing the model to escape local minima for improved convergence and generalization (Smith, 2017). Reduction of learning rates on plateau is a concept that diminishes the learning rate when the validation loss fails to improve for a set number of epochs. These schedules help strike a balance between rapid initial progress and refined optimization towards the end of training, contributing to improved model generalization and performance.

Convergence is achieved when the model reaches a point where further adjustments to the weights do not significantly reduce the loss. To determine if the model converged, it is helpful to display the loss function against the number of epochs, with each epoch signifying a complete iteration through all data instances. During training, a subset of the training data, termed the validation dataset, is set aside and not used for training purposes. Following each iteration, the performance of the model is evaluated on the validation data by computing the loss.

4.8.3.5 Variance and bias

Variance and bias play a crucial role in determining the performance and generalization capability of a neural network (Geman et al., 1992; Livingstone, 2010). Generalization refers to the ability of a model to perform well on data it hasn't been explicitly trained on.

[A] http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.png

Bias is the error introduced by approximating a problem by a simplified model (fig. 25 A). High bias can lead to underfitting, where the model is too simplistic and fails to capture the underlying patterns in the data. Variance is the error introduced by the sensitivity of the model to small fluctuations or noise in the training data (fig. 25 B). A high-variance model is overly complex and captures noise in the training data as if it were a real pattern. High variance can lead to overfitting, where the model performs well on the training data but fails to generalize to new, unseen data. With increasing model capacity, the bias decreases while the variance increases, known as the bias-variance tradeoff (fig. 25 C). Achieving a good balance between bias and variance is essential for building models that generalize well to new, unseen data (fig. 25 D).

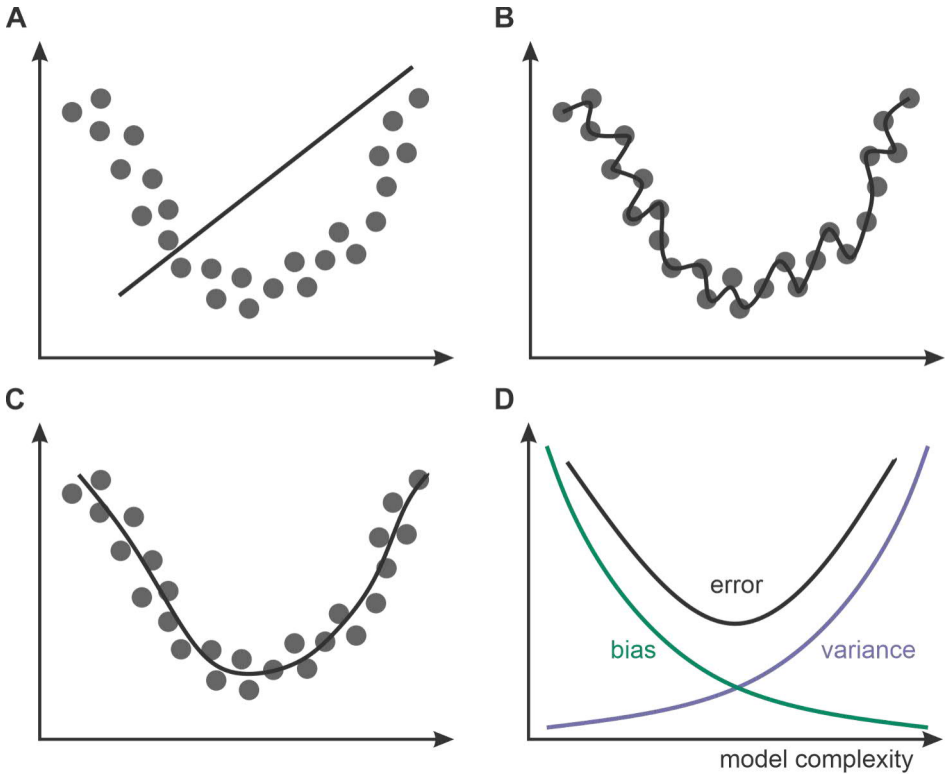

Figure 25: Bias and variance tradeoff. A A too simplistic model is fit to the data, which fails to describe the data properly. The model is underfitting the data and has a high bias. **B** A too complex model is fit to the data, learning the noise. The model is overfitting the data and has a high variance. **C** Bias-variance tradeoff. With increasing model complexity the bias decreases while the variance increases, with an optimal model complexity in the valley of the U-shaped error curve. **D** A model with optimal complexity is fit to the data, which reflects the trend of the data well.

To detect bias and variance, the discrepancies between the training and validation loss can be studied. Underfitting occurs when the model is too simplistic to capture the underlying patterns in the data, leading to poor generalization. The loss curves are relatively high and flat, suggesting that the model is incapable of learning the pattern in the data (fig. 26 A). Underfitting may arise from a model that lacks complexity or has not been trained for a sufficient number of epochs. In contrast, overfitting manifests when the training data is

too small. The model might also be too complex with a larger tendency to overfit. A large gap between the training and validation curve is an indication for overfitting (fig. 26 B). This might also happen if the model is trained for too many epochs and start to learn the noise at a later stage of training.

Besides adjusting the number of training instances and the model complexity, early stopping is a technique to prevent overfitting. This strategy monitors the performance of the model on the validation data, halting the training process if no improvement is observed over a predefined number of epochs. Early stopping not only prevents overfitting but also optimizes computational resources by avoiding unnecessary training time. Early stopping belongs to a toolbox of regularization techniques that prevent overfitting, such as L1 and L2 regularization (Krogh et al., 1991; Weigend et al., 1990), dropout (Hinton et al., 2012; Tompson et al., 2015), data augmentation (Balestriero et al., 2022; Shorten et al., 2019), gradient clipping (Mikolov, 2012), and batch normalization (Ioffe et al., 2015), which will not be discussed here in detail.

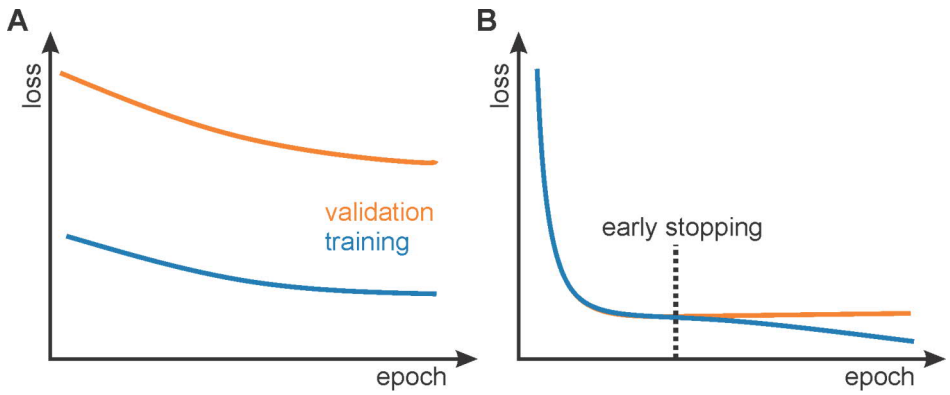

Figure 26: Learning curve examination to detect bias and varianc. **A** The loss function of the training and validation data is displayed per epoch. An underfitted model exhibits relatively high and flat loss values. **B** An overfitted model displays discrepancies between the training and validation loss. The model starts to overfit if trained for too many epochs. A method to prevent overfitting is early stopping, which interrupts the training process if the validation loss has not improved over a predefined number of epochs.

4.9 Network architectures

In this chapter, two pivotal neural network architectures for computer vision are introduced: the U-Net architecture (Ronneberger et al., 2015) and the residual channel attention network (RCAN) (Zhang et al., 2018b).

4.9.1 U-Net architecture

The U-Net architecture by Ronneberger et al., 2015 is based on convolutional layers and was originally demonstrated for segmentation of EM and light microscopy images (fig. 27). The architecture translates input images to desired output images and consists of an encoder and decoder path (fig. 27).

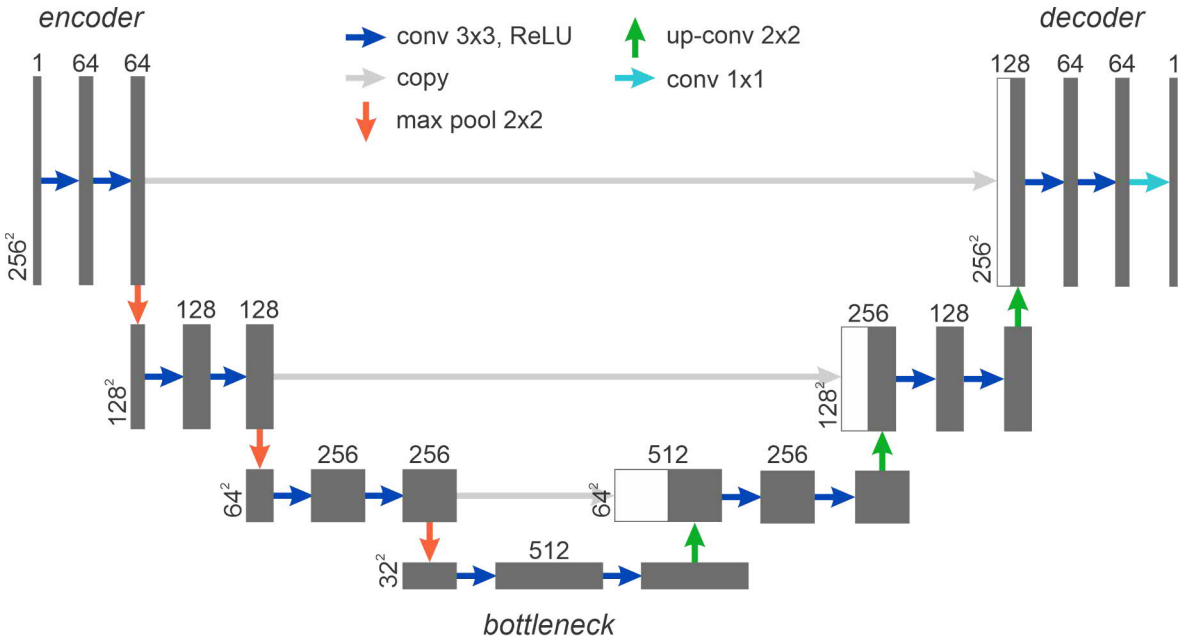

Figure 27: U-Net architecture. An input image is translated to a target image. The network consists of an encoder and decoder path, connected through skip connections. Other building blocks are convolutional layers, activation functions, max pooling layers, and a 1x1 convolution. The number of feature maps and the width and height of the input in the layers is denoted with numbers. Adapted from Ronneberger et al., 2015.

The encoder path of the network downsamples the images in their width and height. It consists of a series of convolutional layers followed by an ReLU activation function and a max-pooling operation. The max-pooling layers reduce spatial resolution to extract features with the convolutional layers at different scales. For each reduction in spatial dimension, the number of filters is increased. This process is repeated until the desired width and height of the images at the bottleneck is reached. The bottleneck serves as a bridge between the encoder and decoder paths. It is constructed of a convolutional layer with a large receptive field, capturing global context information. Due to the compact representation of the input

image, this facilitates the extraction of essential features. The decoder upsamples the images by using up-convolutional layers back to the original input size. Furthermore, the feature maps of the encoder and decoder are combined through residual (skip) connections (He et al., 2015) at multiple levels of the network. Residual connections involve the addition of the original input to the output of a neural network layer (fig. 28), facilitating the flow of information through the network and mitigating issues such as vanishing gradients during training. This approach effectively combines low-level and high-level features, facilitating accurate image prediction at multiple scales. The final layer uses a 1x1 convolution to produce output images.

The U-Net architecture has been widely applied and adapted in the field of (super-resolution) light microscopy, such as image denoising (Buchholz et al., 2020; Ebrahimi et al., 2023; Goncharova et al., 2020; Krull et al., 2018; Weigert et al., 2018), segmentation (Buchholz et al., 2020; Schmidt et al., 2020; Stringer et al., 2021), image translation (Chen et al., 2021; Guo et al., 2020; Ounkomol et al., 2018; Wang et al., 2019), and multi-emitter fitting (Nehme et al., 2018; Speiser et al., 2021).

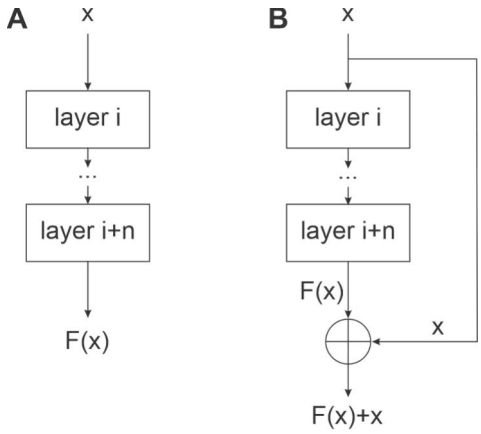

Figure 28: Residual connection. **A** Traditional feed forward connection. **B** Feed forward connection with residual connection. The identity of the input x is added to the feed forward output F(x) (He et al., 2015).

4.9.2 Residual channel attention network

The residual channel attention network (RCAN) was developed by Zhang et al., 2018b and popularized for super-resolution microscopy by Chen et al., 2021. It is specifically designed to map a low-resolution (LR) input to a super-resolution (SR) output (Zhang et al., 2018b). In the field of computer vision, super-resolution means enhancing the resolution or details of an image beyond its original size or quality. In the LR input, low-frequency information is abundant with the goal to recover high-frequency information. Key features of this architecture are residual connections and channel attention layers (fig. 29). The residual connections allow for deep neural network training and provide a large receptive field with information being investigated more globally. Conventional CNN-based approaches often

treat the feature maps in a convolutional layer equally, potentially resulting in a deficiency to emphasize relevant information across various feature channels. These two key designs allow the abundant low-frequency information to be bypassed, emphasizing on the more relevant high-frequency features of the images.

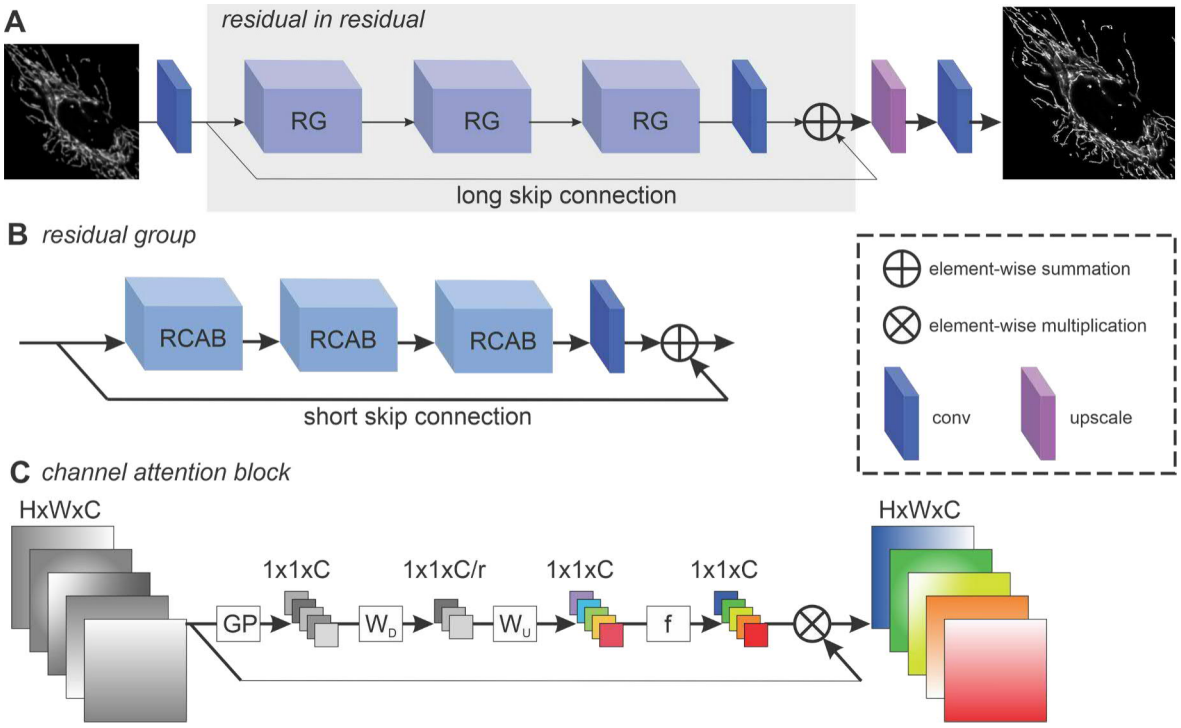

Figure 29: RCAN architecture. **A** The residual in residual block is built from multiple residual groups and is framed by a long skip connection. **B** The residual groups are build from multiple residual channel attention blocks and frames by a short skip connection. **C** Channel attention blocks are trained to yield weights for feature map importances. Adapted from Zhang et al., 2018b.

The RCAN architecture is constructed of four parts: A shallow feature extraction, residual in residual (RIR) deep feature extraction, an upscale module, and a reconstruction part. To extract shallow features, referring to basic features of an image, such as edge, color, or basic shapes, a convolutional layer is used. The RIR consists of G residual groups (RGs) with a long skip connection to allow flow of information across the RGs (fig. 29A). Each RG consists of a residual channel attention block (RCAB) with short skip connections (fig. 29B). Channel attention is designed to adaptively recalibrate the importance of different channels in the feature maps (fig. 29C). By assigning different attention weights to channels, the network can focus on relevant information and suppress irrelevant features, enhancing its ability to capture important details. The mechanism applies global average pooling to the channels to aggregate their values into a single number (GP). This is followed by a convolutional layer ($\mathrm{W_D}$), which downscales the number of channels based on a reduction ratio r. A non-linear function is applied to capture complex channel interdependencies. The low-dimensional signal is upsampled by the ratio r with another convolutional layer

($\mathrm{W_U}$). The output passes a sigmoid-gate, scaling each number to a range between 0-1, generating a scaling factor s. The factor is multiplied with the original channel weights, to give the channels individual importances. With channel attention, the residual component in the RCAB is rescaled. The upscale module is realized with a transpose convolution. The module can be deactivated if LR and SR should have same dimensions. The image is reconstructed with one final convolutional layer.

The residual learning and channel attention mechanisms contribute to the ability of the network to reconstruct finer details in the high-resolution output. Ebrahimi et al., 2023 combined the RCAN architecture with a U-Net to denoise STED images while maintaining high frequency information.

4.10 Quality control metrics

The evaluation of the performance of a network relies on test data, comprising input and corresponding ground truth pairs that were not part of the training process. Compared to the evaluation based on the loss curves, quality control metrics allow an in depth evaluation of the network performance. Various quality control metrics exist to quantify the disparity between the output of the network and the ground truth in the test data. The choice of metrics depends on the nature of the task. An in depth review of metric choices and potential pitfalls has been recently published (Maier-Hein et al., 2024; Reinke et al., 2024). This chapter introduces metrics for assessing image similarity (chapter 4.10.1), highlights metrics which were specifically developed for super-resolution microscopy (chapter 4.10.2), and gives an overview of measures for resolution (chapter 4.10.3) as well as metrics for classification tasks (chapter 4.10.4).

4.10.1 Image similarity metrics

Image similarity metrics describe the difference between two images. In the context of neural networks, the difference of a ground truth image, for example a segmentation or the super-resolution image, and a predicted image are calculated.

4.10.1.1 Pixel-independent metrics

Pixel-independent comparison metrics calculate the difference between the predicted image $\hat{y}$ and ground truth image y per pixel. They are termed pixel-independent metrics, as they solely compare pixels at the same position between two images and neglect broader context information. The overall score is then given by averaging over all pixel values. Typical pixel-independent metrics include the MAE (eq. 43), MSE (eq. 44), and the root mean squared error (RMSE) (eq. 45). In eq. 43-45, ij denotes the pixel position in a 2D image

and N refers to the total number of pixels. The lower the score the more similar the images, where a score of zero means that the compared images are exactly the same.

$$MAE = \frac{1}{N}\sum_{ij}^{N}(y_{ij} - \hat{y}_{ij}) \tag{43}$$

$$MSE = \frac{1}{N}\sum_{ij}^{N}(y_{ij} - \hat{y}_{ij})^2 \tag{44}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{ij}^{N}(y_{ij} - \hat{y}_{ij})^2} \tag{45}$$

The peak signal-to-noise ratio (PSNR) is mainly applied to quantify the loss in compression algorithms (eq. 46), but has gained popularity in the microscopy community (Chamier et al., 2021; Chen et al., 2021; Lequyer et al., 2022). It evaluates the quality of a reconstructed or compressed signal, such as an image, by comparing it to the original, uncompressed signal. It utilizes the decibel scale, where L is the maximum possible pixel value, for example 255 for an 8-bit image, and MSE is the mean squared error as described in eq. 44. Low PSNR values indicate compromised image quality, while elevated PSNR values mean less impact of noise.

$$PSNR = 20 \cdot log_{10}(L) - 10 \cdot log_{10}(MSE) \tag{46}$$

The PEARSON correlation coefficient r describes the linear relationship of two variables (eq. 47) (Pearson, 1895). In the context of comparing two images, it evaluates the relationship of pixel intensities. The individual pixel values are y_{ij} and $\hat{y}_{ij}$ and the averages across all pixels are $\langle y \rangle$ and $\langle \hat{y} \rangle$. A PEARSON correlation coefficient of 1 means perfect linear relationship between the images, a coefficient of 0 means no linear relationship, and a coefficient of -1 means perfect negative linear relationship. If two images are exactly the same, they yield a PEARSON correlation coefficient of 1.

$$r = \frac{\sum_{ij}^{N}(y_{ij} - \langle y \rangle)(\hat{y}_{ij} - \langle \hat{y} \rangle)}{\sqrt{\sum_{ij}^{N}(y_{ij} - \langle y \rangle)^2 \sum_{ij}^{N}(\hat{y}_{ij} - \langle \hat{y} \rangle)^2}} \tag{47}$$

4.10.1.2 Pixel-dependent metrics

Pixel-independent metrics tend to not correlate well with the human perception of image similarity, as they focus on pixel-wise differences without considering the structural information in an image (Wang et al., 2004; Zhang et al., 2018a). A perceptually motivated measure is the structural similarity metric (SSIM), designed to capture perceived similarities in images (Wang et al., 2004). Differences in luminance l, contrast c, and structure s are

taken into account and weighted by the factors α, β, and γ (eq. 48). Terms for the three contributions are derived based on the average pixel intensities μ_y and $\mu_{\hat{y}}$, the standard deviations σ_y and $\sigma_{\hat{y}}$, and correlation coefficient $r_{y,\hat{y}}$ within a sliding window. With assumptions such as equal weighting of the three contributions, the SSIM is expressed with eq. 49. The constants C_1 and C_2 prevent denomination by zero and are based on a small constant $K_i \ll 1$ and the dynamic range of the pixel values L (eq. 50) (for example 255 for a 8-bit image). The overall SSIM is the average across all sliding windows. SSIM ranges from 0 (no similarity) to 1 (maximum similarity). Multi-scale SSIM (MS-SSIM) is an extension of SSIM and determines the metric on different down-scaled versions of the original image (Wang et al., 2003).

$$SSIM(y, \hat{y}) = l(y, \hat{y})^{\alpha} \cdot c(y, \hat{y})^{\beta} \cdot s(y, \hat{y})^{\gamma} \tag{48}$$

$$SSIM(y, \hat{y}) = \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)} \tag{49}$$

$$C_i = (K_i \cdot L)^2 \tag{50}$$

Blur is a disturbing component in a super-resolution image. Blurring causes large perceptual distances, but small discrepancies in pixel-wise metrics (Zhang et al., 2018a). Also SSIM has the tendency to be less sensitive to blur compared to other metrics (Abdullah-Al-Mamun et al., 2021; Zhang et al., 2018a). A metric that punishes blur more strongly is the learned perceptual image patch similarity (LPIPS) (Zhang et al., 2018a). This metric utilizes deep neural networks such as AlexNet (Krizhevsky et al., 2012) or VGG (Simonyan et al., 2014), which were trained for an image classification task. Feature maps from the first convolutional layers are extracted for the ground truth and predicted image, further processed, and compared with Euclidean distance.

4.10.2 Super-resolution microscopy specific metrics

A metric specifically designed for super-resolution microscopy images is the Haar wavelet kernel analysis method for the assessment of nanoscopy (HAWKMAN) metric, which evaluates differences between two SMLM images (Marsh et al., 2021). HAWKMAN disregards intensity information and focuses exclusively on structural disparities. The evaluation involves comparing a ground truth image, often approximated with a low emitter density image, with a test image to determine its quality. HAWKMAN entails several steps, such as flattening, binarization, Gaussian blurring, and skeletonization, to generate various representations of the data that highlight different aspects. One special feature of this evaluation metric is that the images are blurred with varying widths of Gaussians, rang-

ing from a pixel size up to the PSF of the instrument. This allows the determination of image quality at different length scales. The sharpening map, which incorporates information about localization precision, approximates the full width at half maximum (FWHM) of the structure. It reveals biases in dense emitter regions, particularly mislocalizations toward overlapping areas, which are especially present in crowded areas, such as cross sections of tubular structure (Costello et al., 2021; Marsh et al., 2021; Wolter et al., 2011). The structure map, which depicts skeletonized peak positions, exposes instances where structures have merged due to the inability to resolve nearby emitters. For the sharpening and structure map the PEARSON correlation coefficient between the maps of the ground truth and test image is provided for all length scales. The confidence map, derived from the local correlation of sharpening and structure maps, serves as a confidence metric for the reliability of specific regions in the test dataset. Finally, the artifact scale map highlights the minimal scale length without artifacts per pixel.

Another metric specifically designed to evaluate super-resolved images is super-resolution quantitative image rating and reporting of error locations (SQUIRREL), which downscales the super-resolved image I to compare it with a diffraction-limited image I_D of the same FOV (Culley et al., 2018). This makes the metric sensitive to disappearing and false positive structures in the super-resolved image, to bright aggregates, and structure merging on the PSF scale. The downscaling is either achieved by determining a resolution scaling function (RSF) in an iterative optimization process, or by using the PSF of the instrument (measured or simulated). The RSF is then convolved with the super-resolved reconstruction to match the resolution of the wide-field image. The resolution scaled error (RSE) (eq. 51) and PEARSON correlation coefficient (eq. 47 are computed between the RSF-convolved super-resolution and the wide-field image, with N being the total number of pixels and $\bar{I}$ and $\overline{I_D}$ the average values of the images. Spatial discrepancies are displayed by mapping out the pixel-wise absolute difference between these images. A RSE of 0 and a PEARSON correlation coefficient of 1 means that the compared images are equivalent.

$$RSE = \sqrt{\frac{\sum_{ij}^{N}(I_D(i,j) - I(i,j))^2}{N}} \tag{51}$$

While SQUIRREL proves highly useful in quantifying artifacts at the PSF scale, it is important to note that its capabilities are limited to assessing errors occurring at or beyond this specific scale (Costello et al., 2021). Instead of a super-resolved image I and a diffraction-limited image I_D, a ground truth image y and a predicted image $\hat{y}$ can be used. As no diffraction-limited image is used, this allows comparisons also below the PSF scale. Compared to SQUIRREL, HAWKMAN neglects intensity information and focuses solely on the structure and provides artifact quantification on several scales.

4.10.3 Image resolution

The resolution serves as another metric to determine the success of an image prediction, as the prediction should maintain the resolution of the ground truth.

The FWHM of the intensity line profile serves as an estimate for the resolution. However, this approach is laborious as the lines must be inserted by hand and error-prone as the selected lines might be narrowed by noise (Tortarolo et al., 2018).

A more robust measure for the resolution is the FOURIER ring correlation (FRC), which was originally introduced for electron microscopy (Heel et al., 1982; Saxton et al., 1982). To determine the resolution, two images of the same target are required with independent noise. The images are FOURIER transformed and their frequencies cross correlated within concentric rings of increasing radius, where F_1 and F_2 are the FOURIER transforms of the two images and r_i refers to the frequency bins (eq. 52). Noise manifests as high-frequency components in the FOURIER space and does not correlate, as it is random. Therefore, by increasing the radius of the concentric rings, noise plays an increasing role and diminishes the cross correlation value. The resolution is the inverse of the frequency, where the correlation drops below a set threshold (typically 1/7). When comparing the resolution of a predicted and ground truth image, an image pair of the input is required, to generate two predicted images of the same signal. The determination of the resolution of the ground truth also requires an image pair. Ideally, ground truth and prediction should have similar resolutions.

$$FRC = \frac{\sum_{r_i \in r} F_1(r) \cdot F_2(r)}{\sqrt{\sum_{r \in r_i} F_1^2(r) \cdot \sum_{r_i \in r} F_2^2(r)}} \tag{52}$$

Decorrelation analysis also determines the resolution based on FRC, but only requires one image of ground truth and prediction, as it compares the cross correlation curve of a normalized FOURIER transformed image to a low-pass filtered version (Descloux et al., 2019). Rolling FRC (rFRC) uses a sliding window to inform about the resolution locally instead of the entire image (Zhao et al., 2023).

4.10.4 Metrics for classification tasks

In this chapter, metrics are discussed which are commonly used in classification tasks, where it is relevant to quantify how many instances have been detected in the prediction compared to the ground truth. The chapter starts by introducing the basic metrics true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). Furthermore, the confusion matrix and accuracy, recall, precision, and the F1 score are explained. Finally, the intersection over union (iou) is introduced.

4.10.4.1 Basic metrics

True positives (TP) represent the instances correctly classified as positive by the model, while false positives (FP) are instances incorrectly classified as positive and missing in the ground truth (Sokolova et al., 2006). False negatives (FN) are instances incorrectly classified as negative, meaning that ground truth instances are missing in the prediction. True negatives (TN) represent instances correctly classified as negative. These metrics are fundamental for understanding the ability of the model to correctly identify positive and negative cases.

The confusion matrix is a comprehensive table that summarizes a classification performance of the model (eq. 53). For a classification problem with a positive and a negative class it summarizes the four different cases of TP being instances correctly predicted as belonging to the positive class, FP being instances that are incorrectly predicted as belonging to the positive class (actual class is negative), TN being instances that are correctly predicted as belonging to the negative class, and FN being instances incorrectly predicted as belonging to the negative class (actual class is positive).

$$\begin{array}{lcc} & ground\ truth\ positive & ground\ truth\ negative \\ predicted\ positive & TP & FP \\ predicted\ negative & FN & TN \end{array} \tag{53}$$

This can be extended to N multiple classes with an NxN matrix (eq. 54). The diagonal of the matrix refers to the TP counts. All other entries are either FP, FN, or TN. To clarify this, the basic metrics are listed for class 1 (TP = 9, FN = 1+0 = 1, FP = 4+1 = 5, TN = 5+2+3+6 = 16). This can be repeated for all classes.

$$\begin{array}{lccc} & ground\ truth\ class\ 1 & ground\ truth\ class\ 2 & ground\ truth\ class\ 3 \\ predicted\ class\ 1 & 9 & 4 & 1 \\ predicted\ class\ 2 & 1 & 5 & 2 \\ predicted\ class\ 3 & 0 & 3 & 6 \end{array} \tag{54}$$

4.10.4.2 Performance metrics

A variety of performance metrics exist, that are based on the above introduced basic metrics. Dependent on what should be emphasized when evaluating a classification task, different performance metrics can be used.

The accuracy measures the overall correctness of a classification model (eq. 55). It is the ratio of correctly predicted instances to the total instances in the dataset.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{55}$$

The false positive rate quantifies falsely detected instances, as the denominator describes the total number of true negatives (eq. 56).

$$FPR = \frac{FP}{FP + TN} \tag{56}$$

Recall, also known as sensitivity or true positive rate, is a measure of how well the model identifies positive instances (eq. 57).

$$Recall = \frac{TP}{TP + FN} \tag{57}$$

Precision quantifies the accuracy of positive predictions made by the model, calculated with the proportion of true positives among all positive predictions (eq. 58).

$$Precision = \frac{TP}{TP + FP} \tag{58}$$

The F1 score is the harmonic mean of precision and recall, providing a balanced metric that considers both false positives and false negatives (eq. 59).

$$F1 = \frac{2}{recall^{-1} + precision^{-1}} \tag{59}$$

An extension is the Fβscore, where the weighting between precision and recall can be chosen.

All the mentioned performance metrics range from 0 to 1, with a score of 0 indicating the poorest performance and a score of 1 representing the best performance.

4.10.4.3 Spatial metrics

Various metrics offer insights into the spatial accuracy and overlap between predicted and ground truth binary images.

The iou measures the overlap between predicted and ground truth bounding boxes, providing insights into the spatial accuracy of a model (Jaccard, 1912). The score is calculated by taking the ratio between the intersection $|y \cap \hat{y}|$ and the union $|y \cup \hat{y}|$ (eq. 60). Let's consider the case where segmentations of cells are compared in a ground truth and predicted image. The iou quantifies the overlap of cell segmentations in the images and therefore gives further insights into model performance. The iou can also be expressed with the basic metrics and ranges from 0 (=no overlap) to 1 (perfect overlap).

$$iou = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|} = \frac{TP}{TP + FP + FN} \tag{60}$$

The SØRRESNEN-DICE coefficient is a similar score with higher emphasis on the intersection between y and $\hat{y}$ compared to the iou (eq. 61) (Dice, 1945; Sørensen, 1948).

$$Dice = \frac{2 \cdot |y \cap \hat{y}|}{|y| + |\hat{y}|} = \frac{2TP}{2TP + FP + FN} \tag{61}$$

5 Accelerated DNA-PAINT imaging with a neural network

Single-molecule localization microscopy is a super-resolution imaging technique based on the localization of isolated emitter positions over many images. The detected localizations are merged to produce a super-resolved image that enables the visualization of biological structures at the nanoscale. The need for large amounts of frames in an SMLM measurement results in low throughput and poor time resolution, which is particularly critical when imaging living cells. Accelerating acquisition by increasing the emitter density per image requires advanced "emitter" fitting algorithms capable of deciphering overlapping PSF signals. Deep learning approaches have proven to be powerful in restoring super-resolved images for challenging data sets with high density (Nehme et al., 2020; 2018; Speiser et al., 2021).

DeepSTORM is a neural network based approach for obtaining super-resolution images and was demonstrated with STORM data (Nehme et al., 2018). However, *DeepSTORM* relies on suitable emitter densities and STORM inherently suffers from photobleaching during image acquisition, resulting in emitter densities possibly not longer being in the optimal performance window of the trained neural network.

Here, the potential of applying *DeepSTORM* to DNA-PAINT data is demonstrated. DNA-PAINT offers a constant emitter density due to the continuous exchange of labels. A comprehensive workflow to acquire the training data based on experimental measurements and the test data containing both the high emitter density images and the ground truth is presented. By combining DNA-PAINT with neural network based image reconstruction, a rapid image acquisition of multi-color super-resolution images of semi-thin neuronal tissue is enabled within minutes.

5.1 Results and discussion

5.1.1 Training and testing of *DeepSTORM* with DNA-PAINT imaging

To facilitate the training of *DeepSTORM* using experimental data, high density DNA-PAINT images along with a matching list of emitter positions must be retrieved experimentally (fig. 30A) (Narayanasamy et al., 2022). As the precise localization of high density movies is not possible, low density DNA-PAINT measurements were acquired and emitter positions determined. The summation of patches and merging of the localization lists from these images resulted in high density patches, grounded in experimental data, which served as the training data for *DeepSTORM*.

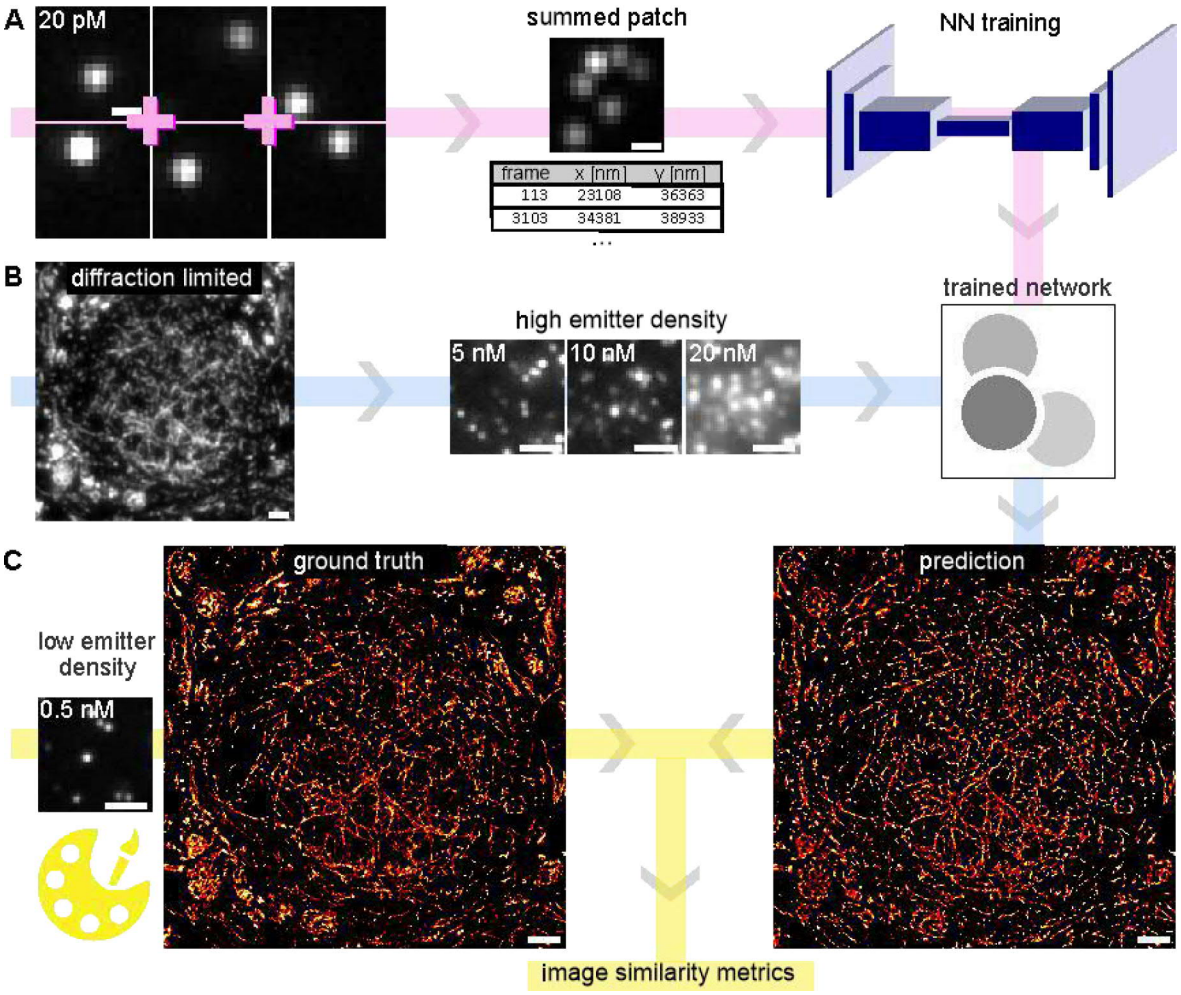

Figure 30: *DeepSTORM* workflow with experimental DNA-PAINT data. **A** Low density patches are summed to generate high density patches along with their emitter positions based on experimental data, which serves as training data for the *DeepSTORM* model. **B** High emitter density data is acquired and used as input for the trained model, which predicts a super-resolved image. **C** A low emitter density DNA-PAINT movie at the same FOV as the high emitter density measurements serves as the ground truth and is compared to the *DeepSTORM* prediction to evaluate the model performance. Scale bars 0.5 µm in **A** and 2 µm in **B** and **C**. Adapted from Narayanasamy et al., 2022, CC-BY-4.0.

Leveraging an encoder-decoder architecture, *DeepSTORM* transforms diffraction-limited input images into super-resolved output images. To generate the super-resolved output image, the emitter positions from the list are rendered into a super-resolved image. The output of *DeepSTORM* consists of a stack of super-resolved images per input frame and one super-resolved image comprising all emitters across the movie. Coordinates can be retrieved in a postprocessing step from the super-resolved stack within the *DeepSTORM* framework.
To quantify the performance of the network, *DeepSTORM* was applied to experimental high density data at varying imager strand concentrations (5 nM, 10 nM, 20 nM) (fig. 30B). To establish a ground truth for the performance evaluation, low density data (0.5 nM imager strand concentration) were acquired from the same field of view to generate a super-resolved image with single-molecule DNA-PAINT reconstructed using the software *Picasso* (fig. 30C). The similarity of the ground truth *Picasso* image and the predicted super-resolved image produced by *DeepSTORM* was assessed by various image similarity metrics.

Acquiring appropriate high emitter density test dataset entails addressing two key considerations: firstly, the imager strand concentration of the test dataset must generate a density that aligns with the training data, and secondly, the required number of frames to fully capture the underlying structure must be ascertained.
As the training data is built from sparse emitters successfully detected by a conventional localization algorithm, its density is known. The sparse emitters are summed into higher density patches along with their corresponding localization lists, resulting in a distribution of emitter densities across the patches. This makes the model resilient to density variations around this distribution. Deviating from these densities could lead to a decline in model performance. Consequently, the optimal imager strand concentration of the test dataset must be systematically explored. Extrapolating the emitter density from low density measurements aids in estimating the density of the high imager strand concentrations (fig. 31A), considering that conventional algorithms are reliantly detecting emitters at low concentrations but falter in high density environments. Furthermore, testing different imager strand concentrations is advisable, as densities might differ depending on the target structure. For instance, in a tubular structure, emitters may exhibit strong overlap due to the thin tubular structure. However, if the tubular structure itself is sparsely distributed across the field of view, the overall density of the image might appear low.

Besides the density optimization, the question of determining the requisite number of high density frames to fully capture the target structure arises. Excessive frames in a DNA-PAINT movie may redundantly image the same position multiple times, yielding no additional information. To ascertain the optimal number of frames, the high density movie is cropped to varying frame counts and processed with *DeepSTORM*. The resulting predicted super-

resolved images are then compared to the ground truth with image similarity metrics. The saturation of these metrics with the addition of redundant frames indicates the point at which an adequate number of frames is achieved to completely capture the structure (fig. 31B).

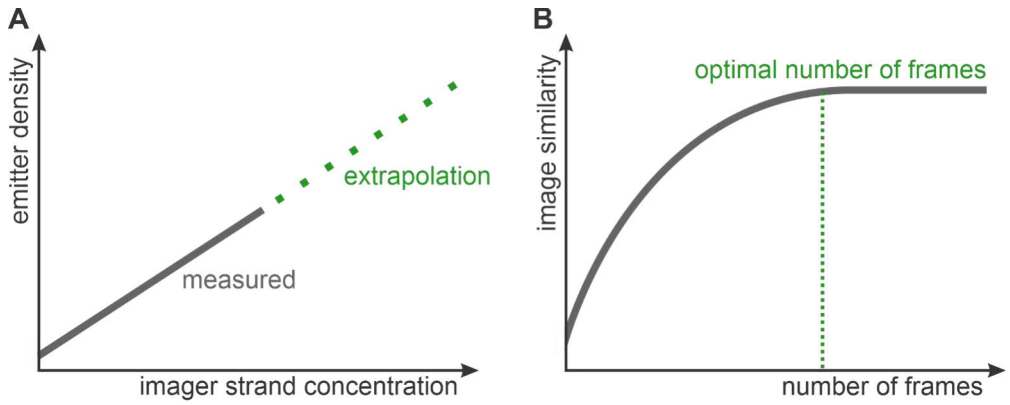

Figure 31: Optimization of the high density movie. **A** Different imager strand concentrations are measured and their resulting emitter densities determined in a concentration range, where conventional algorithms reliably detect emitters. The density of high concentrated measurements is extrapolated. **B** The number of required frames in a high density movie is determined as the saturation point, where more frames do not improve the image similarity between ground truth and prediction.

Table 1: Image similarity metrics for the comparison of ground truth and *DeepSTORM* predicted images of α-tubulin (red) and TOM20 (red) in MNTB tissue.

target	MS-SSIM	RSP	RSE	resolution / nm	resolution / nm
				prediction	ground truth
Tubulin	0.785	0.839	5707	44	32
TOM20	0.750	0.754	2350	41	33

5.1.2 Single-emitter fitting with *DeepSTORM*

To probe the capacity of *DeepSTORM* in handling DNA-PAINT movies, α-tubulin and TOM20 was imaged in MNTB tissue with exchange DNA-PAINT (Jungmann et al., 2014) at imager strand concentrations compatible with single-emitter fitting routines (α-tubulin: P1, 500 pM, 8000 frames, TOM20: P5, 500 pM, 8000 frames). The movies were processed with *Picasso* to generate a ground truth (fig. 32A). A *DeepSTORM* model was trained on patches with the same density acquired on a separate measurement (fig. 32B). The resulting super-resolved images from both methods were compared and yielded satisfying image similarity, meaning that *DeepSTORM* is capable of reconstructing DNA-PAINT data and produces similar results compared to *Picasso* (table 1). The resolution was slightly decreased, but still within acceptable range.

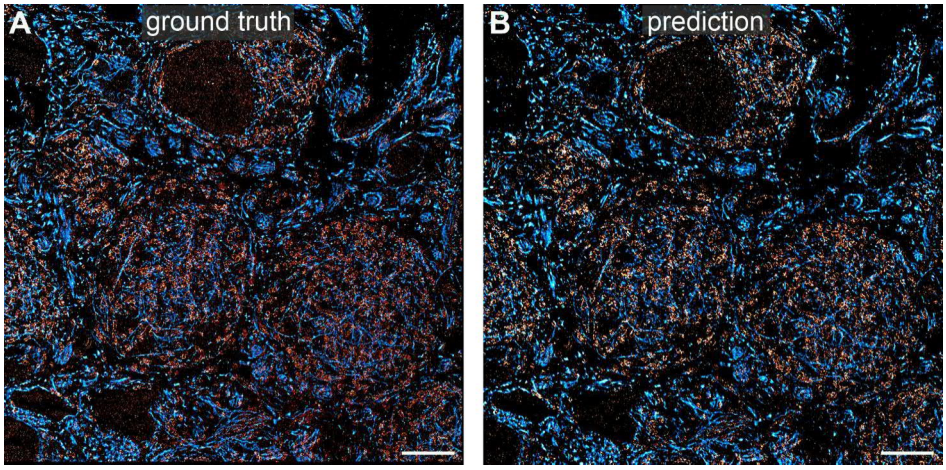

Figure 32: *DeepSTORM* performance on a low density DNA-PAINT movie. The ground truth **A** and *DeepSTORM* prediction **B** of α-tubulin (red, P1 imager strand) and TOM20 (blue, P5 imager strand) in MNTB tissue. Scale bar 5 µm.

5.1.3 *DeepSTORM* reconstruction of DNA-PAINT movies in high density tissue

Next, a *DeepSTORM* model was trained to handle images with high emitter density (∼2 emitters/µm^2) and applied to exchange DNA-PAINT movies of α-tubulin and TOM20 in MNTB tissue (fig. 33A).

For α-tubulin 5 nM P1 imager strand concentration with 400 frames and for TOM20 10 nM P5 imager strand concentration with 400 frames provided the highest similarity with the ground truth (fig. 34 and 35). The tested 20 nM imager strand concentration contained more background, which was disadvantageous for the predicted *DeepSTORM* image. Again, a slight decrease in resolution was observed, but within acceptable range. A low density movie was acquired in the same field of view for both targets, which served as the ground truth (fig. 33B). This acquisition of the ground truth data took 25 minutes per target protein, whereas the acquisition of the high density data only required 1 minute. Comparing the ground truth and predicted super-resolution images revealed an overall agreement in faithfully reconstructed structural features of the five cells in the tissue section, highlighted with dotted circles in fig. 33. The neural network effectively predicted TOM20 structures (fig. 33C, D) and 1D filamentous structures of α-tubulin (fig. 33 C, D i, ii). Despite slightly reduced performance in capturing the density of complex 2D arrangements of α-tubulin (fig. 33C, D iii–vi), the trained *DeepSTORM* model exhibited good quality in predicting the labeled targets in the tissue section.

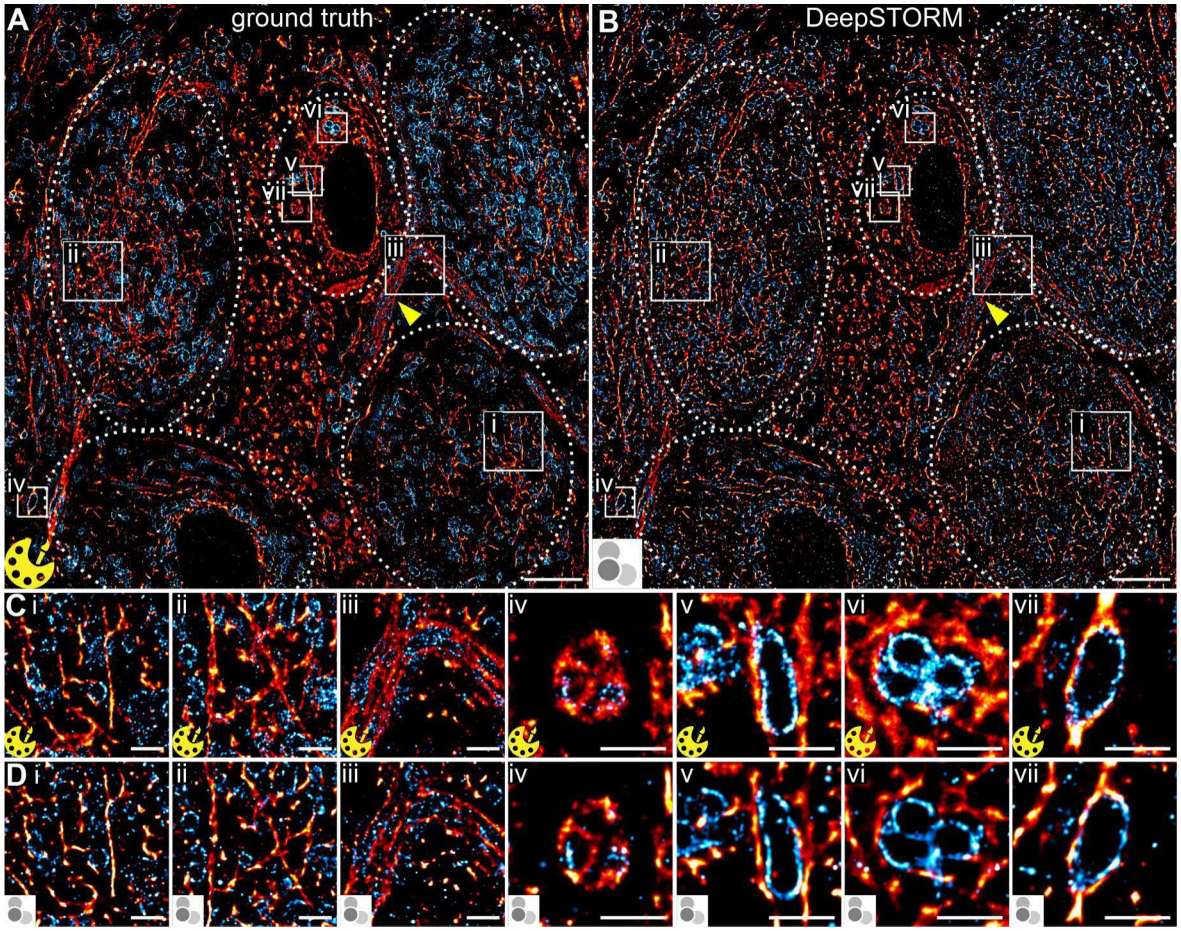

Figure 33: *DeepSTORM* performance on a high density exchange DNA-PAINT movie. Comparison between DNA-PAINT super-resolved image of **A** *DeepSTORM* and **B** ground truth image generated from low density acquisiton with *Picasso*. α-tubulin (red, P1 imager strand) and TOM20 (blue, P5 imager strand) were imaged. The shown MNTB tissue contains 5 cells, highlighted with dotted circles. Scale bars 5 µm. Magnified regions are shown in **C** and **D**. Scale bars 1 µm. Adapted from Narayanasamy et al., 2022, CC-BY-4.0.

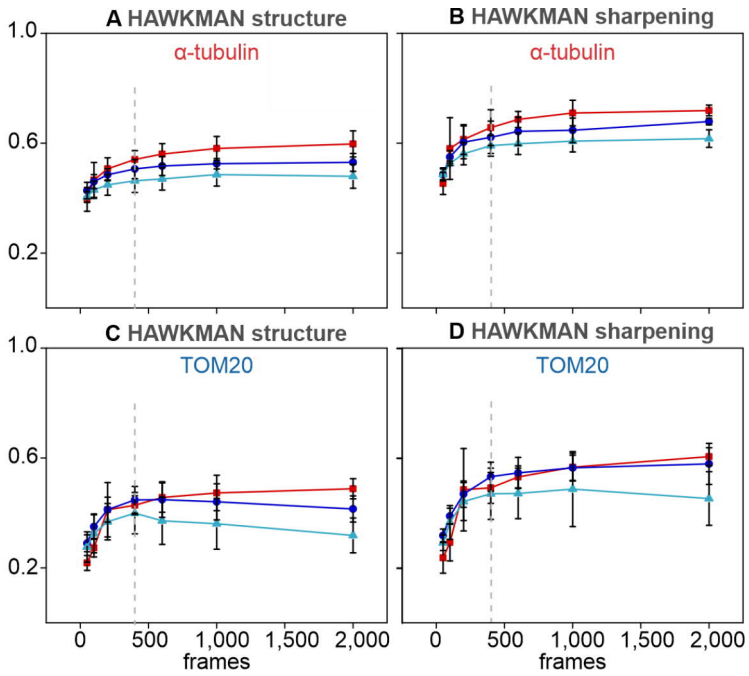

Figure 34: Number of frame optimization for complete image reconstruction with *DeepSTORM*. High density movies with 5 nM, 10 nM, or 20 nM imager strand concentrations were predicted with varying frame numbers. The predicted super-resolved image by *DeepSTORM* was compared to the ground truth per frame number by calculating HAWKMAN to determine the optimal frame number, which is sufficient to reconstruct all content in the image. Dashed lines indicate the optimal frame length at 400. N = 3, error bars = SD, adapted from Narayanasamy et al., 2022, CC-BY-4.0.

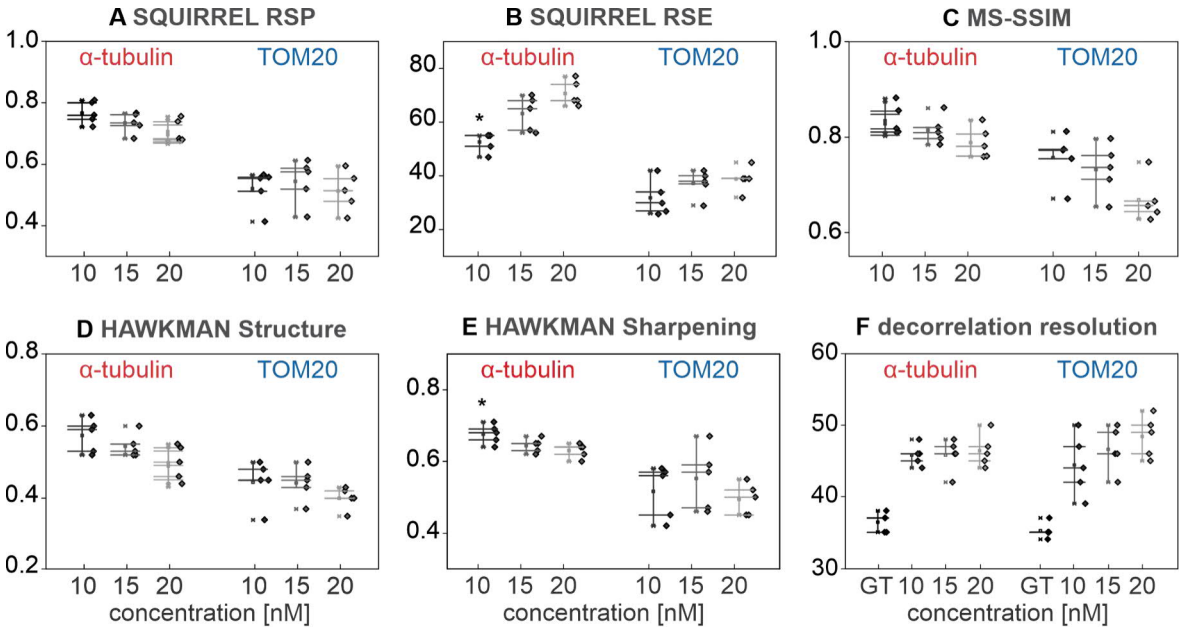

Figure 35: DeepSTORM predicted image of high density DNA-PAINT quality assessment against the ground truth. The similarity metrics include **A** SQUIRREL RSP, **B** SQUIRREL RSE, **C** MS-SSIM, **D** HAWKMAN structure, **E** HAWKMAN sharpening, and the **F** resolution. Asterisks indicate statistical significance at $p < 0.05$ using one-way ANOVA, N = 5 Adapted from Narayanasamy et al., 2022, CC-BY-4.0.

5.1.4 Retrieving emitter positions in postprocessing

In the case where emitter positions are additionally required besides the super-resolved image, *DeepSTORM* offers a postprocessing step where emitter positions are obtained based on the predicted stack of super-resolved images. In this algorithm, the brightness of a point needs to surpass a certain threshold and be of maximum intensity in a local neighborhood to qualify as a localization. Three parameters must be adapted: "threshold", "neighborhood size", and "local averaging". With local averaging activated, a center of gravity estimator is applied to the neighborhood, to potentially refine the emitter positions. The threshold was set to ensure a roughly equivalent number of localizations as the ground truth. The neighborhood size and local averaging were set to maximize image similarity when comparing the rendered *DeepSTORM* localization image to the ground truth. This resulted in a neighborhood size of 3 and local averaging being activated. The *DeepSTORM* super-resolved image and rendered image produced nearly the same image similarity metrics, suggesting that the method on how to determine the algorithms parameters works robustly on DNA-PAINT images (fig. 36).

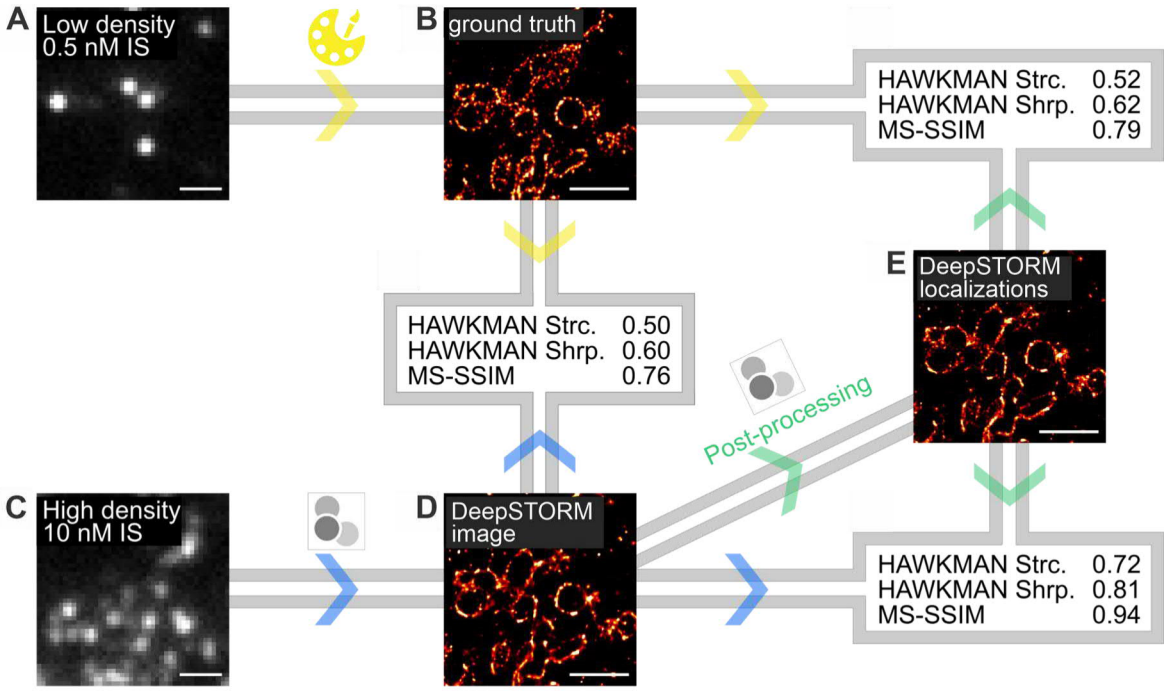

Figure 36: Comparison between *DeepSTORM*'s super-resolved image and rendered image from *Picasso*. A TOM20-labeled MNTB section was imaged with **A** low emitter density (0.5 nM, P5, 10,000 frames), from which **B** ground truth image was reconstructed using *Picasso*. **C** High density acquisition with 10 nM imager strand, which was used to **D** generate a super-resolved image with *DeepSTORM*. **E** With *DeepSTORM*'s postprocessing method, emitter positions were retrieved, which were rendered in *Picasso*. HAWKMAN and MS-SSIM were computed to determine image similarity between the ground truth image, *DeepSTORM* predicted image, and *DeepSTORM* postprocessed image. Scale bar 1 µm. Adapted from Narayanasamy et al., 2022, CC-BY-4.0.

In conclusion, *DeepSTORM* has been demonstrated on STORM data to obtain super-resolved images under the challenging condition of high emitter density (Nehme et al., 2018), outperforming other methods such as *ThunderSTORM* (Ovesny et al., 2014), *FALCON* (Min et al., 2015) and *CEL0* (Gazagnes et al., 2017). The here demonstrated application of *DeepSTORM* to DNA-PAINT measurements enhances its versatility. The constant emitter density in DNA-PAINT, contrary to STORM measurements, keeps the data in the optimal density performance window of the network. Moreover, DNA-PAINT enables the utilization of a single fluorophore across multiple targets, necessitating just one trained neural network as a digital enhancement to a microscope.

The workflow presented herein is rooted in experimental data, offering an alternative to simulation based training. With this work, *DeepSTORM* within the *ZeroCostDL4Mic* framework (Chamier et al., 2021) was extended to accept experimental data in a user friendly way. Both approaches, simulation and experiment, have their merits and drawbacks. High density movies based on experimental data closely resemble images obtained on the microscope, but might be more tedious to generate. Conversely, simulated data can be generated extensively, but may lack specific behaviors of the experimental setup. Adjusting the parameters of a simulation might be challenging and the direct acquisition of training data on the microscope might be the straight forward route. Furthermore, generating a dataset with many instances is easily achievable in this case, as small patches from large FOVs are randomly summed together, resulting in nearly endless patch combinations with only a few acquired frames.

For training data acquisition, choices must be made regarding the PAINT-strand, target, fluorophore, and setup. As it has been demonstrated here, the type of PAINT-strand and target does not affect the generalization capability of the network. The target structure is not visible in one patch of a single frame and the emitters appear to be randomly distributed, independently which structure is underneath. Furthermore, in one frame, the different kinetics of the PAINT-strand are not relevant, as signal must solely be emitted by some strands being bound. However, the fluorophore and setup influence the appearance of the PSF. If a model is trained with a different fluorophore or PSF, careful attention must be paid to quality control metrics, and retraining may be necessary if predictions are unsatisfactory.

Optimal imager strand concentration varies with structure type. Dense structures need lower concentrations than less dense ones. This has been shown that for α-tubulin 5 nM imager strand concentration was optimal and for TOM20 10 nM. *DeepSTORM* handles a range of emitter densities (5-10 nM) but deteriorates at 20 nM due to lower SNR and emitter overlap. Increasing the imager strand concentration will eventually cause a saturation of the binding sites, beyond which only background fluorescence rises. Furthermore, *DeepSTORM* predicted 2D shapes with lower precision than 1D structures. A next step in improving the capability of *DeepSTORM* in dealing with different structures and local

densities could be to estimate the local density and then choose a model that was trained on this density.
Moreover, the efficiency enhancement of deep-learning assisted imaging of only 1 minute per target protein compared to 25 minutes with conventional DNA-PAINT allows for improved statistical robustness, as more instances can be measured within a certain time. Fast imaging offers an additional advantage in capturing large views with super resolution as demonstrated in Narayanasamy et al., 2022. For this, smaller super-resolution views of α-tubulin in MNTB tissue were stitched together to create a 55x55 μm^2 super-resolved image within only 16 minutes. This facilitates information spanning from the microscale down to the nanoscale.
The user-friendly *ZeroCostDL4Mic* environment aims to make neural networks accessible without the need for substantial computational resources or programming skills. This environment offers simplicity, documentation, support, sample data, and access to decentralized computing resources through *Google Colab*. The here demonstrated workflow uses only freely available tools and the custom code offers a user interface and is well documented. Access to the trained model and data is given to allow reproducibility and reuse of the model. This lowers the barrier of using high-speed DNA-PAINT and SMLM in everyday super-resolution microscopy.
While *DeepSTORM* is designed for 2D data, adaptation to 3D is feasible using neural networks such as *DeepSTORM3D* (Nehme et al., 2020) and *DECODE* (Speiser et al., 2021). In *DeepSTORM3D* a phase mask is added to the optical path in the microscope, that alters the shape of the PSF depending on the z-depth. Using a CNN, a 3D volume is generated, from which 3D localizations can be retrieved. In *DECODE*, emitters are simulated based on a calibration measurement of a z-stack with beads and an astigmatism length in the light path encoding the z-information. Based on that measurement, training data is simulated and the network learns to localize 3D emitters.
The here demonstrated workflow was successfully applied to measurements using exchangeable HaloTag ligands (xHTLs) (Kompa et al., 2023), further extending the application capability of *DeepSTORM*. Jang et al., 2023 demonstrated a reduced acquisition time by ∼25 fold using xHTLs in high density movies of vimentin, TOM20, and calreticulin-KDEL, which were processed with *DeepSTORM*. The combination with *DBlink* (Saguy et al., 2023), a neural network generating continuous structures from patchy SMLM movies, enabled the observation of the endoplasmic reticulum in a live cell for 12 minutes with a temporal resolution of 1 second (Jang et al., 2023). This underscores the potential of synergistically combining different neural network tools to push the boundaries of super-resolution microscopy even further.

5.2 Methods

For a more detailed description of the methods refer to Narayanasamy et al., 2022.

5.2.1 Sample preparation

All animal experiments adhered to the applicable laws and institutional guidelines of Baden-Württemberg, Germany (protocol G-214/20), and received approval from the Regierungspräsidium Karlsruhe. Sprague-Dawley rats obtained from Charles River were anesthetized and transcardially perfused with PBS at postnatal day 13, followed by 4% paraformaldehyde (PFA) in PBS (Sigma-Aldrich). The brains were dissected and further fixed in 4% PFA overnight at 4 °C. 200 µm thick sections of the brainstem, containing the medial nucleus of the trapezoid body (MNTB), were prepared using a vibratome (SLICER HR2, Sigmann-Elektronik, Germany). The MNTB regions were excised, infiltrated with 2.1 M sucrose (Sigma-Aldrich) in 0.1 M cacodylate buffer at pH 7.4, and left overnight at 4 °C. The tissue was then mounted on a holder, plunge-frozen in liquid nitrogen in 2.1 M sucrose, and 350 nm semi-thin sections were cut using a cryo ultramicrotome (UC6, Leica). These sections were transferred to 35 mm glass bottom dishes (MatTek, USA) pre-coated with 30 µg/ml of fibronectin from human plasma (Sigma-Aldrich). Nanodiamonds (100 nm; Adamas Nanotechnologies, USA) were added as fiducials. The dishes containing sections were stored at 4 °C before use. For tissue staining, the thawed tissue sections were washed three times with PBS, with each incubation lasting 15 minutes, to remove sucrose droplets.

5.2.2 Labeling

Secondary antibodies of donkey anti-mouse (715-005-151), donkey anti-rabbit (711-005-152), and donkey anti-chicken (703-005-155), were purchased from Jackson ImmunoResearch. DNA strands with thiol or azide modifications on the 5' end for docking strands and a Cy3B dye on the 3' end for imager strands were used (Metabion). Conjugation of antibodies to thiol-DNA docking strands was prepared using a maleimide linker, while azide-DNA conjugation utilized the DBCO-sulfo-NHS ester linker (CLK-A124-10; Jena Bioscience). The antibody-DNA solutions were stored at 4 °C.
For tissue labeling, primary antibodies against α-tubulin-mouse (T6199, Sigma-Aldrich; clone DM1A; dilution 1:500) and TOM20-rabbit (sc-11415, Santa Cruz Biotechnology; dilution 1:80) were used. Tissue samples in dishes underwent a three-time, 10-minute wash with PBS to remove the sucrose-methylcellulose layer, followed by blocking with 5% foetal calf serum (FCS; Gibco) for 30 minutes. The primary antibodies, diluted in 0.5% FCS, were applied to the tissue section for 1 hour at room temperature and washed off three times with PBS. Subsequently, the secondary antibody-DNA docking strand (5.8 mg/mL stock;

dilution 1:100) in 0.5% FCS was applied onto the tissue for 1 hour at room temperature, followed by three washes with PBS.

5.2.3 Setup

DNA-PAINT microscopy was conducted using a home-built SMLM setup featuring an Olympus IX81 inverted microscope frame and an Olympus 150× TIRF oil immersion objective (UIS2, 1.49NA). Illumination was carried out in HILO mode with a 561 nm laser line (Coherent Sapphire LP) at 0.88 kW/cm2, using a 4 L TIRF filter (TRF89902-EM, Chroma Technology) and an ET605/70 M bandpass filter (Chroma Technology). An Andor iXon EM+ DU-897 EMCCD camera (Andor, Ireland) collected signal with a pixel size of 107 nm.

5.2.4 DNA-PAINT imaging

DNA-PAINT imaging was performed in Buffer C (2.5 M NaCl; S7653, Sigma-Aldrich in 5x PBS; 14200-059, Gibco Fisher Scientific) supplemented with 1 mM ethylenediaminetetraacetic acid (EDTA; E6758, Sigma-Aldrich), 2.5 mM 3,4-dihydroxybenzoic acid (PCA; 03930590, Sigma-Aldrich), 10 nM protocatechuate 3,4-dioxygenase pseudomonas (PCD; P8279, Sigma-Aldrich), and 1 mM (±)-6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (Trolox; 238813-5G, Sigma-Aldrich). Oxygen scavenging buffers PCA and PCD were used to reduce site-loss labeling due to DNA docking strand damage by ROS42. To acquire training data, 20 pM P5 imager strands were imaged in TOM20-labeled tissue samples for 5,000 frames with an acquisition rate of 150 ms. For imaging the ground truth, P1 (α-tubulin) and P5 (TOM20) strands were recorded with a concentration of 0.5 nM for 10,000 frames with an acquisition rate of 150 ms. High density emitter DNA-PAINT datasets for *DeepSTORM* prediction were obtained by imaging with concentrations of 5 nM, 10 nM, and 20 nM for 400 frames. Exchange-PAINT was performed manually by washing five times with 1x PBS. Five different FOVs were imaged.
Ground truth DNA-PAINT movies were processed with *Picasso* (v0.2.8) (Schnitzbauer et al., 2017). Events in each frame were localized with integrated Gaussian and maximum likelihood estimation. The localizations were drift corrected and rendered using 'one pixel blur'. Localizations were linked to merge localisations appearing in multiple consecutive frames. Movies from the same FOV with different concentrations were aligned by a linear transformation using nanodiamonds as a registration reference. Super-resolution large-sample imaging on α-tubulin was performed using DNA-PAINT imaging with 10 nM P1 imager strands. For the large-scale imaging of α-tubulin (10 nM, P1), a grid-like acquisition strategy of 4x4 was employed for each imaging area for 400 frames, with imately 10% overlap between adjacent images. Image stiching was performed using *Inkscape* based on structural similarity.

5.2.5 Generating training patches

For *DeepSTORM* model training, high density emitter images with matching emitter position are required. Due to the challenging nature of achieving precise localizations at high concentrations, low density emitter frames were summed to artificially generate high density emitter frames, along with a list of emitter positions. This process was executed using a custom script (*ImageSumming* version 220306, https://github.com/HeilemannLab/ImageSumming). For the low density data, 5000 sparse emitter frames (20 pM, P5, TOM20) were acquired with a density of 0.028 emitters/μm^2 and a size of 512x512 px^2. *Picasso* was used for localizing emitters. The imaged frames were randomly cropped to 17x17 px^2 and summed, resulting in a high density movie of 1.9 emitters/μm^2 with 30,000 frames and matching localization list. Patches devoid of emitters were excluded from the summation process. As the summing of n patches induces n camera offsets, the offset was corrected by subtracting its value n-1 times. The camera offset was estimated through the average pixel value in a measurement conducted with a closed shutter.

5.2.6 DeepSTORM training and prediction

DeepSTORM model training was performed in *Google Colab* using the *ZeroCostDL4Mic* (Chamier et al., 2021) notebook (v1.12). The assigned resources entailed a Tesla V100 GPU with *CUDA* version 11.0 and *Tensorflow* version 2.5.0. Training for the SEM model took 17 min on 1782 instances and for the MEM model 35 min on 30,000 instances. To directly use the summed image patches as input, some parameters must be adjusted, as the notebook is constructed to expect images, which are patched within the notebook (number of patches per frame = 1, patch size = 16 px, maximum number of patches = 30,000, minimum number of emitters per patch = 1). The parameters for training were set to number of epochs = 100, batch size = 256, percentage of validation = 15 %, initial learning rate = 1E-5. Default values were used if not stated otherwise. One *DeepSTORM* model was trained on the P5 TOM20 dataset and used for all high density emitter images. The trained model and data is available at https://zenodo.org/records/6966132. For prediction, 400 frames with a size of 512x512 px^2 were processed, which took 7-25 minutes. *DeepSTORM* outputs images, which must be localized with a built in postprocessing step to obtain emitter positions. This entails the adaption of three parameters, which were set per imaged target to achieve high image similarity between ground truth and rendered image from predicted localizations. The neighborhood size was always set to 3 and local averaging was activated. The threshold was adapted per target.

5.2.7 Image analysis

Picasso-rendered ground truth and *DeepSTORM* predicted super-resolution images were compared using different control metrics in *Fiji*. The resolution was determined using decorrelation analysis (Descloux et al., 2019) with default values. For MS-SSIM analysis (Wang et al., 2003), the images were intensity-normalized. For the SQUIRREL analysis (Culley et al., 2018), the magnification factor was set to 1. For the HAWKMAN analysis (Marsh et al., 2021), the images were converted to 8 bit depth. The PSF FWHM in super-resolution pixels was set to 23 nm, the PSF of the instrument to 300 nm, the reconstruction magnification factor to 8, and 10 scales were analyzed. A pixel scale of 69 nm was finally chosen, corresponding to the resolution of the predicted images.
To determine a sufficient number of frames of the high density movies, movie lengths of 50, 100, 200, 400, 600, 1000, and 2000 frames predicted with *DeepSTORM* to generate a super-resolution image. Per frame length, this image was compared to the ground truth. With increasing number of frames, the similarity improved, but saturates for large numbers of frames. The number of frames before saturation was chosen. This analysis was conducted per measured target.

6 Fast and long-time STED imaging of ER nano-structural dynamics in living cells

Live cell imaging is essential for investigating dynamic biological processes by providing valuable insights into the spatiotemporal behavior of cellular events. STED microscopy has served as a powerful tool to study processes with high resolution, including dynamic processes in living cells (Stockhammer et al., 2021). However, the high laser intensities of this technique lead to fluorophore photobleaching and sample phototoxicity, which limits the number of recordable fluorescence images for a given sample.

There are various approaches to tackle these challenges. Stable fluorophores extend the STED measurement window to several minutes (Liu et al., 2022b). Additionally, transient fluorophores prevent photobleaching, allowing for longer measurements (Glogger et al., 2022; Liu et al., 2024; Spahn et al., 2019). Although stable and transient fluorophores can circumvent photobleaching, the high laser power remains the same. Event-triggered STED microscopy monitors subcellular events in less photodamaging imaging modes, such as wide-field microscopy, and switches to STED once the event of interest is detected (Alvelid et al., 2022). This is applicable for investigating short scenarios with the high resolution of STED and the photon budget is reduced. Monitoring long-term scenarios with high resolution remains a challenge.

Further efforts to mitigate photobleaching and phototoxicity encompass denoising, which has been vastly applied to several fluorescence microscopy techniques (Jin et al., 2020; Krull et al., 2018; Weigert et al., 2018). When denoising super-resolved images, maintaining the resolution can be challenging, as convolutional neural networks (CNNs) often introduce blurring due to the commonly employed pixel-level loss function (Kim et al., 2019a). Switching to a perceptual loss or an adversarial loss might improve the denoising performance. The introduction of attention layers to the network also prevents blurring, as the flow of high-frequency information is enhanced (Vaswani et al., 2017; Zhang et al., 2018b).

Ebrahimi et al., 2023 demonstrated improved performance in denoising STED images while maintaining the resolution, by combining a U-Net (Ronneberger et al., 2015) and RCAN (Zhang et al., 2018b) architecture with attention layers. Different cell organelles were imaged with a pixel dwell time of 1 µs for minutes and imaging quality restored with the network.

In this work, the UNet-RCAN network (Ebrahimi et al., 2023) is utilized to denoise STED images of the endoplasmic reticulum (ER) recorded with ultra-low irradiation intensity. This enables the observation of the ER for several hours with high spatiotemporal resolution. Drug treatment with torin1 (Kim et al., 2015) and bafilomycin A1 (Mauvezin et al., 2015) initiate early ER autophagy in cells, leading to morphological changes in the ER. So far,

ER-phagy has only been studied with super-resolution in fixed cells (Berkane et al., 2023; González et al., 2023). To quantify the dynamical changes, sheets and tubes were segmented using *ERnet* (Lu et al., 2023). *ERnet* is a neural network based segmentation tool of the ER, which has been applied to SIM images. Its translation to STED images is demonstrated here without the need for model re-training. Additional descriptors were applied to the segmented structures, providing a quantitative analysis of the drug effects.

6.1 Results and discussion

6.1.1 Denoising pipeline for live cell ER STED images

For denoising, the UNet-RCAN network was chosen, as it preserves most of the resolution of the ground truth STED images compared to other methods (Ebrahimi et al., 2023; Rahm et al., 2024). UNet-RCAN is a supervised neural network, thereby requiring matching pairs of low-intensity (noisy) and high-intensity (ground truth) images. The noisy-ground truth image pairs were acquired with the acquisition switching between lines of the noisy and ground truth image to mitigate any spatial offset between the image pairs, which would dampen the quality of predictions. The rapid acquisition of noisy images at exceptionally high speeds (0.5 µs dwell time per pixel) minimized discernible spatial shifts, while any residual offsets were drift corrected using the registration method of *NanoJ* (Laine et al., 2019) (fig. 37A). The acquired noisy-ground truth image pairs were then cropped into smaller patches (128^2, 256^2 or 304^2 px^2) and served as training data for the network (fig. 37B). Cropping has the advantage of generating numerous training instances with only a few measurements. Furthermore, processing the entire recorded image is not possible due to limited RAM size on the GPU. To enhance the resilience of the network to potential variations in samples across measurement days, inputs stemmed from at least three measurement days.
Initial attempts to train the UNet-RCAN model using default parameters proved unsuccessful. The model did not converge and a large loss was observed, indicating gradient explosion and rendering the model not useful. Despite adjustments of hyperparameters, this issue persisted, necessitating the introduction of additional hyperparameters which were included in the code base of UNet-RCAN. An exploration of more than a hundred parameter combinations ensured the identification of a model that converged (fig. S1) and performed adequately (fig. 37C).
Model performance was probed with a test dataset, which was acquired on a separate measurement day (fig. 37D). The generated predictions exhibited normalized pixel intensities, which were rescaled to match the pixel distribution of the ground truth data (fig. 37E). The rescaling procedure exclusively leveraged information from the training data, preventing any influence from the test dataset. Finally, quality control metrics were computed on the rescaled predictions and ground truth data, facilitating the identification of an optimal model within the screened parameter space (fig. 37F).

The importance of the hyperparameters was investigated with a random forest regressor, that ranks the effects of the hyperparameters to model performance (fig. S2). The two most important hyperparameters for model training were an appropriate learning rate and the size of the input. A patch size of 304x304 px^2 was the optimal choice for this dataset (the size must be divisible by 8 due to the three down-sampling layers in the U-Net),

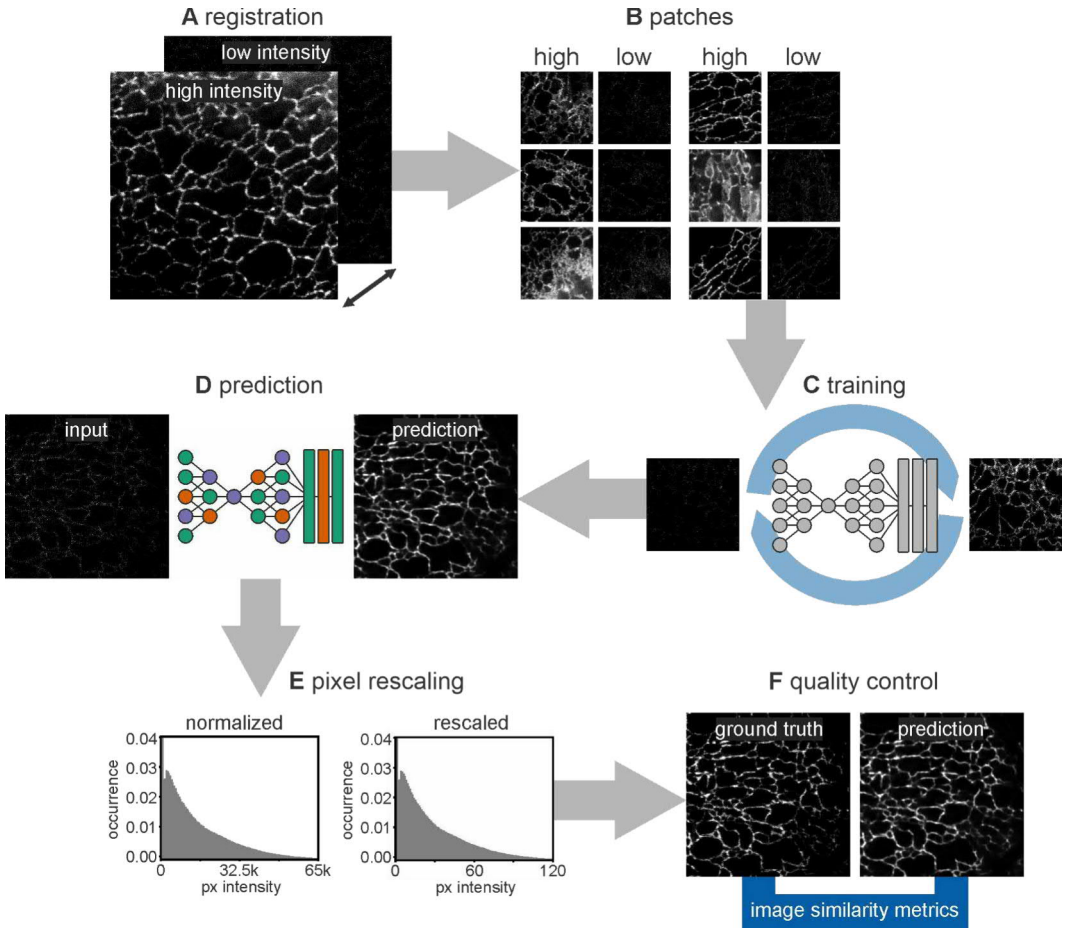

Figure 37: STED denoising workflow. A Ground truth and the corresponding noisy images are aligned in the xy plane, **B** from which small patches are generated. **C** A UNet-RCAN model is trained, wherein the patches from noisy images function as input and the rest of the patches serve as the ground truth. An exploration of various hyperparameters is undertaken to identify a suitable model. **D** The trained model is employed for prediction. **E** The normalized predictions rescaled to align with the pixel distribution of the ground truth. **F** Quality control metrics are then computed on a designated test dataset. Scale bar 1 µm, N = 15 (Rahm et al., 2024).

while larger patch sizes were limited by memory restrictions. Given the pivotal role of this parameter, its optimization is recommended in conjunction with considerations of image content size, filter size, and downsampling extent. The importance of different parameter values within a parameter group are listed in SI table S1 and S2. This approach gives an overall indication of appropriate parameters. However, the best performing models across the hyperparameter search had slightly different combinations (SI table S3).

6.1.2 Quality control of STED image restoration

To visualize the ER in living cells, U2-OS cells expressing calreticulin-KDEL as an ER marker were used. The fluorophore SiR, which was attached to the target structure using a covalent HaloTag, was used as a label. The ground truth is imaged in a small region (12x12 µm^2) to prevent movement artifacts between the subsequent line acquisitions (fig. 38A). The noisy images appear spotted and contain only a few different low pixel intensity (fig. 38B).

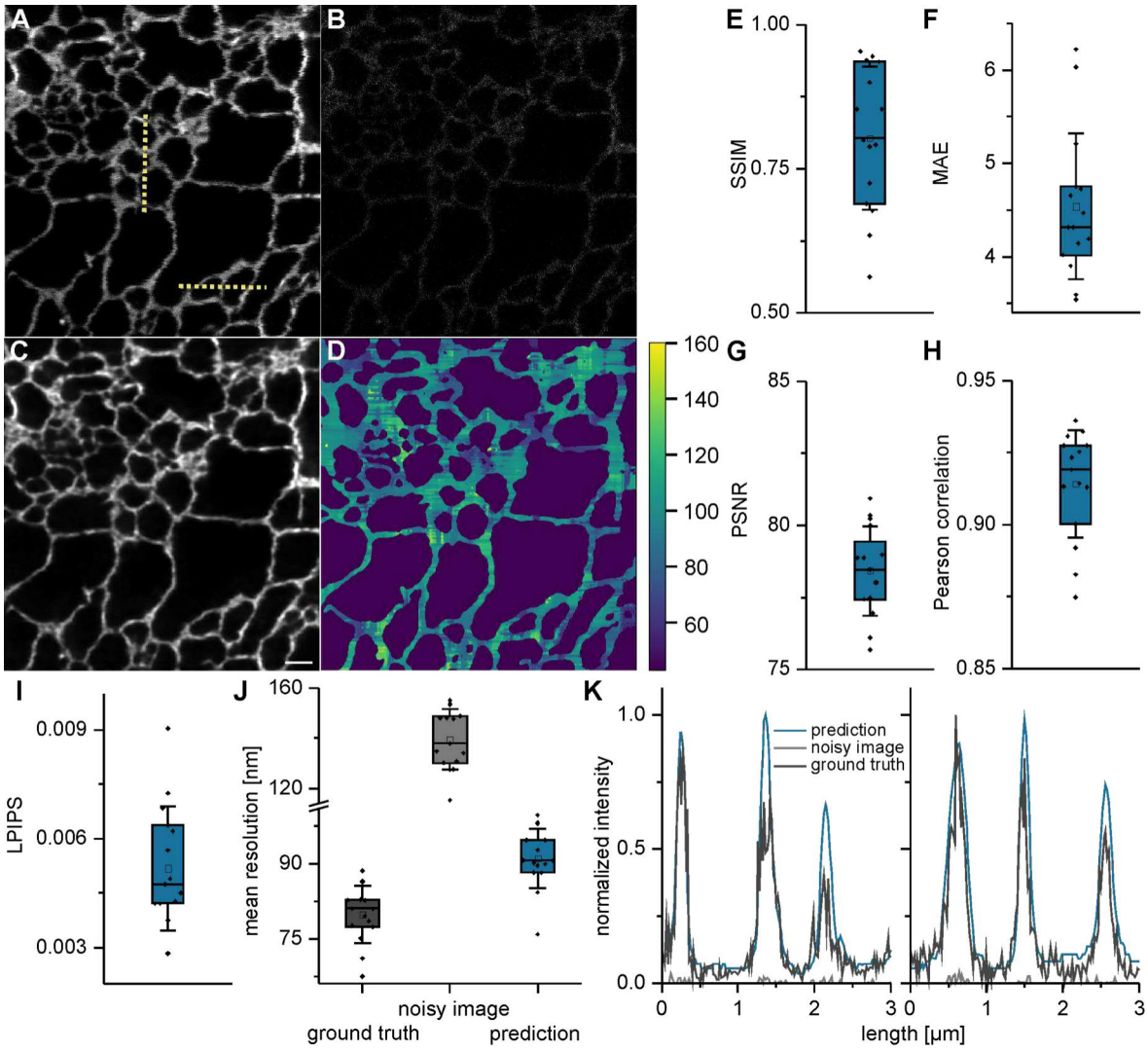

Figure 38: STED denoising of ER in living cells. **A** Ground truth STED image of the ER of a U2-OS cell. **B** Noisy STED image of the ER acquired in parallel with the ground truth in line acquisition mode. **C** Predicted image of **B** with a trained UNet-RCAN network. **D** rFRC resolution map of the predicted image. **E** SSIM, **F** MAE, **G** PEARSON correlation, **H** PSNR, and **I** LPIPS quality control metrics comparing ground truth and predicted images. **J** Mean resolution of ground truth, noisy, and predicted image calculated using rFRC. **K** Intensity line profiles of ground truth, noisy and predicted image, the intensities were normalized globally. Scale bar 1 µm, N = 15 (Rahm et al., 2024).

The resultant prediction of neural network restoration recovers the structural details (fig. 38C). A rolling FOURIER ring correlation (rFRC) map (Zhao et al., 2023) of the predicted image shows how the resolution varies spatially, with resolutions up to 42 nm and an average resolution of 89 nm (fig. 38D). The corresponding ground truth image has an average resolution of 82 nm. The average resolutions were 90 $\pm$ 6 nm for the predicted and 80 $\pm$ 6 nm for the ground truth images. All predicted images were further scrutinized using various quality control metrics (fig. 38E-I). The MAE focuses on the pixel-wise absolute difference between two images. The PSNR evaluates the quality of a reconstructed signal, by comparing it to the original, uncompressed signal. The PEARSON correlation evaluates the relationship of pixel intensities. Contrary to these metrics, SSIM (Wang et al., 2004) and LPIPS (Zhang et al., 2018a) consider spatial pixel dependencies. SSIM evaluates the similarity of the images in the aspects of structure, contrast, and luminescence. In LPIPS, feature representations extracted from pretrained neural networks are compared between prediction and ground truth images. This mix of metrics, along with the resolution, is a robust toolbox for comprehensive image quality assessment.
The QC metrics showed that the predictions recover the structural details of the ground truth image. In addition, the predictions maintained the resolution although not to full extend compared to the ground truth images (fig. 38J). Intensity profiles across the image also indicate that the prediction retained the underlying details of the ground truth image and the relative intensities (fig. 38K).

6.1.3 Effects of morphological changes on model robustness

To study ER-phagy dynamics, cells were treated with torin1 and bafilomycin A1. Torin1 induces stress and autophagy in cells (Kim et al., 2015). The cell degrades its organelles and proteins through the enclosure in autophagosomes, a spherical structure with a double layered membrane (Vinod et al., 2015). For further degradation, the autophagosome fuses with lysosomes. By adding bafilomycin A1, the fusion is blocked and the autophagy process stopped at autophagosome formation (Mauvezin et al., 2015).

As the induction of ER-phagy causes morphological changes in the ER structure, the performance of a model trained with images from both resting and induced cells (mixed model) was compared with a model solely trained on resting cells. The mixed model performed well with the matching test data from autophagy induced cells and the resting model with the resting cell test data. The performance was probed in terms of structure recovery (fig. 39A-C), resolution (fig. 39D), intensity fluctuations (fig. 39E) and QC metrics (fig. S3 and S4). Furthermore, test data of resting cells were predicted with the mixed model and test data of autophagy-induced cells with the resting model. The resting model performed slightly better im terms of QC metrics compared to the mixed model for the resting cell test dataset. For the autophagy dataset, no significant improvement was observed when

using the mixed model compared to the resting model. This shows that the data from both conditions was within the generalization capability of the resting model. For further analysis, the resting model was applied.

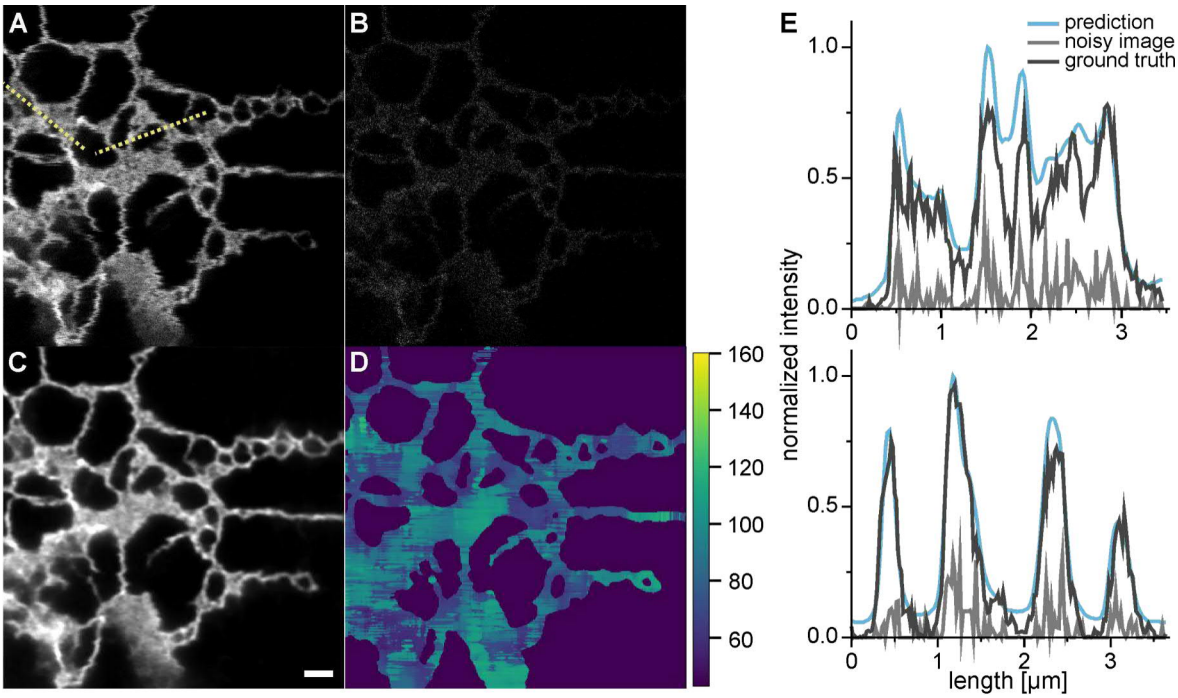

Figure 39: STED denoising of living cells induced for stress. **A** Representative ground truth and **B** noisy STED image of the ER in a stress-induced live cell and **C** its corresponding prediction. **D** rFRC resolution map of the predicted image pair. **E** Intensity line profiles of ground truth, noisy and predicted image at two highlighted regions of interest. The intensities of all three datasets were normalized together as a single group. Scale bar 1 µm (Rahm et al., 2024).

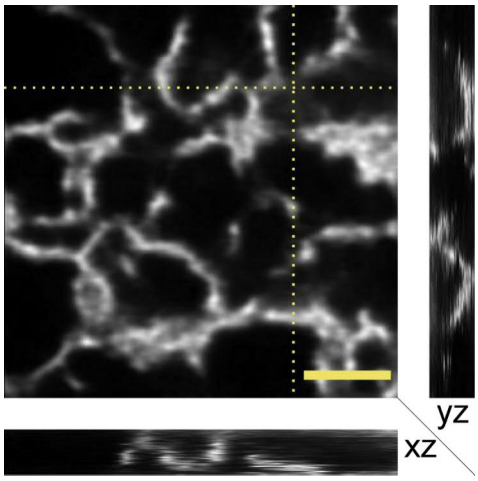

Figure 40: Prediction of a 3D stack. Prediction of the resting model for an input containing low-intensity volumetric STED images of the ER (calreticulin-KDEL). An exemplary xy plane is visualized and the yz and xz planes corresponding to the yellow dotted lines are shown on the left and bottom respectively. Scale bar 1 µm (Rahm et al., 2024).

As an additional query for the robustness, the resting model was used to predict noisy images of 3D stacks acquired from living resting cells. A top-hat PSF pattern was used for the depletion laser to gain axial resolution. Fig. 40 shows a predicted 3D stack of an exemplary cell visualized as xy, yz and xz planes. Since only noisy images were measured in the 3D stack, additional single plane noisy and ground truth images were acquired with the same settings including a top-hat depletion laser PSF. The 3D test dataset was evaluated and showed that the resting model does indeed denoise images acquired with a top-hat depletion laser PSF (fig. S5).

6.1.4 Strategies to discern structural signals from hallucinations

To ensure the quality of predictions, it is crucial that the low-intensity input contains signal that is not completely bleached away. In the absence of signal in the FOV, hallucinations occur (fig. 41A, B). To distinguish hallucinated predictions from those stemming from structure, the similarity of adjacent frames in a movie can be probed. Moving organelles exhibit a degree of similarity between adjacent frames, resulting in higher structural similarity when compared to adjacent predictions derived from empty FOVs (fig. 41C). By tracking the similarity over time, the authenticity of predictions can be validated.
Areas of uncertain predictions might not only arise temporally, but also spatially; for instance, when a minute signal was collected from an out of focus plane, which is not sufficient for a robust prediction (fig. 41D, E). By thresholding the pixel intensities of the low-intensity image, the pixels can be roughly grouped into background, uncertain, and signal pixels. The grouping is not perfect, but enough to gear the attention to areas with potential artifacts. Fig. 41F contains a hallucinated structure and an area with dim signal predicted from areas labeled as "uncertain" in the low-intensity image. This method is also an alternative for tracking the similarity over time, in cases where large structural changes occur frame to frame due to the acquisition time or mobility of the structure.

6.1.5 Denoising and segmentation to describe fast and long-term ER dynamics

The ER-dynamics under drug treated condition (torin1 and bafilomycin A1) was imaged for four hours with a temporal resolution of 4.6 seconds per image. Imaging time was restricted by the cell moving out of focus. The low-intensity signal remained constant during that time and would have allowed for even longer imaging. To quantify morphological changes, the denoising pipeline was coupled to a neural network based segmentation pipeline, *ERnet*, which segments the structure and distinguishes between sheets and tubes (Lu et al., 2023) (fig. 42). The intensity of the input images had to be unified as a preprocessing step to enhance the continuity of the segmentation. With torin1 and bafilomycin A1 treatment, the aggregation of ER fragments in autophagosomes is expected. The ER structure changed

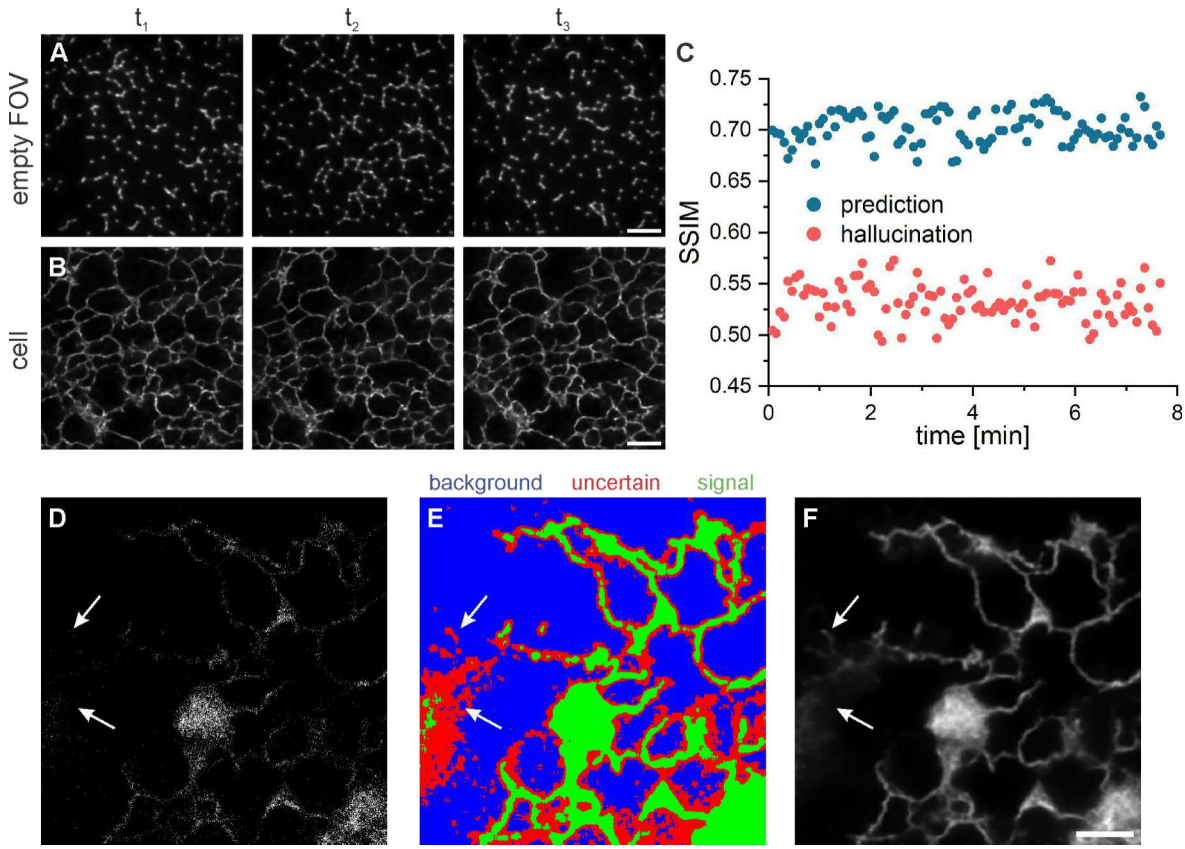

Figure 41: Hallucination artifacts. **A** Predictions of a time series of background signals, which contain hallucinations. **B** Predicted images of a time series of ER in living cells. **C** To distinguish both cases, the SSIM for adjacent frames is calculated. The hallucinated structures appear randomly due to the random nature of noise and have a lower SSIM than a movie containing signal of interest. **D** Low-intensity input from treated movie condition. **E** Classified pixels as background, uncertain, and signal. **F** Predicted image, partly predicting structure, where hardly any signal was collected in the low-intensity image. The arrows highlight two areas of uncertainty. Scale bar 2 µm (Rahm et al., 2024).

the most within the first hour, where the density of the ER decreased and small sheet regions accumulated to larger regions.

Besides the segmentation, *ERnet* additionally provides descriptors of the segmented structure, including the quantification of pixels belonging to tubes or sheets per image and a description of the tube network via graph theory. For the treated movie, the fraction of tubes decreased within the first hour, with the amount of pixels segmented as sheets surpassing the number of pixels segmented as tubes. An exception occured at approximately two hours, where sheets move out of the field of view, predominantly revealing tubes.
The degree of nodes, representing the number of connections per node, was predominantly 1 followed by 3. A degree of 1 corresponds to the ends of tubular connections, while a degree of 3 indicates nodes within the network structure. Interestingly, the number of degree 1 nodes rose within the first hour, reflecting an increased discontinuity of the network.
Clustering denotes grouping of nodes and edges, with values ranging from 0 indicating no

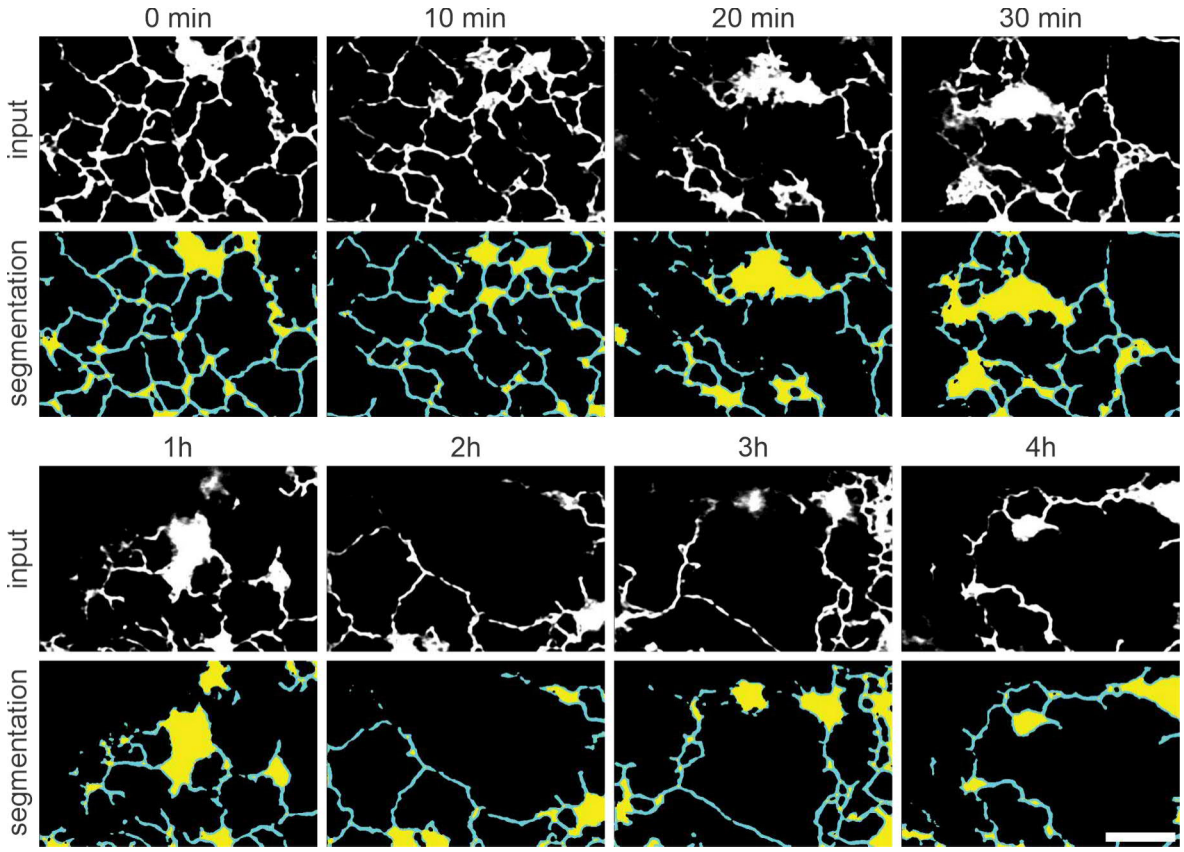

Figure 42: Development of ER structure after drug treatment. Input and segmentation results of *ERnet*. Tubes are depicted in blue and sheets in yellow. Scale bar 3 µm (Rahm et al., 2024).

clustering to 1 indicating complete clustering. The clustering decreased within the first hour and stays at very low numbers, suggesting that clustering is not a prominent feature of the ER network. The assortativity, which is the tendency of nodes to connect with nodes of the same degree, fluctuated around 0, meaning that no clear tendency to connect to similar or dissimilar nodes is observed. Clustering and assortativity values around zero are in alignment with the values reported in Lu et al., 2023.

Additional descriptors were explored for further quantification. The iou of segmented adjacent frames informs about the similarity of the segmentation and therefore the speed of rearrangement (fig. 44A). The iou fluctuated, but had an apparent increase in similarity within the first hour. This is reflected in the segmented images, where small sheet areas aggregated within this time window. Compared to a cell measured in resting condition (fig. 41C), the iou was smaller, meaning that drug treatment induces more pronounced morphological changes.

The overall density of ER structure decreased, evident from the reduction in the number of foreground pixels (fig. 44B). Additionally, sheet instances were quantified by connecting regions with the same pixel values. This allowed the counting of sheets per frame, revealing a decrease in sheets (fig. 44C). The observations are in alignment with the expected ag-

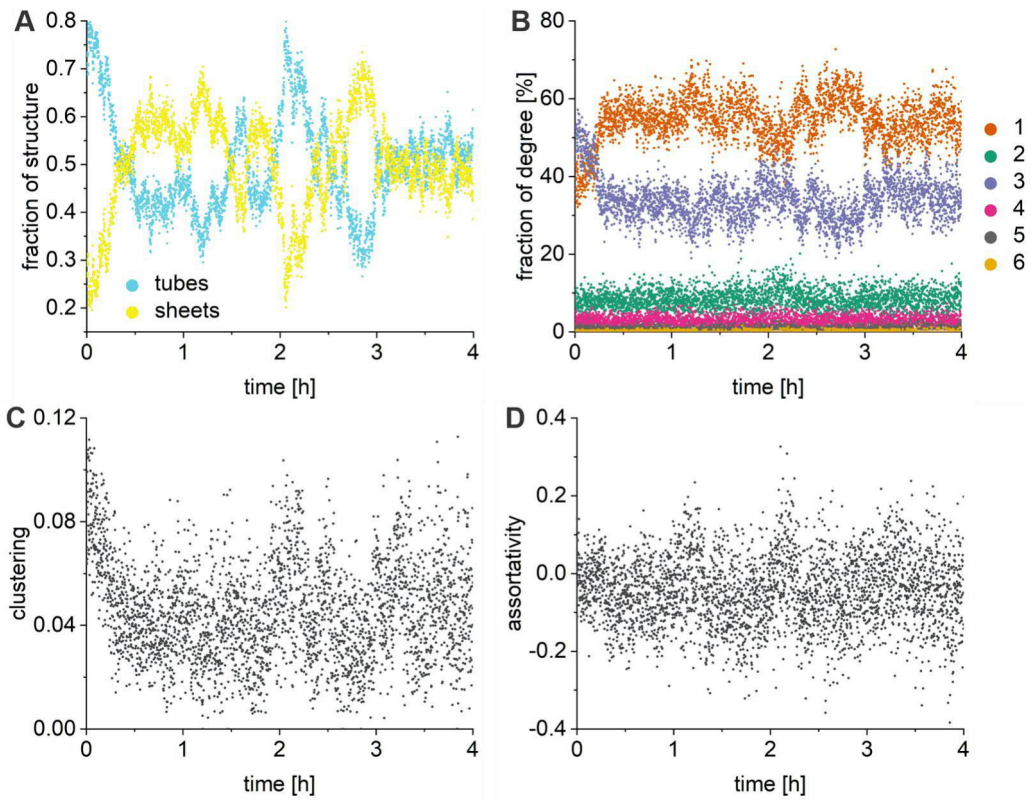

Figure 43: *ERnet* structure quantification. **A** Fraction of tubes and sheets. **B** Fraction of the node connectivity degree. **C** Clustering of the tube network. **D** Assortativity of the tube network (Rahm et al., 2024).

gregation of ER fragments in autophagosomes, with most effects of the drug taking place within the first hour.

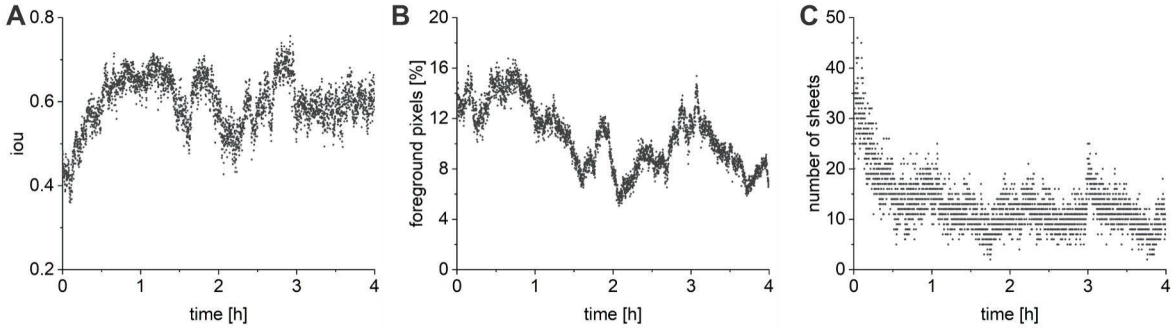

Figure 44: Additional descriptors for dynamic ER structure quantification. **A** Intersection over union between adjacent segmented frames. **B** Foreground pixel ratio. **C** Number of sheets per frame (Rahm et al., 2024).

The here demonstrated pipeline shows the potential of denoising STED images to record live cell organelle dynamics with high spatiotemporal resolution for a long period of time. The observation of the ER structure for several hours with a resolution of 90 nm and 0.5 µs pixel dwell time pushes STED to previously unknown limits. Other studies that denoised STED images used a pixel dwell time of 1 µs pixel dwell for imaging for several minutes

(Ebrahimi et al., 2023). Hajiabadi et al., 2022 denoised RNA polymerase II clusters in live zebrafish embryos and acquired 3D stacks within seconds. As the biological process of interest is on the scale of a few seconds, no long-time measurements were explored. Furthermore, the nucleus does not contain apparent high frequency features and it remains unclear how well the used N2V network (Krull et al., 2018) restores the resolution. The ER as a study object requires both fast and long-time measurements, as the structure is highly dynamic and drug effects take place on a larger time scale (González et al., 2023). The here achieved spatiotemporal resolution with an observation window of hours fulfills these demands.

As calreticulin-KDEL is assosiated with ER-phagy, functioning as a quality control chaperone for protein folding and known to stimulate the formation of autophagosomes (Yang et al., 2018), the macro level effects during ER-stress were of interest. Sheet accumulation and and overall decrease in structure density was mostly apparent within the first hour, although it should be stated that more measurements and comparisons to a mock condition would be required to draw conclusions. The decrease of structure density is in alignment with other known responses to torin1 and bafilomycin A1 treatment. An increased cluster size of FAM134B, a protein involved in ER reshaping among others, was observed after 6 hours of ligand treatment in fixed U2-OS cells measured with DNA-PAINT (Berkane et al., 2023).

Imaging live cell data with high quality is challenging due to the phototoxicity of light and the movement of the organelle (Wäldchen et al., 2015). Organelle movement induces movement artifacts, visible as horizontal lines in the images. As the ER dynamics are surprisingly fast, movement artifacts could not be fully avoided when acquiring the training and test ground truth data. Mapping low-intensity live cell images to high-intensity fixed cell images could mitigate this issue. The fixed cell images should be cautiously evaluated with regards to fixation artifacts. The same marker compatible for both the live cell and fixed cell condition could be used. To further improve the quality of the target image, deconvolution could be applied as a preprocessing step to sharpen the ground truth images. The elimination of background signal and non-specific binding could be achieved through a neural network based technique known as label2label (Kölln et al., 2022). This method involves training a network to convert images with two distinct labels that target the same cellular structure. Fluorescence signals that exhibit variation in the images are excluded, thereby restoring correlated structural signals.

To enhance predictive accuracy and mitigate the risk of structural hallucination, multiple frames could be employed as input to improve the accuracy of the model. Using temporal resolution to improve prediction quality has been exemplified by other works, for example in the multi-emitter fitting model *DECODE*, where adjacent frames are incorporated as input to derive localizations for the central frame (Speiser et al., 2021). It is imperative to fine-tune the quantity of added frames, as the inclusion of frames exhibiting significant structural alterations may not assist in improving results.

The translation of images from confocal to STED microscopy offers an alternative method for facilitating rapid and long time observations of live cell structure. However, Ebrahimi et al., 2023 demonstrated superior prediction quality of denoising compared to cross-modality translation, particularly in terms of regained resolution. Bouchard et al., 2023 show that the accuracy of confocal to STED translation can be improved by adding a specific task during training, for example the segmentation of structure. However, this approach requires manually annotated data for the additional task. Cross-modality translation is an interesting choice for extending the capabilities of conventional light microscopes. If a STED microscope is available, denoising will likely preserve the resolution more effectively, as the low-intensity input contains more information about the high-frequency features of the structure compared to a diffraction limited image.

All in all, this work demonstrates a neural network based workflow accessing the fast ER dynamics with long time observation at high resolution. This robust framework could be further applied to different scenarios, where fast organelle dynamics with super-resolution are of interest. Future work might include the extension of describing other cellular structures along the ER, for example studying the contact sites between the ER and mitochondria (Sassano et al., 2022). This would draw a more holistic understanding of organelle dynamics and interactions in living cells.

6.2 Methods

6.2.1 Sample preparation

U2-OS cells that stably express calreticulin-KDEL-HaloTag7 were seeded onto 8-well chambered coverglass at an amount of $1 \cdot 10^4$ cells per well. The cells were incubated overnight at 37 °C and 5% CO_2 and were induced the next day with 250-500 µg/mL doxycycline (Sigma Aldrich, Germany) in culture media. Two days post-induction, the cells were incubated for 15-20 min with 300 nM of HaloTag ligand conjugated to the fluorophore SiR diluted in culture media followed by 3x washing with culture media. The cells were incubated in culture media at 37 °C with 5% CO_2 until taken out for imaging, using live-cell media (reduced culture media with HEPES, Thermo Fisher Scientific, Germany) as an imaging buffer. 250 nM torin1 and 250 nM bafilomycin A1 were added to the live-cell media prior imaging for the treated condition.

6.2.2 Image acquisition

STED imaging was performed on an Abberior expert line microscope (Abberior Instruments, Germany) with an Olympus IX83 body (Olympus Deutschland GmbH, Germany). Imaging was conducted with a UPLXAPO 60x NA 1.42 oil immersion objective (Olympus Deutschland GmbH, Germany). An excitation laser of 640 nm and a depletion laser of 775 nm laser was used. The PSF shape was either a donut (for planar and long-term imaging) or a top-hat (for volumetric imaging). Fluorescence was collected in the spectral range of 650 nm to 760 nm using an avalanche photo diode (APD). For ground truth imaging, the excitation laser power was set to 56.35 µW with 6 line accumulations. Low-intensity images were conducted with 4.83 µW excitation laser power and 1 line accumulation. The depletion laser power was 166.02 mW, the pinhole diameter 0.81 a.u., the pixel dwell time 0.5 µs, and the pixel size 20 nm for all conditions.

6.2.3 Image processing and dataset generation

High-intensity and low-intensity images were aligned with the registration method by *NanoJ-Core* (Laine et al., 2019). For the resting condition, training images were acquired from 3 measurement days. To generate a dataset suitable for both the treated and resting condition, 26% of images were added from the treated condition. The same amount of data was used for training to allow comparison between the resting and mixed model performance. Images were cropped into patches of 128x128 px^2, 200x200 px^2, 256x256 px^2, and 304x304 px^2 to determine an optimal patch size of the training data. A training and validation split of 90%-10% was made. Per condition, 15 test images of size 600x600 px^2 were acquired at a different measurement day.

6.2.4 Network training

UNet-RCAN networks (Ebrahimi et al., 2023) were trained on a NVIDIA RTX 3090 24 GB GPU. As the default hyper parameters led to exploding gradients, some additional options were added to the published codebase, and the parameter space was explored to find a robustly working model. This included different loss functions (leaky ReLU, tanh), kernel initialization strategies (glorot uniform, lecun uniform, orthogonal), gradient clipping (0.1, 0.01, 0.001), and l2 regularization (0, 0.01, 0.001). Different learning rates were also tested during the search (0.001, 0.0001, 0.00001 for the resting and 0.001, 0.0001, 0.0005, 0.00005, 0.00001, 0.000001 for the mixed condition), along with different patch sizes (128x128 pixels, 200x200 pixels, 256x256 pixels, and 304x304 pixels). The batch size was chosen as the maximum possible number per patch size before an out-of-memory error occurred on the GPU (patch size: batch size, 128: 36, 200: 16, 256: 10, 304: 6).
A parameter space search was executed to identify possible parameter combinations for models that perform well on the test dataset. To diminish the search time, 1000 data instances were included and training was interrupted after 10 epochs or if the loss exceeded a value of 1000. 147 models were evaluated for the resting condition and 393 models for the mixed condition. Quality control metrics were calculated on a test dataset after every model run to determine the model performance. The metrics SSIM (Wang et al., 2004), MAE, and the spatial resolution (determined with decorrelation analysis (Descloux et al., 2019) in NanoPyx v0.2.2 (Saraiva et al., 2023)) were each linearly transformed between 0-1 across all model runs per condition. A "total score" was computed by summing the transformed scores and dividing them by three to assess performance across multiple metrics. The top 5 models, based on the highest total score, were trained until convergence on the full dataset with 1764 instances.
For the final calculation of quality control metrics, the pixel intensities of the predicted test images must be rescaled, as the model outputs normalized intensities. A scale and offset were calculated based on the pixel values of predicted training instances and their ground truth images. The test dataset is deliberately excluded from this process. A linear scaling operation was applied to the pixel values of the predicted images. Values above the 99th quantile and below the 1st quantile were clipped. The scale and offset values were determined per model and applied to the test data before calculating quality control metrics. The final model was selected based on the highest total score, with parameters detailed in SI table S3.

6.2.5 Quality control metrics

The MAE was calculated using the mean absolute difference of pixel values. SSIM (Wang et al., 2004) and PSNR were computed with the scikit-image library v0.19.3 (Walt et al., 2014). Pearson correlation was determined with SciPy v1.9.3 (Virtanen et al., 2020). The learned perceptual image patch similarity (LPIPS) (Zhang et al., 2018a) was ascertained

with a VGG model. Resolution during model training was determined with decorrelation analysis (Descloux et al., 2019) in NanoPyx v0.2.2 (Saraiva et al., 2023). Resolution for the test data was determined using rolling FOURIER ring correlation v0.2.5 (Zhao et al., 2023) with a 1/7-threshold and a background intensity of 20.
To discern predictions stemming from FOVs with signal and FOVs containing only noise, the SSIM was calculated between adjacent predicted frames of the treated movie and a movie recorded of only dark noise of the detector (in an empty image plane without sample). To spatially group pixels into the categories background, uncertain, and signal, a sliding window of size 10x10 px^2 was used to sum the pixel intensities within this window per pixel. Borders of the image were padded with reflected pixels to maintain the input size. A pixel was labeled as background, if the sum was equal or below 5, as uncertain if in between 6 and 25, and as signal if above 25. The window size and thresholds must be adapted per use case.

6.2.6 Hyperparameter importance

The RandomForestRegressor from scikit-learn v1.0.2 (Pedregosa et al. 2012) was used to determine hyperparameter importance based on the 147 model runs for the resting and 393 model runs for the mixed condition. The approach ranks the hyperparameters based on their importance in achieving a high "total score". Furthermore, suitabilities are assigned to the different parameter options within a parameter group.

6.2.7 *ERnet*

The intensity of the images were unified by clipping values below 30 to 0 and above 40 to 40 prior to *ERnet* processing. The pretrained 20220306_ER_4class_swinir_nch1.pth *ERnet* v2.0 model was used in Google Colab (Lu et al., 2023). The segmentation outputs the structures tubules, sheets, and sheet-based tubules (SBTs). As the latter was not visible in the movies and *ERnet* just segmented some minor spots as SBTs, these spots were merged into the sheet segmentation. To segment sheet instances, regions with the same pixel values were connected by assuming 8-neighborhood connectivity.

6.2.8 Statistics

As some distributions rejected the test for normality (SHAPIRO-WILK test, $\alpha = 0.05$), MANN-WHITNEY-U tests were used to compare the metrics of ground truth and predicted images. Levels of significance were defined as: $p > 0.05$ no significant difference (n.s.), $p < 0.05$ significant difference (*), $p < 0.01$ very significant difference (**), $p < 0.001$ highly significant difference (***). Datasets of all treatment groups were probed for normality by applying SHAPIRO-WILK tests ($\alpha = 0.05$). All tests were performed in Origin Pro 2018b.

7 Virtual staining of cellular landmarks

The artificial generation of fluorescent labels based on the information of a label-free channel is termed "virtual staining" and is typically achieved using deep learning approaches. Compared to traditional fluorescent staining, artificial staining can be observed over a long period of time, as no photobleaching occurs. Multiple cellular landmarks can be detected at once and temporal and spectral bandwidth is saved for other fluorophores and their targets of interest.
Previous studies have used the bright-field channel or qualitative phase information as model inputs for fluorescence prediction (Christiansen et al., 2018; Ounkomol et al., 2018). However, cells absorb light weakly and therefore produce little contrast in the bright-field images (Zuo et al., 2020). In addition, the contrast in the focal plane is minimal, reducing available information for the model to learn a transformation between label-free and fluorescence channel. Jo *et al.* showed that three-dimensional refractive index tomograms acquired with a holographic microscope, a quantitative phase imaging (QPI) method, improved the predictions for multiple organelles and cell types compared to using bright-field or qualitative phase images as model inputs (Jo et al., 2021). QPI therefore serves as a better contrast mechanism for cross-modality inference.

In the first chapter, a deep learning model is trained to virtually stain the nucleus and membrane. The inputs are label-free quantitative phase images, reconstructed from a stack of defocused bright-field images, acquired with a commercial bright-field microscope (Soto et al., 2017). The robustness of the model performance for the nucleus channel is probed under different experimental conditions, a step towards a generalist model that is applicable for different microscopes and cell types. The nucleus and membrane were chosen as targets, as the analysis of high throughput imaging often commences by identifying nuclei and cell boundaries (Liu et al., 2023). Furthermore, segmentation of nucleus in the predicted fluorescence images is demonstrated using cellpose (Stringer et al., 2021).
The second chapter probes the application of virtual staining to recover the structure of the endoplasmic reticulum and draws a comparison to denoising (see chapter 6), which uses low-intensity fluorescence instead of the label-free images as input.

7.1 Robust virtual staining for cell instance segmentation

7.1.1 Results and discussion

7.1.1.1 Virtual staining and nucleus segmentation workflow

The workflow of image processing for phase reconstruction and deconvolution with waveOrder (Yeh et al., 2021), model training for virtual staining with *VisCy* (Guo et al., 2020), and instance segmentation of the nucleus channel with cellpose (Stringer et al., 2021) is visualized in fig. 45.

Before using the bright-field images and matching nucleus and membrane fluorescence images as model inputs, several preprocessing steps must be applied. These include parameter optimization for the phase conversion and fluorescence channel deconvolution, registration of the label-free and fluorescence channels in x, y and z, and tiling of the images to make them suitable training inputs (SI chapter 13.2.1). For retrieving the quantitative phase information, a 3D stack of bright-field images must be acquired. The following results concentrate on the focal plane of the stack and use a 2D U-Net neural network architecture (Ronneberger et al., 2015) for virtual staining.

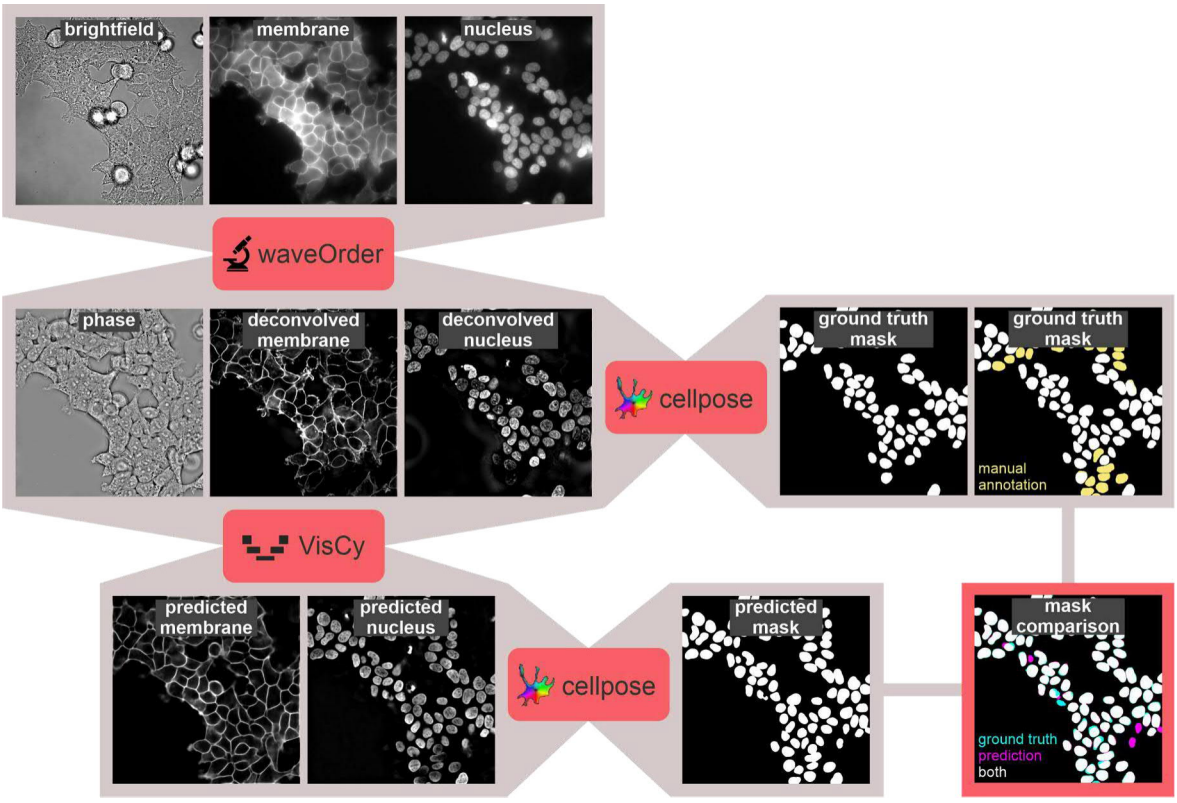

Figure 45: Virtual staining and segmentation workflow. The bright-field channel is converted to phase and the fluorescence membrane and nucleus channels are deconvolved using waveOrder. The phase and deconvolved fluorescence channels serve as input-target pairs for training a model for virtual staining with *VisCy*. Masks of the predicted nuclei are segmented using cellpose. Ground truth masks of the nuclei are created by applying cellpose to the ground truth fluorescence image and manually refining the mask (yellow regions in the ground truth mask image). The predicted and refined ground truth masks are compared to determine segmentation quality and virtual staining model performance.

Virtually stained images show are more homogeneous fluorescence signal compared to their ground truth images (fig. 46A, B vs E, F). Signals that are weak or absent in the ground truth fluorescence images due to the polyclonal labeling are recovered in the predictions. The phase images and their refractive indices may not contain information about the fluorescence intensity, yielding an averaged intensity across all observed organelles. This averaged prediction is beneficial for the task of segmentation, as otherwise cell instances would have been missed.

Most segmentation models specialize in specific cell types and experimental conditions (Ivanov et al., 2023). Recently, generalist models for segmenting diverse cell types have been developed, but adapting them to label-free datasets demands extensive human annotation. With virtual staining, cell segmentation models like cellpose (Stringer et al., 2021) are readily applicable to the predicted fluorescence images and the necessity for human annotations is eliminated.

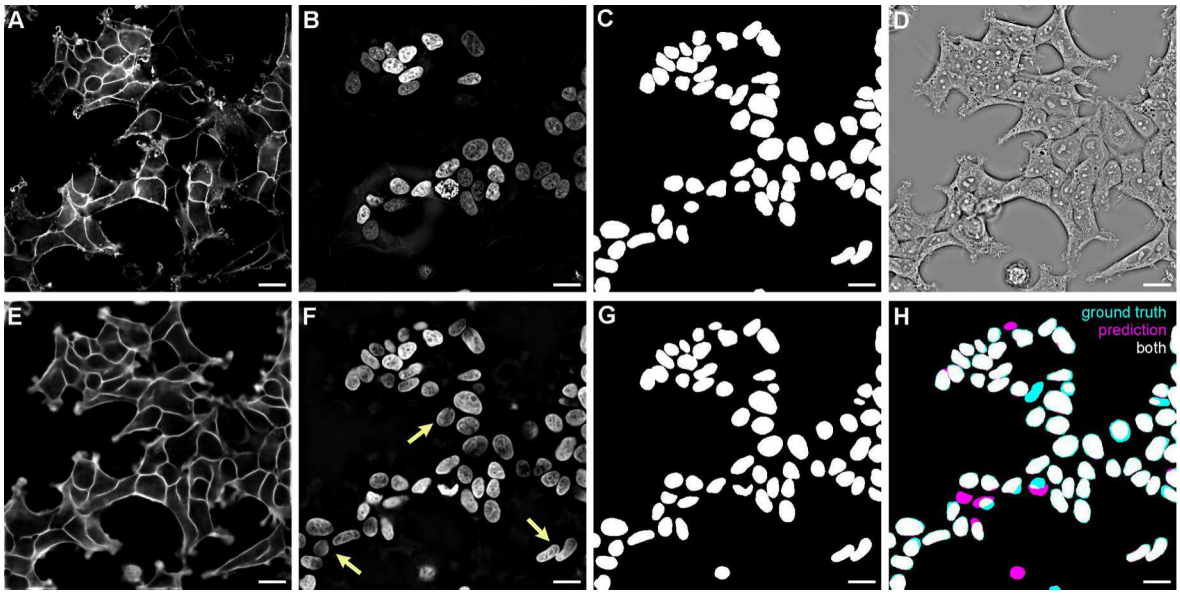

Figure 46: Predictions of fluorescence signal of cell membrane and nucleus with nucleus segmentation. **A** Deconvolved membrane fluorescence. **B** Deconvolved nucleus fluorescence. **C** Ground truth nucleus masks based on segmentations of the deconvolved nucleus fluorescence image with cellpose plus manual annotation. **E** Predicted membrane fluorescence. **F** Predicted nucleus fluorescence (yellow arrows show regions of recovered fluorescence signal). **G** Predicted nucleus segmentation with cellpose. **H** Overlay of ground truth mask (cyan) and predicted mask (magenta) with white showing overlap of the masks. Scale bars 20 µm.

The property of averaged fluorescence in the predictions leads to a hurdle in evaluating the model performance. The ground truth and predicted images can not be directly compared with image similarity metrics because some signal is missing in the ground truth images. Therefore, nucleus segmentation masks were created with cellpose based on the ground truth fluorescence channel and missing nuclei were drawn in by hand based on information from the phase image (fig. 46C, D). Even though parts of the labeling had to be done

manually, the utilization of the ground truth fluorescence channel for segmentation lowered the labor of creating ground truth masks drastically. The virtually stained images were segmented using cellpose with minor retraining of the nucleus segmentation model and quality control metrics were calculated based on the segmented masks (fig. 46G, H).

7.1.1.2 Generalizability towards different experimental conditions

The trained model for virtual staining was applied to validation data of the same experiment and to test data from a different measurement day, where the robustness was examined towards different experimental conditions and data augmentation strategies (Liu et al., 2023):

- generating phase with varying regularization parameters (training data: $1 \cdot 10^{-3}$, test data: $1 \cdot 10^{-4}$, $5 \cdot 10^{-4}$, $5 \cdot 10^{-3}$, $1 \cdot 10^{-2}$)
- different magnifications (training data: 63x, test data: 40x or 100x)
- different condenser apertures (training data: opening of 17, test data: opening of 12 or 6)
- different bright-field exposure times (training data: 10 ms, test data: 5 ms)
- different cell densities
- imaging at the border of the well
- different levels of augmentation
- no augmentation of zoom / no augmentation of intensity
- changed modalities of input and target (bright-field to non-deconvolved fluorescence channel, bright-field to deconvolved fluorescence channel, phase to non-deconvolved fluorescence channel, and phase to deconvolved fluorescence channel)

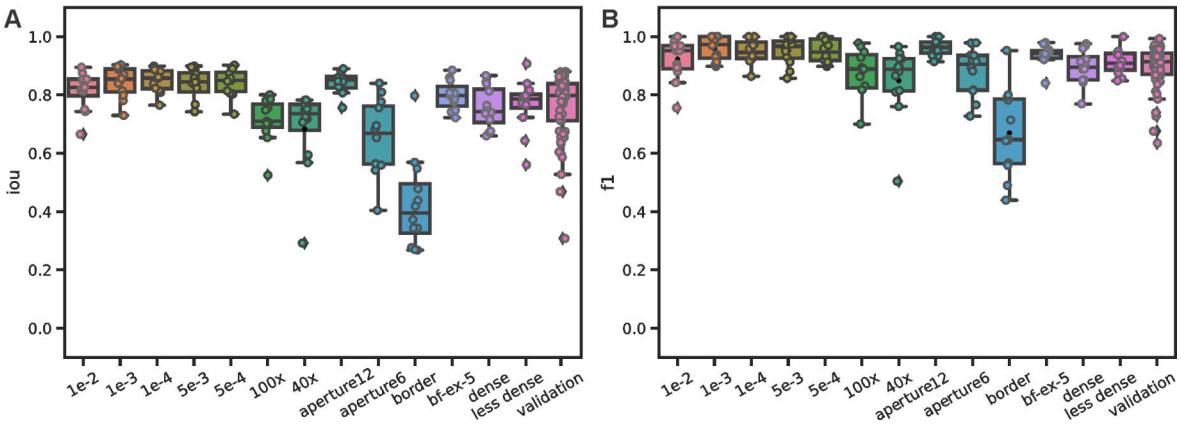

Figure 47: Generalizability towards different experimental conditions. A Intersection over union and **B** f1-score of different regularizations parameters for the phase ($1 \cdot 10^{-2}$, $1 \cdot 10^{-3}$, $1 \cdot 10^{-4}$, $5 \cdot 10^{-3}$, $5 \cdot 10^{-4}$), different magnifications (100x, 40x), condenser aperture openings (12, 6), imaging at the border of the well, half exposure time, FOVs with dense cells, FOVs with less dense cells, and the validation data. The model was trained with MSE loss, a z-range of 12-74, "moderate augmentation", and membrane and nucleus as target channels, N = 12.

Fig. 47 illustrates the model performance across the different experimental conditions of the test dataset as well as the validation dataset as a reference. The iou and the f1-score were chosen as metrics. The reported scores are calculated based on the similarity of ground truth and predicted nucleus segmentations. The iou measures the overlap between predicted and ground truth segmented pixels. The f1-score reports on the capability of detecting segmented instances. If nucleus segmentations are completely missed or falsely appear in the predicted image without a matching instance in the ground truth, the score drops.

The model generalizes robustly towards varying regularization parameters for phase reconstruction and cell densities. The model also generalizes towards different magnifications, but performs worse compared to the validation dataset (iou(40x) $= 0.68 \pm 0.15$, iou(100x) $= 0.71 \pm 0.08$, iou(val) $= 0.76 \pm 0.12$). The robustness towards varying magnifications could be improved by rescaling the test images to the known magnification of the training dataset, instead of covering for varying magnifications by augmentation. Other critical conditions were an extreme aperture closing of 6 (iou(ap_6) $= 0.65 \pm 0.14$) and measuring at the border of the well (iou(border) $= 0.43 \pm 0.16$). These conditions may have affected the image intensities beyond the range of the augmentation process. As cell confluency is not covered by augmentation, the robustness towards different confluence levels could be further enhanced by pooling data from multiple experiments with varying cell densities.

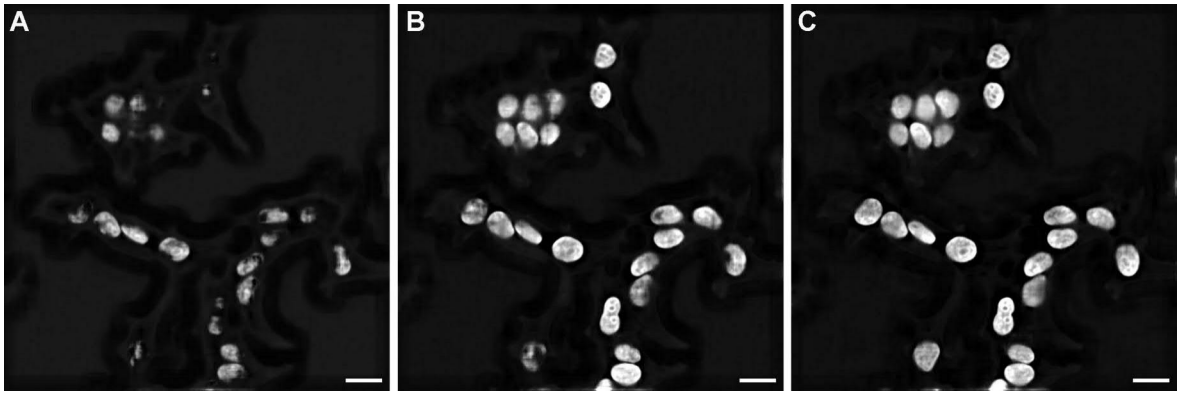

Figure 48: Prediction results of different augmentation strengths. Predictions of the apeture 6 test dataset with **A** "light augmentation", **B** "moderate augmentation", and **C** "strong augmentation". Scale bars 20 µm.

To evaluate the impact of data augmentation, three different levels of augmentation were tested: "light", "moderate", and "strong" augmentation, with a ∼linear increase in augmentation strength (tab. 2, fig. S9). The model with "strong augmentation" slightly outperformed "moderate augmentation", followed by "light augmentation", particularly for the aperture of 6 and the border of the well test images. Fig. 48 shows examples of predictions from the aperture of 6 test dataset with different augmentation levels. "Light

augmentation" generates incomplete nucleus shapes, whereas "moderate augmentation" and "strong augmentation" recover the nucleus shapes more completely. The decreased metrics primarily result from inaccurate predictions and are not due to the cellpose segmentation performance. These results suggest that increased augmentation strength benefits generalization. However, the performance gap between "moderate augmentation" and "strong augmentation" is smaller than between "moderate augmentation" and "light augmentation", indicating that the benefit may saturate if the augmentations cover the sample space.

Disabled augmentation of intensity and zoom

Intensity augmentation is crucial for the model generalization capability as many experimental conditions alter the recorded intensity, such as the exposure time of the bright-field image, different openings of the condenser aperture, and imaging at the border of the well (fig. 49). In addition, a model should be robust across different magnifications so that it does not have to be retrained for each individual magnification setting. The model should be invariant to intensity alterations (same goes for different noise levels). Therefore, the augmentation is solely applied to the input images. In the case of magnification, the model should be equivariant to size and the augmentation is applied to both the input and target image pairs.

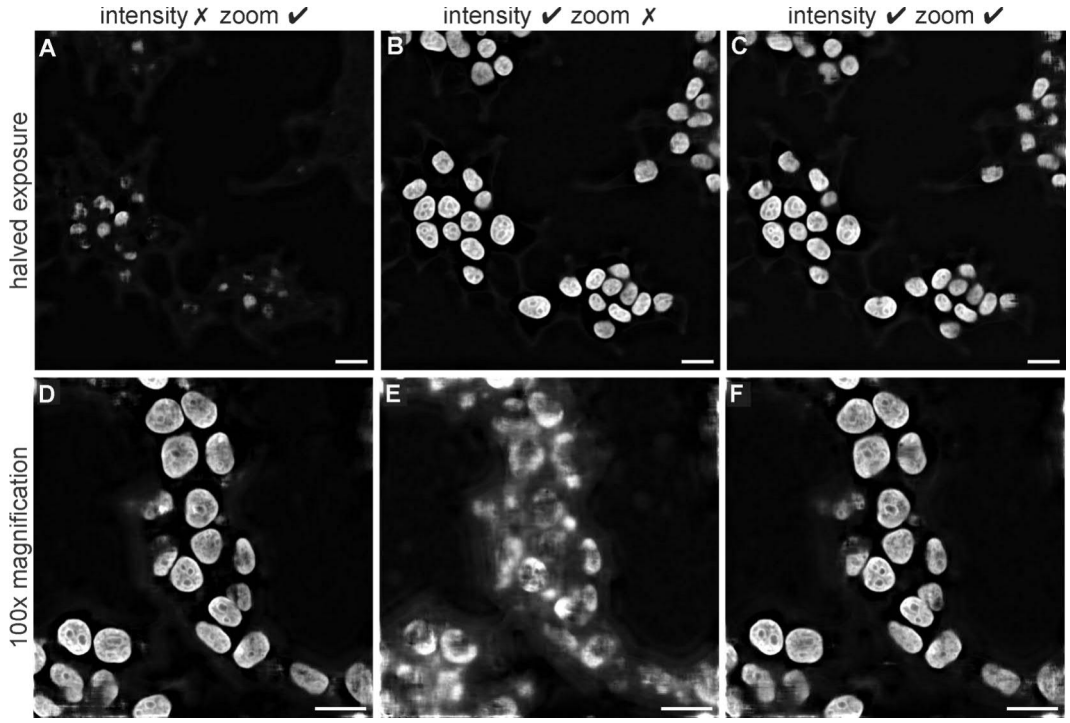

Figure 49: Prediction results of augmentation without intensity or zoom. Predictions of a FOV with halved bright-field exposure time of a model with **A** "moderate augmentation" with disabled intensity augmentation, **B** "moderate augmentation" with disabled zoom augmentation, **C** "moderate augmentation". Predictions of an 100x FOV of a model with **D** "moderate augmentation" with disabled intensity augmentation, **E** "moderate augmentation" with disabled zoom augmentation, **F** "moderate augmentation". Scale bars 20 µm.

A model with disabled intensity augmentation does not recover the fluorescence signal of images with halved exposure time (fig. S10A). With activated intensity augmentation the nucleus fluorescence is recovered, regardless of whether augmentation of zoom is applied or not (fig. S10B, C). Images with deviating magnification are recovered by a model with disabled intensity and enabled zoom augmentation (fig. S10D). Augmentation without zoom prevents generalization towards different magnifications (fig. S10E). Hence, it is crucial to incorporate augmentation techniques that cover the range of intensities and other deviations present in the datasets to which the model will be applied to (Möckl et al., 2020).

7.1.1.3 Changed input and output modalities

A series of experiments involving different combinations of input and target modalities were conducted, including bright-field to deconvolved fluorescence, phase to non-deconvolved fluorescence, bright-field to non-deconvolved fluorescence, and phase to deconvolved fluorescence (fig. 50, fig. S12). The performance of nucleus segmentation was notably enhanced when using the deconvolved fluorescence channel as the target modality. Using phase images as the input did not yield significant improvement. The nucleus is an organelle which is well visible in the bright-field channel. The phase and other descriptors like the retardance (optical anisotropy) (Guo et al., 2020) might be more beneficial for organelles that are less visible in the bright-field channel, such as the membrane.

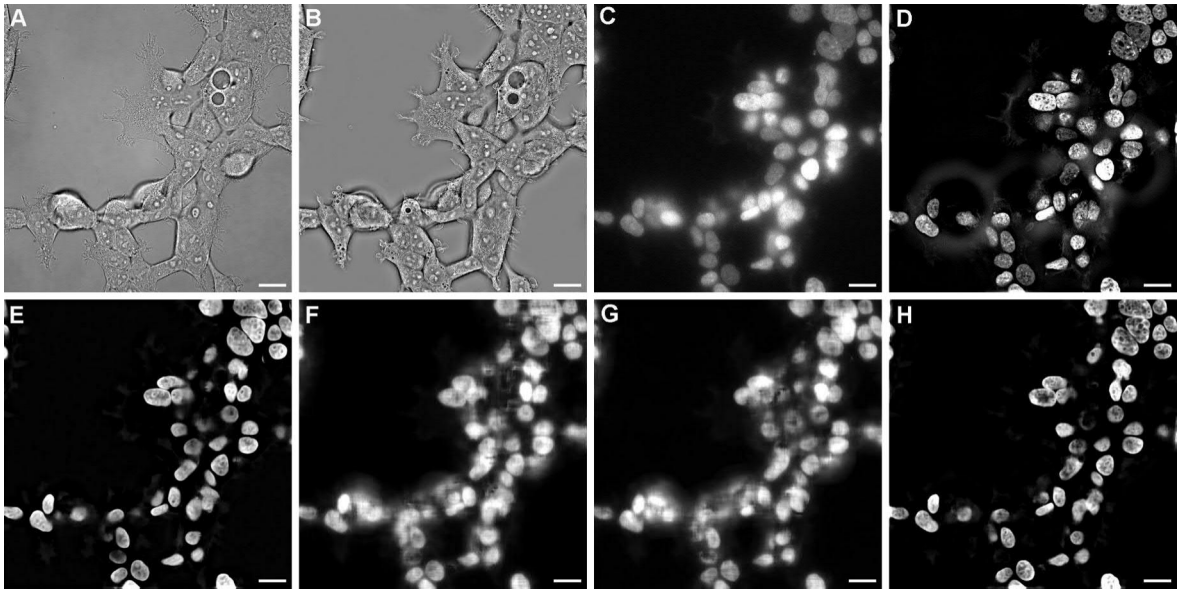

Figure 50: Predictions of different imaging modalities. **A** Bright-field, **B** phase, **C** non-deconvolved nucleus fluorescence, **D** deconvolved nucleus fluorescence, **E** prediction of bright-field → deconvolved fluorescence, **F** prediction of phase → non-deconvolved fluorescence, **G** prediction of bright-field → non-deconvolved fluorescence, **H** phase → deconvolved fluorescence. Scale bars 20 µm.

In summary, investigations have been conducted to evaluate the performance of virtual staining under various experimental conditions. Augmentation that coverages the sample space is a key element for recovering the fluorescence signal. Furthermore, predicting deconvolved instead of raw fluorescence images is advantageous for the down streaming task of segmentation.
Recently, a model has been trained on the same dataset with a 2.5D U-Net (Guo et al., 2020) instead of the here used 2D U-Net, improving virtual staining quality of the nucleus and membrane fluorescence, as information of neighboring planes in the 3D image stack are utilized during training. Furthermore, instead of a MSE loss, a mixed loss function (MAE and MS-SSIM) has been used. This model has been successfully applied to virtually stain the nucleus and membrane of images of the same cell line acquired on a Mantis microscope, a custom designed microscope with label-free and oblique light-sheet fluorescence microscopy (Ivanov et al., 2023; Liu et al., 2023). Compared to the quantitative phase images of the training data, which were generated using unpolarized illumination of a 3D bright-field stack, quantitative label-free imaging with phase and polarization is used on Mantis. This demonstrates that the model is applicable to different microscopes.
As virtual staining is not occupying fluorescence channels, the multiplexing capabilities are enhanced. The segmentation of cell nuclei and membranes facilitates the extraction of cell properties such as the volume and shape. The ability to accurately quantify cell properties becomes particularly significant when studying virus-infected cells. Understanding the effects of viral infections on cellular structures and dynamics is of importance in virology and immunology research. For this, the model generalizability towards confluency and infection must be further tested, and could potentially be improved by pooling data from different experimental conditions.

7.1.2 Methods

7.1.2.1 Sample preparation

Sample preparation details can be found in the OpenCell study (Cho et al., 2022). In brief, HEK293T cells (ATCC CRL-3216) were cultivated using DMEM high-glucose medium (Gibco, cat. #11965118) containing 10% fetal bovine serum (Omega Scientific, cat. #FB-1), supplemented with 2 mM glutamine (Gibco, cat. #25030081), penicillin and streptomycin (Gibco, cat. #15140163). All cell lines were kept at 37 °C and 5% CO_2. HEK293T cells were genetically tagged with split-fluorescent proteins using CRISPR/Cas9-based techniques. The nucleus was tagged with H2B-mIFP and the membrane with m-Scarlet-I-Caax (construct number pML442). Cells were transferred from a 10 cm plate to a 24 well plate coated with 50 µg/mL fibronectin (Corning, cat. #356008) one day before measurement with a density of 39K cells per well.

7.1.2.2 Setup and data acquisition

Brightfield and fluorescence images were acquired using a widefield DMi8 automated microscope by Leica in Köhler illumination (Köhler, 1893). For the training dataset, the bright-field channel was illuminated with an LED with an exposure time of 10 ms. The membrane channel was illuminated with an exposure time of 250 ms. The emitted light had a peak at 695 nm and passed a TRITC 554-595 filter. The nucleus channel was illuminated with an exposure time of 250 ms. The emitted light had a peak at 704 nm and was filtered with a Cy5 635-730 nm filter. The 63x magnifying objective (Leica HC PL APO 63x/1,30 GLYC CORR CS) with a numerical aperture of 1.3 was used with immersion oil with a refractive index of 1.47. The camera (Teledyne Photometrics Prime BSI Express Scientific CMOS camera) had a pixel area of 6.5 x 6.5 μm^2. The image size was 2048 x 2048 px^2 with a pixel size of 103.2 nm. The condenser aperture was set to an opening of 17 and the field aperture to 4. An incubation chamber by Oktolab was used, which kept the CO_2 concentration at 5% and the temperature at 37 °C.
Per well a 4x4 grid of 16 images was recorded resulting in 336 positions. Per position an axial range of ±12 µm around the focal plane with increments of 0.25 µm was measured resulting in 97 images per position. With µManager (Edelstein et al., 2010) (v2.0.0-gamma1 20201112) and an automated stage the xyz-positions could be defined prior to the measurements and then recorded automatically.

For the test dataset, different experimental settings were altered. A well was chosen per change in experimental condition and 12 positions imaged per well. Images were recorded 100x magnification with a numerical aperture of 1.47 (Leica, HC PL APO 100x/1,47 OIL CORR TIRF) and a pixel size of 65 nm, operating with an immersion oil with a refractive index of 1.518 or with 40x magnification with a numerical aperture of 1.10 (Leica, HC PL IRAPO 40x/1,10 W CORR) with a pixel size of 162.5 nm, used with water. The condenser

aperture was reduced to 12 (medium closed condenser aperture) or 6 (extreme closed condenser aperture). The bright-field exposure time was reduced to 5 ms.

7.1.2.3 Data preprocessing

Phase reconstruction and fluorescence deconvolution

The phase was reconstructed and the fluorescence channels deconvolved using waveOrder (v1.0.0, commit 683bcea) (Yeh et al., 2021). 3D phase deconvolution with TIKHONOV regularization of $1 \cdot 10^{-3}$ and z-padding of 15 slices at the volume borders was applied to the bright-field stacks. For fluorescence deconvolution, a 3D deconvolution with TIKHONOV regularization of $8 \cdot 10^{-4}$ and z-padding of 15 slices at the volume borders was applied to the fluorescence-stacks.

Registration

Registration offsets of the label-free and fluorescence channels along z-axis were determined using the BRENNER gradient (eq. 62) (Yazdanfar et al., 2008). The BRENNER gradient is an edge detector that measures the difference of the pixels in the image between their neighbors, which are two pixels away in horizontal and vertical direction. The FOV in a z-stack that had the highest BRENNER gradient was chosen as the focus FOV for the fluorescence and phase channels. Theoretically, the FOV with the lowest BRENNER gradient could be chosen as the focus for the bright-field channel, since bright-field images have minimal contrast in the focal plane. However, this led to unstable results and the bright-field images were assigned the same focal planes as the corresponding phase images.

$$B = \sum_{i=1}^{N} \sum_{j=1}^{M} [s(i,j) - s(i+m,j)]^2 \tag{62}$$

The focus of the label-free and membrane fluorescence channel was aligned and the focus distance between the membrane and nucleus fluorescence channel was maintained. Positions with a z-focus index in the membrane channel of < 25 and > 60 were filtered out, as the whole cell volume was not captured for these positions.

Tiling and normalization

OTSU's thresholding was applied to the membrane and nucleus fluorescence channels to distinguish between background and foreground pixels with *VisCy*, formerly known as *microDL* (v1.0.0, commit b9d6045 - 1a4218e) (Guo et al., 2020). In case of multiple fluorescence targets, the masks were merged across channels. The images were tiled into 256 x 256 px^2 patches with a sliding window of 126 x 126 px^2. Patches with a foreground fraction in the mask channel above 5% were kept as training data. The intensities of the images were normalized across the dataset.

7.1.2.4 Neural network architecture

A 2D U-Net was chosen as the neural network architecture (Guo et al., 2020; Ronneberger et al., 2015). Each convolution block in the encoding and decoding path consisted of two repeats of a convolution layer, a non-linear activation function ReLU, a batch normalization layer, and a dropout layer set to 20%. In the encoding path, 2x2 downsampling after each block was achieved by 2x2 convolutions with a stride of two. The filter size was 3 x 3 px^2. The number of filters per block was 16, 32, 64, 128, and 256, meaning that the size of the input was downsampled from 256 x 256 px^2 to 16 x 16 px^2 in total. In the decoding path, upsampling was achieved with bilinear interpolation. Feature maps of the encoding and decoding paths were concatenated at every level. The final output block only had a convolutional layer and a linear activation function.

A residual connection was added from the input of a convolution block to its output, which is known to promote faster convergence of the model (Drozdzal et al., 2016; Guo et al., 2020; Milletari et al., 2016). In the downsampling path, if downsampling followed after the convolution block, the input of the convolution block was 2x2 downsampled with average pooling and added to the input of the next convolution layer. If another convolution block follows instead, the input was simply added to the next convolution block's input. In the upsampling path, no resizing was applied to the residual connection, as all inputs were added within a layer level.

7.1.2.5 Model training

Model training was performed with *VisCy* (v1.0.0, commit b9d6045-1a4218e), (Guo et al., 2020). Phase or bright-field images were used as the input and either fluorescent nucleus, membrane, or both as target(s). The fluorescence channels were either deconvolved or used directly.

The images were split into a training and validation set in a 80/20 ratio. The batch size was set to 64 patches. The MSE was used as the loss functions. The optimizer was Adam (Kingma et al., 2014). A learning rate scheduler was chosen (Guo et al., 2020; Smith, 2017). The learning rate varied between a minimum and maximum bound and decayed with an exponential factor of $gamma^{epochs}$ after every cycle (min_learning_rate $= 5 \cdot 10^{-5}$, max_learning_rate $= 0.006$, scale_mode $=$ cycle (learning rate is scaled after every cycle), step_size $= 2$ (number of epochs per cycle), gamma $= 0.5$) . Early stopping was defined based on the minimum validation loss with a patience of 10 epochs. The maximum number of epoch was set to 70. The model trained between 12 and 70 epochs with a training time of 20-60 hours on an NVIDIA A40 GPU with 48 GB of memory.

7.1.2.6 Augmentation

Different augmentations were applied to each patch sampled in the training process including zoom, rotation, shear, blur, intensity variation, and noise. The magnitude of augmentation was sampled uniformly from a defined range per augmentation type (tab. 2). Gaussian blur was added to the image, where the sampled value defined the kernel standard deviation. Noise was added per pixel and sampled from a Gaussian distribution with mean $= 0$, where the drawn value defined the standard deviation of the distribution. The intensity of a normalized patch was shifted to a new mean and standard deviation. Intensity variation, noise and blur were only applied to the input.

Table 2: Uniform sampling ranges of augmentation parameters.

Augmentation	blur	int. mean	int. std	noise	rotation	shear	zoom
heavy augmentation	0-10	±0.5	±0.5	0-0.5	±180	±10	0.6-1.6
heavy augmentation x2	0-20	±1	±1	0-1	±180	±20	0.6-1.6
light augmentation	0-5	±0.2	±0.2	0-0.2	± 180	0	0.8-1.2
no intensity	0-10	0	0	0-0.5	±180	±10	0.6-1.6
no zoom	0-10	±0.5	±0.5	0-0.5	±180	±10	0

7.1.2.7 Segmentation with cellpose

The nucleus channel was segmented using cellpose (v2.2.1, commit b709064) (Stringer et al., 2021). The cellpose nucleus model was used as a basis and fine-tuned by re-training with 10 FOVs from the training dataset. Cellpose created one mask per nucleus instance and a FOV contained multiple individual nucleus masks. Masks were created for the predicted nucleus fluorescence images and ground truth images. Some nuclei in the ground truth fluorescence images did not emit a fluorescence signal due to labeling heterogeneity. Therefore, the ground truth masks were refined by adding missing nuclei using information from the label-free channel.

7.1.2.8 Evaluation metrics

The similarity of ground truth and predicted images was determined based on the nucleus mask channels. Masks were compared on an object-based level by creating matches of nucleus mask instances from the ground truth and predicted images with a linear assignment problem solver, and on a pixel-based level by distinguishing only between background (=not in a mask) and foreground (=in a mask) pixels in a FOV.

For object-based comparison nucleus masks were matched between ground truth and prediction using a linear assignment problem solver (lapsolver v1.1.0, commit 69f6e55). Masks that were matched but exceeded a distance threshold equal to the average long-axis length divided by 2 (corresponds to the radius for a perfectly round object) were not regarded as successful matches. True positive were mask instances that were present in the ground truth and had a match in the prediction. False positive were masks that were only present in the prediction and had no match in the ground truth. False negative were masks that were only present in the ground truth with no match in the prediction. On this basis the f1-score was calculated, with 0 being the worst and 1 the best possible value.

The intersection over union was calculated on a pixel-based level and describes how well the pixels of a ground truth mask and a predicted mask overlap, where 0 means no and 1 complete overlap.

7.1.2.9 Statistics

Mean values are listed with their respective standard deviation. Datasets were tested for normality by applying the Shapiro-Wilk test ($\alpha = 0.05$). As most datasets rejected the hypothesis of being normally distributed, the unpaired non-parametric Mann-Whitney-U-test ($\alpha = 0.05$) was used to determine significant differences between evaluation scores of different model runs. Levels of significance were classified as: $p \geqslant 0.05$ no significance, $p < 0.05$ significant difference (*), $p < 0.01$ very significant difference (**), $p < 0.001$ highly significant difference (***). Tests were performed using scipy version 1.10.1.

7.2 Virtual staining and denoising of the endoplasmic reticulum

7.2.1 Results and discussion

This section explores the potential of utilizing *VisCy* for generating fluorescence images of the endoplasmic reticulum (ER), as well as leveraging the intrinsic augmentation provided by the reorganization of organelles in living cells. A dataset of transmitted-light images as well as matching low-intensity and high-intensity fluorescence images of the ER in living cells imaged on a confocal microscope was used (fig. 51A-C). Inputs to the model are either transmitted-light, low-intensity fluorescence, or both types of images. The target is the high-intensity fluorescence channel.

When the model is trained with transmitted-light images as input, the predicted images roughly show the cell volume, but lose the fine structure of the ER (fig. 51D). This happens due to the fact that the ER structure is not clearly visible in the transmitted-light image and is similar to results reported in literature (Ounkomol et al., 2018).
When using the low-intensity fluorescence image as input, the fine structure of the ER is recovered. The resolution, which was determined with decorrelation analysis (Descloux et al., 2019), is 380 $\pm$ 10 nm for the high-intensity ground truth fluorescence images and 390 $\pm$ 14 nm for the predicted fluorescence images. Using both low-intensity and transmitted-light images does lead to a significant difference ($p = 0.158$) in SSIM values (fig. 52A). This renders low-intensity images as the appropriate input for predicting ER fluorescence.

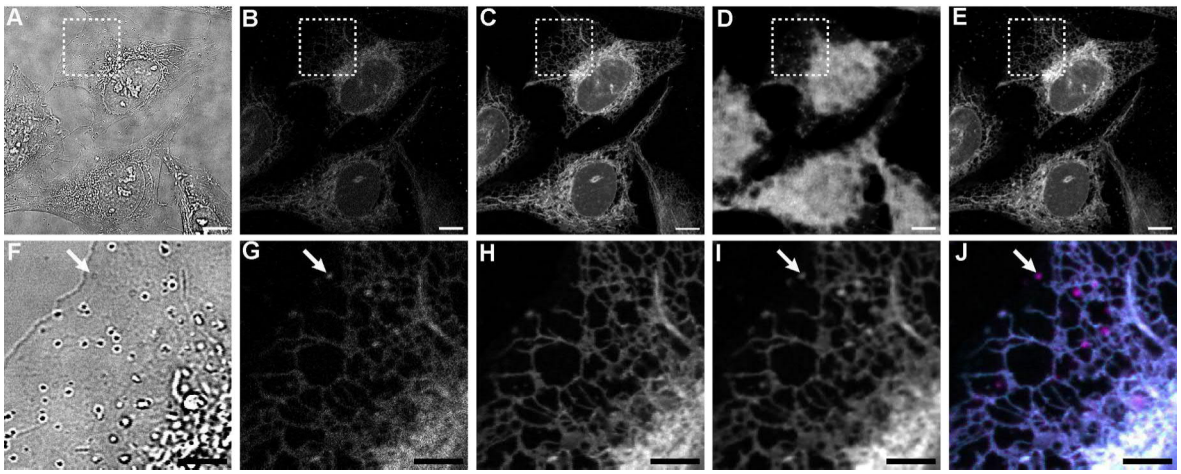

Figure 51: Prediction of the ER with different input modalities. **A** Transmitted-light image, **B** ground truth low-intensity fluorescence image, **C** ground truth high-intensity fluorescence image, **D** prediction with transmitted-light image as input. **E** Prediction with transmitted-light image and low-intensity fluorescence image as input. **I-J** Zoom-ins of **A-E**. The white arrows highlight spots of dye accumulation visible in the transmitted-light and low-intensity fluorescence image. These spots are amplified in the prediction in **I** and **J**. Scale bars 10 µm in **A-E** and 5 µm in **F-J**.

The organelle reorganization in living cells could be used as an intrinsic augmentation of the structure and potentially reduce the amount of FOVs that must be recorded to curate a training dataset. To explore this possibility, a movie with 30 frames (dt = 5 s) was acquired at each FOV. When comparing the SSIM values between using all time steps and only the first time step in a FOV the gain in performance is improved by 6% ($SSIM(t_{all})$ = 0.86 ± 0.06, $SSIM(t_0)$ = 0.81 ± 0.02) (fig. 52A). The ER structure probably did not change enough over the observed time window of 140 seconds and the images within a time series maintain a high level of similarity (fig. 52). The potential of recording movies of reorganizing organelles could be further explored with larger time steps.

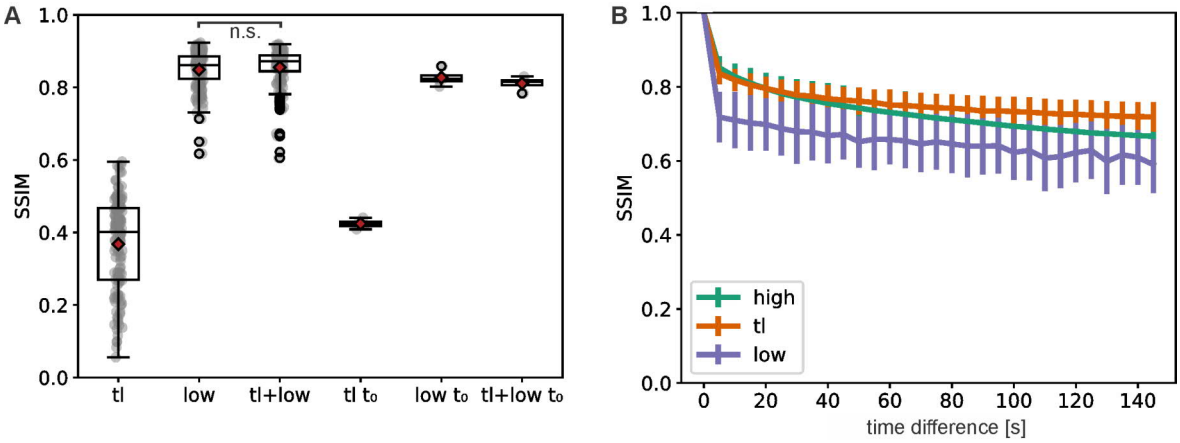

Figure 52: SSIM of the ER of different input modalities and over time. **A** SSIM (prediction and ground truth comparison) of predictions with transmitted-light (tl), low fluorescence intensity (low), and transmitted-light and low fluorescence intensity (tl+low). t_0 indicates that only the first image of the time series was used. N = 600 for t_{all} and N = X for t_0. **B** SSIM between different time points in a time series ($t_0 \rightarrow t_x$), N = 20, error bars = SD.

Reconstructing high-intensity fluorescence images from low-intensity fluorescence images using a U-Net has been demonstrated by Weigert et al., 2018. The approach is termed denoising, as the signal-to-noise ratio is improved by translating low-intensity to high-intensity fluorescence images. Compared to virtual staining, denoising maintains the different levels of intensity as heterogenic brightness information is available in the low-intensity fluorescence channel. However, denoising requires the usage of fluorescent labels which elevates experimental complexity. Fluorescence microscopy additionally suffers from photobleaching and -toxicity. It should be emphasized that these effects are strongly attenuated because denoising diminishes the required photons for imaging. Denoising is a vital alternative for targets that are not visible in the label-free channel and denoising of STED images of the endoplasmic reticulum is discussed in detail in chapter 6.

7.2.2 Methods

7.2.2.1 Cell preparation and imaging

U-2 OS cells (CLS Cell Lines Service GmbH) were treated with 1 µM ER-tracker Red stain (Thermo-Fisher #E34250), which is compatible with living cell and binds to the sulfonylurea receptors on the endoplasmic reticulum (ER). Cells were incubated for 15 min at 37 °C with 5% CO2 in LCIS (Sigma Aldrich), washed 2x with LCIS, and tempered (10 min) before image acquisition. Transmitted-light, high-intensity fluorescence, and low-intensity fluorescence images were acquired on a Leica SP8 confocal microscope. The fluorescence images were excited with a 561 nm diode laser (low-intensity: 0.1%, high-intensity: 0.5-1%). The transmitted-light image was created with the remaining light of the laser. The emission light was detected in a range of 570-700 nm. The images were recorded with a hybrid PMT (HyD3, Leica) with 10% gain in line accumulation mode and had a size of 1024 x 1024 px^2 with a pixel size of 92 nm. The dataset consists of 20 positions with 30 time points 5 s apart.

7.2.2.2 Data preprocessing

Foreground masks were created with unimodal thresholding with the *VisCy* pipeline (Guo et al., 2020). A 128 x 128 px^2 sliding window with stride 64 tiled the images and patches with $\geq$5% foreground pixels were kept. The intensity was normalized across the dataset.

7.2.2.3 Neural network architecture and model training

The inputs to the model were either transmitted light images, low-intensity fluorescence images, or both. The target images were the high-intensity fluorescence images. For training, all time positions or only the first time position of a FOV was used. The data was split into a 80% training and 20% validation set. The training and validation split was made across the spatial positions, meaning that different time points of a position were all either in the training or in the validation set. A 2D U-Net (Ronneberger et al., 2015) was used as the model architecture. The number of filters per block was 16, 32, and 128. The batch size was set to 128 and MAE was used as a loss function. Heavy augmentation with deactivated zoom was applied. The maximum allowed epochs was set to 50 with a patience based on the minimum validation loss of 20 epochs. Besides that, all model configurations were equal to the settings described in chapter 7.1.2.5. Training was executed on an NVIDIA 2080 Ti GPU with 12 of GB memory and took 1-2 hours for the complete dataset and around 5 minutes with the first time position only.

7.2.2.4 Evaluation

The SSIM was used to compare the similarity between ground truth and predicted images as well as the similarity between images within a time series (Wang et al., 2004). The resolution was determined using decorrelation analysis (v1.1.8, commit db536a3) (Descloux et al., 2019).

8 A pipeline for the analysis of single-particle tracking data

Single-particle tracking is a powerful method for investigating biomolecule mobility in living cells, with nanometer spatial and millisecond temporal resolution. This technique generates single-particle trajectories, from which valuable descriptors such as the diffusion coefficient and diffusion state can be extracted. This can serve as a proxy to report on protein activity in cells. Segmentation of a trajectory into parts of uniform diffusion behavior further aids in the analysis of changes in particle mobility within these trajectories, shedding light on transitions between distinct functional states of biomolecules. However, conducting such analyses is intricate and time-consuming. To streamline this process, a comprehensive pipeline named *SPTAnalyser* was developed. It comprises mean-squared displacement analysis to calculate diffusion coefficients of the single molecules and allows distinguishing between immobile, confined, and freely diffusing molecules. Occurrences of transitions between different diffusion modes are characterized with transition counting and hidden MARKOV modeling.

The methodology was applied to single-molecule trajectories of the hepatocyte growth factor receptor, also known as the MET receptor and human epidermal growth factor receptor 2 (HER2) in living HeLa cells. Both receptors are members of the receptor tyrosine kinase family and play a crucial role in neurodegenerative diseases and cancer (Lemmon et al., 2010; Majumder et al., 2019). The changes in mobility of receptors treated with different ligands were studied, demonstrating that this analytical pipeline facilitates rapid access to information regarding receptor functional states.

8.1 Results and discussion

8.1.1 Analysis pipeline for single-particle tracking data

Large amounts of SPT data need to be acquired and processed in order to make a meaningful biological statement. This is why the pipeline *SPTAnalyser* was built to assist in particle localization, tracking, and post-hoc analyses (fig. 53) (Rahm et al., 2021; 2022). *SPTAnalyser* is compatible with the localization softwares *ThunderSTORM* (Ovesny et al., 2014) and *RapidSTORM* (Wolter et al., 2012) and creates batch processing scripts for automated analysis. The localizations are connected to trajectories with the global BAYESIAN tracking software *swift* (Endesfelder et al., n.d.), and *SPTAnalyser* estimates the *a priori* tracking parameters. This is a crucial step, as the performance of *swift* is highly dependent on meaningful initial parameters (Martens et al., 2024). *Swift* splits individual trajectories into segments of uniform motion. Further data analysis in *SPTAnalyser* are applied to either trajectories or segments. MSD-based analysis determines diffusion coefficients and discerns between immobile, confined diffusing and freely diffusing molecules. Analysis over time is furthermore possible to find time-dependent trends. Transition counting quantifies the number of diffusion state changes within trajectories. Hidden MARKOV modeling reveals underlying states of diffusion coefficients and transition probabilities between them.

The pipeline was applied to single-particle tracking data of the membrane receptors MET (Harwardt et al., 2017; Rahm et al., 2021; 2022) and HER2 (Catapano et al., 2023) in live HeLa cells. MET was either labeled with a monoclonal Fab fragment that binds but does not activate or with the bacterial ligand InlB that binds and activates the receptor. HER2 is an orphan receptor without own ligand. It was labeled with a nanobody and treated with the ligands EGF, TGFα, EREG, and NRGβ1 that interact with different receptors of the ErbB sub-family.

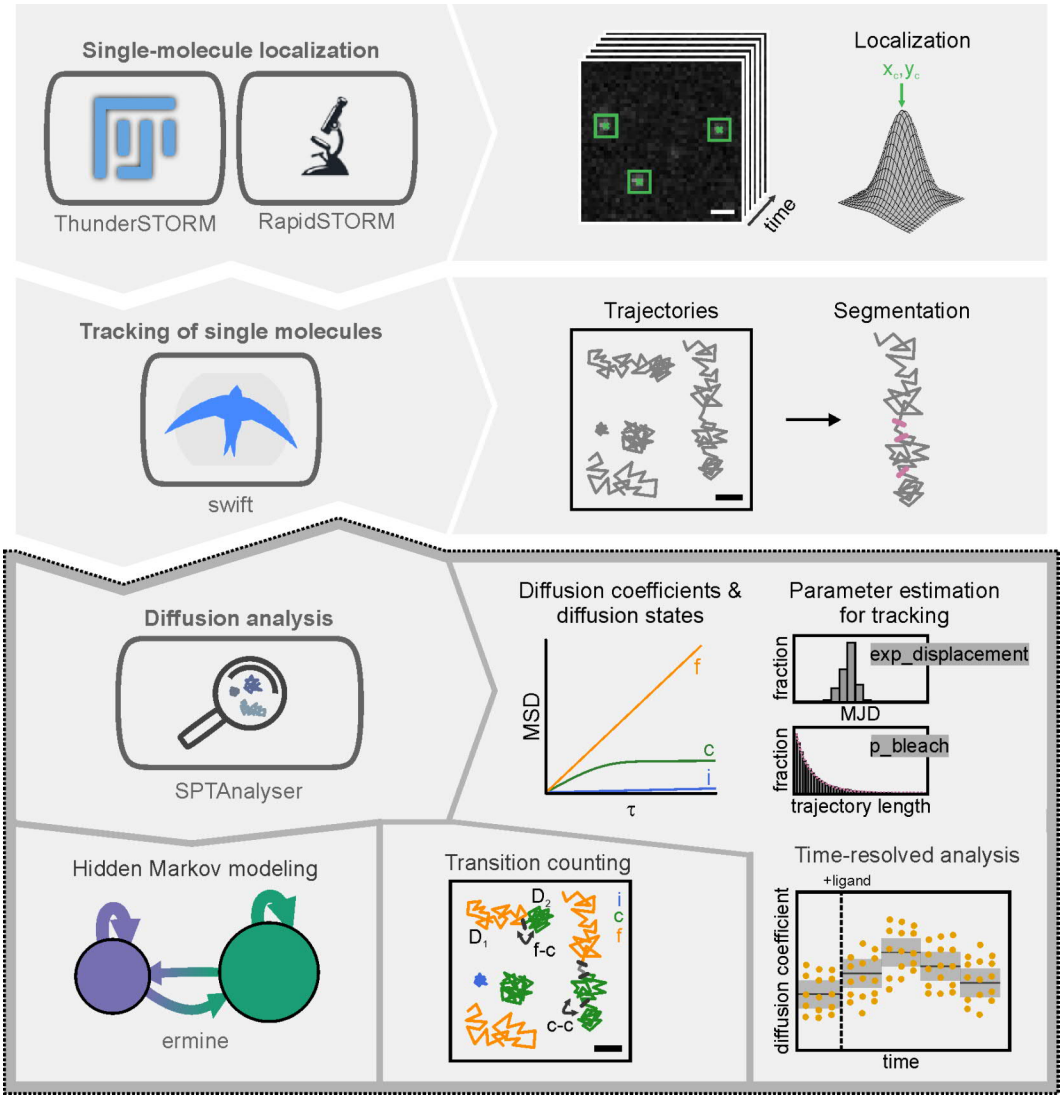

Figure 53: Workflow of the analysis of single-particle tracking data with *SPTAnalyser*. Single molecules are localized with either *ThunderSTORM* or *RapidSTORM*. *Swift* connects the localizations to trajectories and segments. Several input parameters of *swift* can be determined with *SPTAnalyser* such as the expected displacement of the particle and the bleaching probability of the fluorophore. Based on MSD-analysis, *SPTAnalyser* extracts diffusion coefficients and diffusion states (immobile = i, confined = c, free = f) and can display these properties over time. Transition counting quantifies the transitions of diffusion states between segments within a trajectory. Hidden MARKOV modeling can be applied to further analyze transitions between different mobility states (Rahm et al., 2021; 2022).

8.1.2 MSD-based analysis of diffusion states from trajectory segments of the MET receptor

Diffusion coefficients were calculated for the segments of single-molecule trajectories and the segments were classified as immobile, confined diffusing molecules or freely diffusing molecules (Rahm et al., 2021; 2022). InlB-bound receptors move slower than receptors in the resting state (fig. 54C) (confined diffusing molecules, Fab: 0.094 ± 0.007 $\mu m^2/s$, InlB: 0.051 ± 0.003 $\mu m^2/s$; free diffusing molecules Fab: 0.134 ± 0.004 $\mu m^2/s$, InlB: 0.084 ± 0.003 $\mu m^2/s$). They exhibit more confined motion (fig. 54D) with a smaller confinement radius (Fab: 0.18 ± 0.02 µm, InlB: 0.131 ± 0.009 µm) (fig. 54E). Interestingly, the confinement radii of the freely diffusing molecules are in the order of magnitude of the cell sizes. For both labels, the segment lengths of the confined state is shorter than those of the immobile and freely diffusing molecules. For InlB-labeled receptors, the segments classified as confined diffusing molecules last for 0.64 ± 0.02 s on average compared to 1.13 ± 0.03 s for immobile and 1.00 ± 0.04 s for freely diffusing molecules.

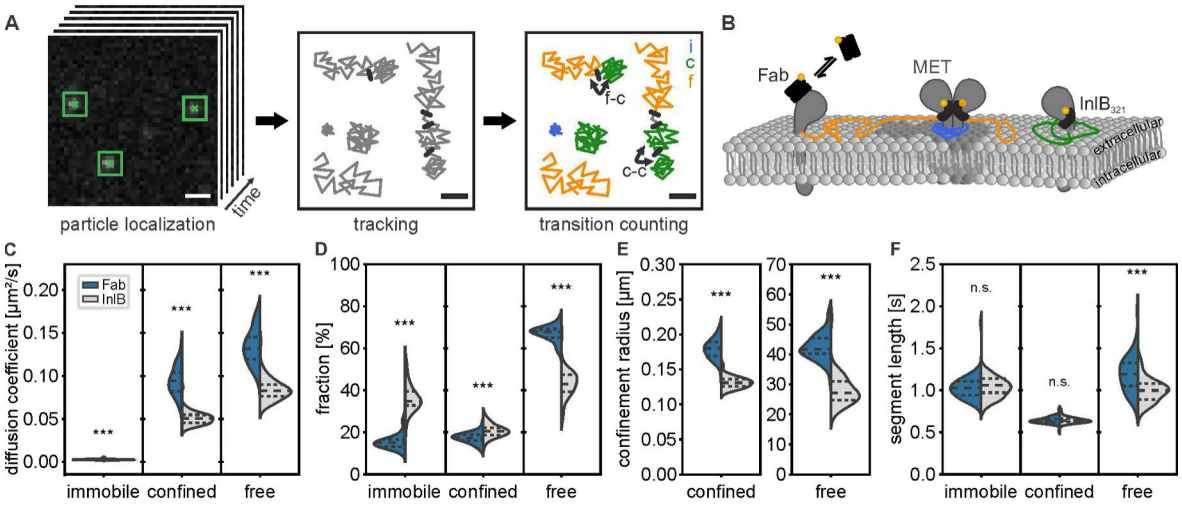

Figure 54: Analysis pipeline and MSD-based results of single-particle tracking data of the MET membrane receptor. A Diffraction-limited signals of membrane-bound fluorescently labeled receptors are localized and connected over time to trajectories (scale bar 1 µm) (left). Changes in diffusion behavior within a trajectory generate segments (schematic scale bar 100 nm) (middle). Segments are assigned to the diffusion states free (orange), confined (green), and immobile (blue) (right). **B** The transmembrane receptor MET is extracellularly labeled with a monoclonal ATTO 647N-Fab fragment that binds but does not activate the receptor and with ATTO 647N-InlB that binds and activates the receptor. The movement of bound labels is recorded and analyzed. The segment properties **C** diffusion coefficient, **D** population of diffusion states, **E** segment length, and **F** confinement radius are displayed as violin plots; dashed lines mark the quartiles; for each condition, Fab (blue) and InlB (gray), the average segment values of 60 cells are displayed; MANN-WHITNEY U test, $p > 0.05$ = n.s., $p < 0.05$ = *, $p < 0.001$ = *** (Rahm et al., 2021; 2022, CC-BY).

To probe the accuracy of diffusion state classification, trajectories of varying length of freely diffusing molecules were simulated and classified (fig. 55). Parts of the trajectories were

falsely classified as confined motion. The false negative rate, which describes the percentage of freely diffusing molecules that were falsely classified as confined, drops with increasing trajectory length. In addition, trajectories of freely diffusing molecules were simulated with trajectory lengths corresponding to the distribution in the Fab experiments, resulting in an error rate of 15%. This suggests that trajectories of freely diffusing molecules contribute partly to the confined population and reduce the average trajectory length, as a misclassification is more likely with shorter trajectories. Misclassification of other trajectory types are also possible, but have not been investigated further. The short trajectory lengths of the confined population (fig. 54F) may also hint at the possibility that the confined state acts as an intermediate state between free and immobile molecules (see chapter 8.1.3 for more details).

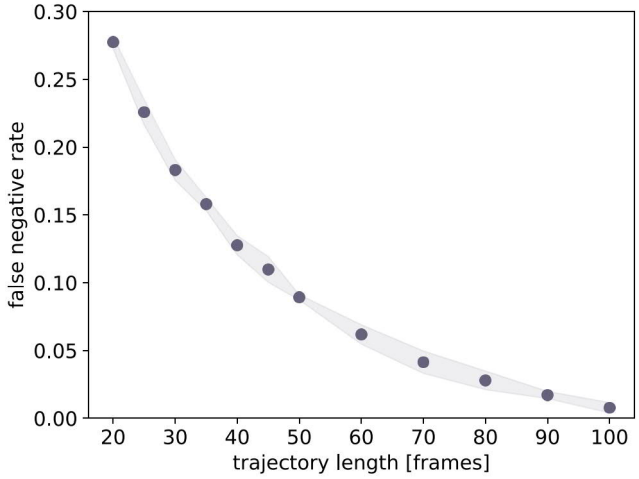

Figure 55: False negative rate of freely diffusing receptors that were classified as confined. Per defined trajectory length 1000 simulated trajectories of freely diffusing particles were classified into the diffusion states confined or free. The false negative rate of falsely assigned confined movement is depicted per trajectory length. The averages of five simulations per trajectory length are visualized as dots with the shaded areas denoting the 95% confidence interval (Rahm et al., 2021, CC-BY).

The results of the MSD-based analysis are in alignment with previous analyses of this dataset (Harwardt et al., 2017). In their work an MSD-based analysis based on the trajectories was executed and reported similar diffusion coefficients. Furthermore, the diffusion coefficients found for MET are within range of reported diffusion coefficients for other RTKs (Bag et al., 2015; Catapano et al., 2023; Chung et al., 2010; Harwardt et al., 2017; Marchetti et al., 2013). Addition of the InlB ligand activates the MET receptors and they are slowed down possibly due to receptor oligomerization and interaction with signaling proteins, which increases the molecular weight of the diffusing species and therefore decreases diffusion. An increase of MET dimers after InlB treatment was shown in Dietz et al., 2013. Analogously,

ligand activation of EGFR (Chung et al., 2010; Low-Nam et al., 2011) were reported to show slower diffusion. Particle confinement can stem from nanodomains, which are 0.01-0.2 µm (Pike, 2006) in size, such as clathrin-coated pits that have been shown to be the main endocytosis pathway of MET (Harwardt et al., 2017; Li et al., 2005). The actin cytoskeleton may also play a role in confinement (Harwardt et al., 2017) and has a mesh size of 0.2-0.6 µm (Kusimi et al., 1996). The observed confinement radii of MET are in the range of both the domain size of actin and nanodomains.

8.1.3 Transition counts reveal differences in diffusion state transitions for activated MET receptors

The transition counts analysis focuses on diffusion state changes within trajectories (Rahm et al., 2021). The occurrence of diffusion state transitions between segments within a trajectory are counted. Exemplary cells with color-coded segments are shown in fig. 56A and B. The zoom-ins show trajectories that consist of at least two segments and therefore exhibit transitions.

For comparison, the transition counts were normalized to sum to one across all transition types per cell (fig. 56C). The probability of a particle remaining in the freely diffusing state is higher for Fab-bound receptors compared to InlB-bound receptors (Fab: 31 ± 2%, InlB: 13.2 ± 0.8%). The probability that a receptor remains immobile in the following segment is higher for InlB-bound receptors compared to unactivated cells (Fab: 22 ± 1%, InlB: 42 ± 2%) (fig. 56C). Homogeneous transitions of the confined diffusion state are less probable in both cases (Fab: 2.2 ± 0.3%, InlB: 2.6 ± 0.2%). Among heterogeneous transitions, transitions from free to immobile (Fab: 11.4 ± 0.4%, InlB: 9.8 ± 0.3%) and immobile to free (Fab: 8.9 ± 0.4%, InlB: 9.1 ± 0.3%) are most probable. Frequencies were further normalized per diffusion state and cell to compare the transitions outgoing from a diffusion state in more detail (fig. 56D). For both conditions, the predominant homogeneous transition is from immobile to immobile followed by free to free. The confined state appears as an intermediate state, with transitions from one confined state to another being less likely than heterogeneous transitions outgoing from the confined state to an immobile or free state. Moreover, for receptors bound by InlB, the probability of transitioning to the immobile state is more likely, while the population of the free state is less pronounced when compared to resting cells.

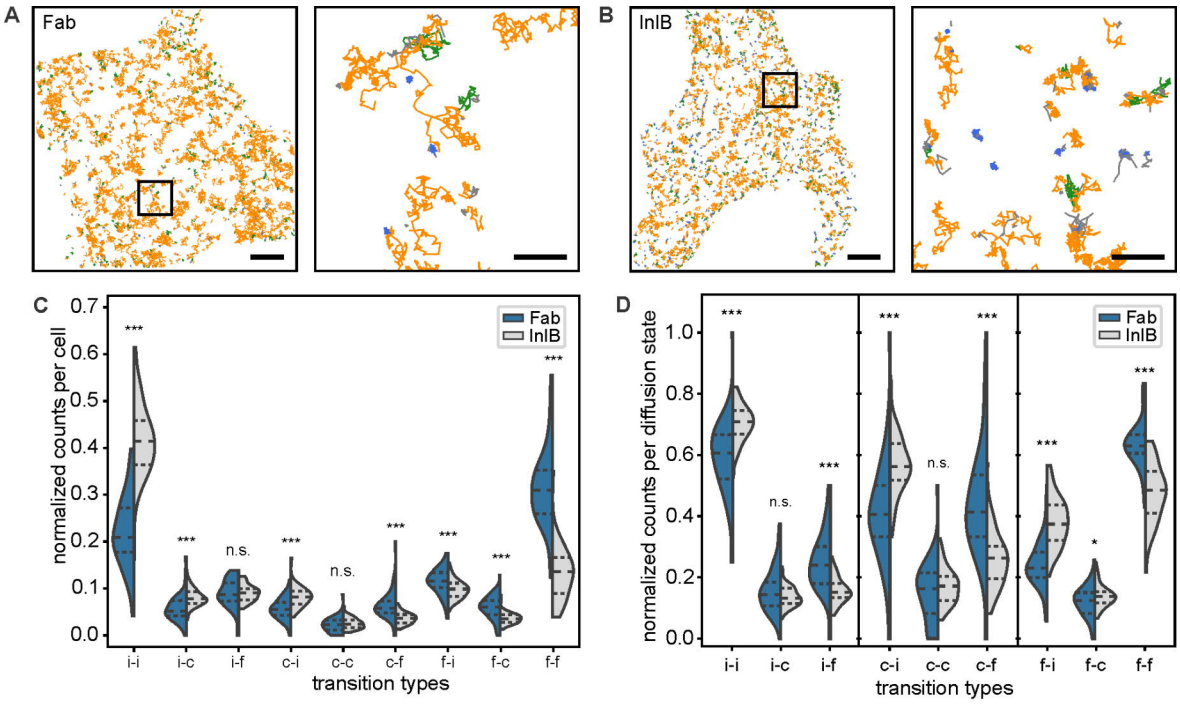

Figure 56: Single-molecule trajectories and quantification of transitions within trajectories. Single-molecule trajectories of **A** Fab-bound and **B** InlB-bound MET receptors within exemplary cells. Diffusion states of segments are highlighted in colors (free: orange, confined: green, immobile: blue; gray segments are below the necessary length to be classified and have no diffusion state label). Highlighted regions in the overview images are shown as zoom-ins (right). In the zoom-ins only trajectories showing at least one transition are displayed. Scale bars 5 µm, zoom-ins 1 µm. **C** Comparison of normalized transition counts per cell between Fab- and InlB-bound MET receptors (i = immobile, c = confined, f = free). **D** Comparison of normalized transition counts per diffusion state between Fab- and InlB-bound MET receptors. For each condition 60 cells were analyzed (Rahm et al., 2021).

On average 184 ± 11 switching events in all trajectories of a cell were observed. In contrast to the average number of 1440 ± 90 trajectories per cell, this number appears relatively small, showing that switching events rarely occur within the observed time window of a trajectory (1.36 ± 0.06 s) and the measurement time per cell (20 s). Longer trajectories contain more transitions (fig. S13A). However, it is notable that 70%, of trajectories remain static in their diffusion mode and consist of only one segment (fig. S13B).

The number of transition events can be increased by up to 30% by masking unclassified segments, as segments with a length below the threshold of 20 frames were not classified due to the small amount of data points (fig. S13C). In the masking approach, the transitions of adjoining segments with a defined diffusion state around the short, unclassified segment are counted. If multiple unclassified segments followed in a row, no transitions were counted, as the unclassified segments span over a too large time window of unknown diffusion behavior. Without masking, 22.2 ± 0.6% of transitions only contain segments with a defined diffusion state. With masking, this value increased to 71.6 ± 0.7%. Some unclassified transitions still remain in the case where multiple short segments are adjacent. Except for the free to

immobile transitions, masking does not significantly influence the relative frequency counts of diffusion state transitions (fig. S13D, E).

8.1.4 Hidden Markov modeling of single-particle trajectories of MET receptors

Hidden MARKOV modeling was applied to single-particle tracking data of the MET receptor as an alternative analysis to examine diffusion state transitions within trajectories (Malkusch et al., 2022; Rahm et al., 2022). The number of states for the model was selected based on the Bayesian information criterion (BIC). For both InlB- and Fab-bound receptors, a three-state model was found to be most appropriate. The three-state model consists of an immobile state (D_{Fab} 0.0 $\mu m^2/s$, D_{InlB} 0.0 $\mu m^2/s$), a slow state (D_{Fab} 0.065 $\pm$ 0.002 $\mu m^2/s$, D_{InlB} 0.029 $\pm$ 0.002 $\mu m^2/s$), and a fast diffusing state (D_{Fab} 0.234 $\pm$ 0.004 $\mu m^2/s$, D_{InlB} 0.199 $\pm$ 0.004 $\mu m^2/s$) (fig. 57A, B). InlB-bound receptors exhibit a higher population of the slow and immobile state (mobile: Fab 52.8 $\pm$ 0.5%, InlB 36.8 $\pm$ 0.05%; slow: Fab 30.8 $\pm$ 0.6%, InlB 38.5 $\pm$ 0.4%; immobile: Fab 16.3 $\pm$ 0.4%, InlB 24.8 $\pm$ 0.5%) and have elevated transition probabilities between the states. The proportion of transitions towards slower diffusion states compared to faster diffusion states increases for InlB-treated cells. The majority of heterogeneous transitions involve the slow state.
As the hidden MARKOV model is based on a BROWNIAN diffusion model, a discrimination between different diffusion types is not directly possible. Further investigation was conducted to ascertain whether the decreased diffusion coefficients of the states could be correlated with confined motion. The jump angles of trajectories within diffusion states were calculated fast: Fab 89.4 $\pm$ 0.3°, InlB 86.0 $\pm$ 0.3°; slow: Fab 85.4 $\pm$ 0.4°, InlB 82.1 $\pm$ 0.3°; immobile: Fab 79.9 $\pm$ 0.6°, InlB 79.3 $\pm$ 0.3°) (fig. 57C, D). Trajectories belonging to slower diffusion states have smaller angles, meaning a more confined motion. Additionally, activated receptors exhibited slightly smaller jump angles compared to their unactivated counterparts.

The slow state, which is linked to confined motion based on the jump distance analysis, is involved in most heterogeneous state transitions and could be regarded as an intermediate state between the mobile and immobile state. Similarly, the transition counts analysis showed a high chance of switching the diffusion state outgoing from the confined state (chapter 8.1.3). Receptor activation results in slower diffusion and an increased population in the immobile state, both shown with hidden MARKOV modeling and MSD-based analysis. These findings align with other observations of receptor tyrosine kinase activation and internalization (Bag et al., 2015; Catapano et al., 2023; Harwardt et al., 2017; Marchetti et al., 2013). Receptor dimerization is enhanced with activation, which leads to slower diffusion (Váradi et al., 2019). It is likely that receptors entering an intermediate diffusion state either become immobilized, a preliminary step before endocytosis, or revert to a mobile state, potentially for interaction with other proteins.

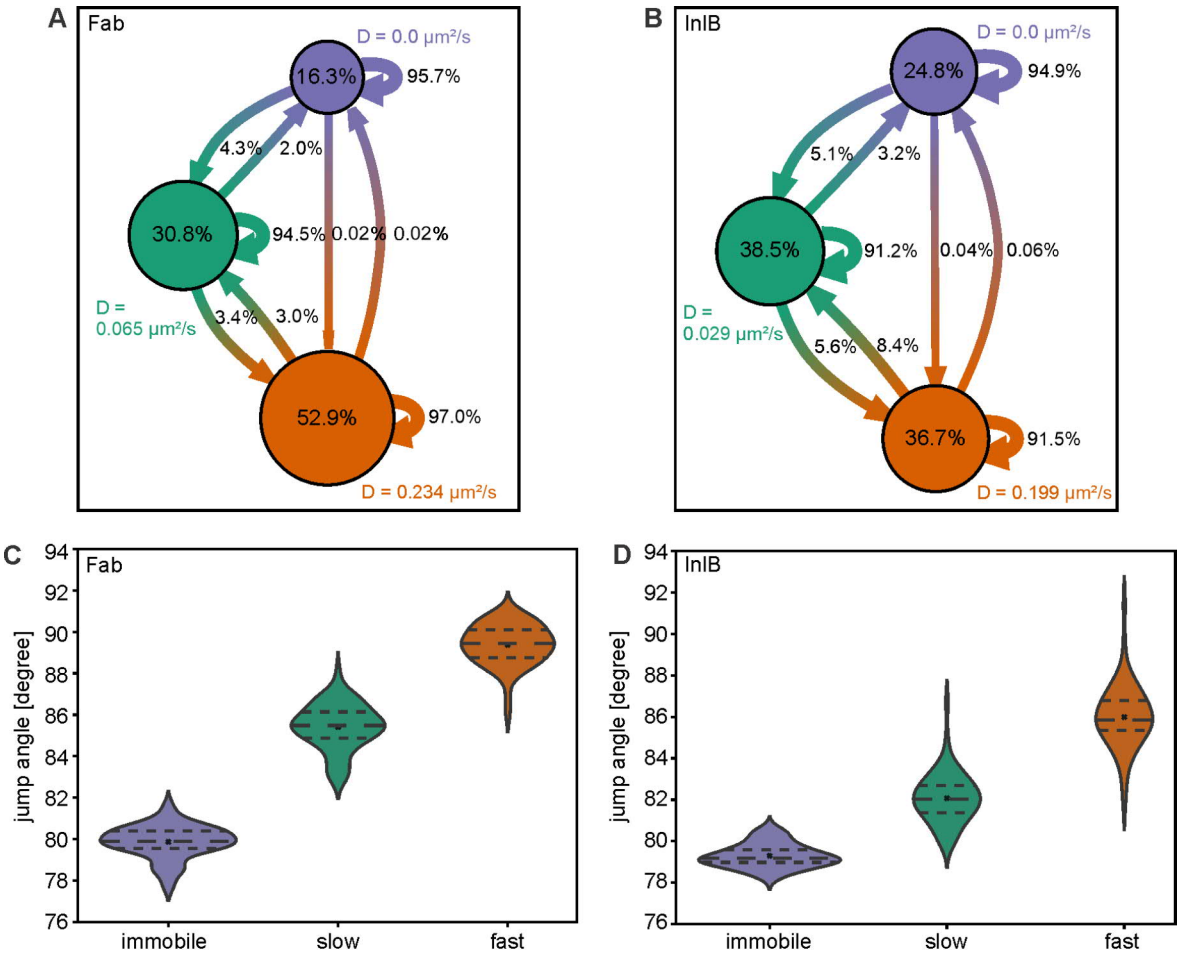

Figure 57: State transition diagrams of MET receptor trajectories from hidden Markov modeling. Average state population, transition probabilities, and diffusion coefficients of the three different states of **A** Fab-bound and **B** InlB-bound MET receptor trajectories. Jump angles of the **C** Fab-bound and **D** InlB-bound MET receptor trajectories of the three different states. For each condition 60 cells were analyzed (Rahm et al., 2022).

8.1.5 Temporal response of HER2 in ligand treated cells

SPT is limited by photobleaching, allowing to observe molecules only for a few seconds, but ErbB receptor signaling initiation is on a minute-scale (Hass et al., 2017). To bridge this gap, sequential measurements of HER2 mobility in various cells within the same coverslip were conducted (Catapano et al., 2023), following a strategy shown for VEGFR-2 (Rocha-azevedo et al., 2020). This allows monitoring of receptor activation after ligand stimulation for longer time periods. The analysis was integrated in *SPTAnalyser* to facilitate rapid data processing.

HER2 has no known ligand and activation occurs through stimulation of other receptors and the formation of heterodimers. Studied ligands include epidermal growth factor (EGF) and transforming growth factor α (TGFα), which activate EGFR exclusively; Epiregulin (EREG), which activates HER4 and EGFR; and neuregulin (NRGβ1), which activates HER3

and HER4. Cells were analyzed before ligand treatment to obtain the "dynamic fingerprint" of the resting state. The different activating ligands were added after 5 min and cells were pooled in 5 min intervals to study the response over 20 min. The fraction of immobile receptors (dot plots in fig. 58) and the level of phosphorylated HER2 (boxplots in fig. 58) was determined. The maximum immobile fraction occurred after 5 min for EGF, TGFα, and EREG. NRGβ1 showed a maximum after 10 min of stimulation. The activation strength can be correlated to the increase of the immobile population and was strongest for EGF, followed by TGFα, EREG, and NRGβ1. The maximum population of phosphorylated HER2 was after 2-5 min and correlates to the maximum percentage of immobile molecules.

The observed temporal diffusion patterns of HER2 closely align with the phosphorylation strength from western blots and reported kinetics of signaling activation (Hass et al., 2017). This suggests that diffusion characteristics are indicative of HER2 activation within heteromeric receptor complexes. The analysis furthermore enables the characterization of ligand-specific activation profiles associated with various heteromeric receptor complex formations, each showcasing distinct activation kinetics, strength, and diffusion patterns.

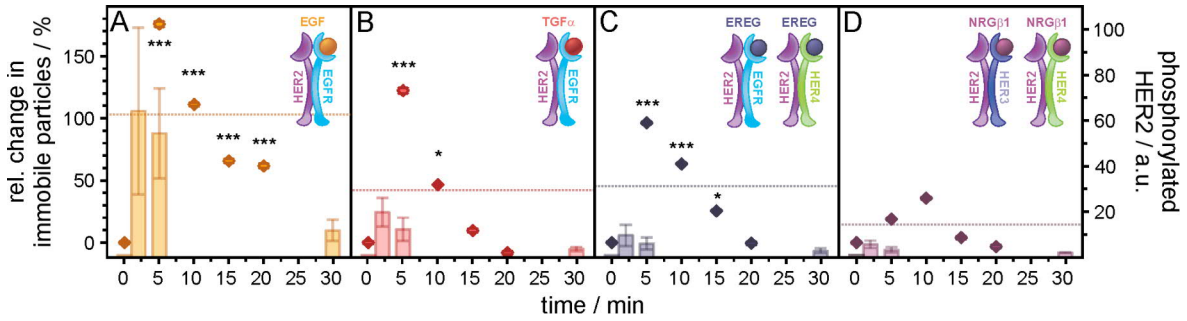

Figure 58: HER2 temporal responses to ligand treatment. Relative change of the percentage of immobile molecules are displayed as dot plots over time with error bars being the standard error of the difference (SED) (N = 200). WILCOXON signed-rank tests probed the differences of stimulated vs. untreated cells. Barplots show HER2 phosphorylation levels from western blots (N = 3) with error bars depicting the SEM (Catapano et al., 2023, CC-BY).

The analysis of temporal diffusion patterns and diffusion state transitions is crucial for understanding receptor dynamics. *SPTAnalyser* is a straightforward applicable toolbox incorporating these aspects, which allows the rapid analysis of single-particle tracking data of different molecules. Information on protein mobility can then be correlated with functional states. This expands the existing analyses that focus on transitions in molecular motion (Hubicka et al., 2020; Karslake et al., 2021; Requena et al., 2023; Vink et al., 2020; Weron et al., 2017). Future developments may concentrate on extending trajectory lengths to enhance the observation frequency of transitions between distinct mobility states. This could be achieved by using more photostable fluorophores, transiently binding labels based on the DNA-PAINT approach (Stehr et al., 2021), or by denoising SPT movies recorded at lower laser intensities (Kefer et al., 2021). These techniques could also be utilized to extend the observation window of protein mobility in a single cell, eliminating the need for measuring multiple cells at various time points to capture time-dependent effects. Requena et al., 2023 demonstrated the characterization of diffusion properties at every time step of the trajectory with a neural network. Another exciting extension is dual-color SPT (Subach et al., 2010; You et al., 2016), which can potentially relate changes in diffusion states to molecular interactions such as the formation of temporary complexes between two receptors. The analyses of spatiotemporal hotspots are further exciting developments that might point to dynamic protein interactions as nanoclusters on the cell membrane (Wallis et al., 2023).

8.2 Methods

8.2.1 A pipeline for the analysis of single-particle tracking data

8.2.1.1 Data acquisition

The dataset analyzed in this study was previously published by Harwardt et al., 2017 and can be found on the EMBL/EBI BioStudies server (https://www.ebi.ac.uk/biostudies/studies/SBSST712). In brief, MET receptors were labeled with the uPAINT technique in HeLa cells. The non-activating anti-MET antibody fragment Fab and the bacterial protein $InlB_{321}$ (named InlB in the following), which is fully functional (Dietz et al., 2013; 2019b), were labeled with ATTO 674N. Serum-free imaging medium containing high glucose Dulbecco's modified Eagle medium (DMEM) (Gibco, Life Technologies, Waltham, MA, USA) with 50 nM HEPES buffer and 1% GlutaMAX (Gibco, Life Technologies) was used. Imaging was executed at 23 °C in TIRF mode with an N-STORM microscope (Nikon, Düsseldorf, Germany). Fluorophores were excited with a 647 nm laser with an intensity of 0.1 kW/cm^2. The image size was 256 x 256 px^2 with a pixel size of 158 nm. SPT movies of 1000 frames were recorded with an EMCCD and a camera integration time of 20 ms. 60 cells on three different measurement days were imaged for Fab- and InlB-bound MET receptors respectively.

8.2.1.2 Single-molecule localization

Single-molecule localization was performed by fitting the fluorescence signal with a normal distribution and maximum likelihood estimation using the *ThunderSTORM* (Ovesny et al., 2014) plugin (version dev-2016-09-10-b1) of the imaging software *Fiji* (Schindelin et al., 2012). "Multi-emitter fitting analysis" was enabled with a maximum of 3 emitters and an intensity range of 73 to 1225 photons per emitter. The "remove duplicates filter" was applied.

8.2.1.3 Single-particle tracking

The localizations were connected to trajectories using the *swift* tracking software (version 0.4.2) (Endesfelder et al., n.d.). Parameters for *swift* were estimated using the *SPTAnalyser* software (Rahm et al., 2021). The parameters "diffraction_limit" = 14 nm, "exp_displacement" = 85 nm (Fab) / 75 nm (InlB), "p_bleach" = 0.01 (Fab) / 0.014 (InlB), and "p_switch" = 0.01 were set globally for all cells. The parameters "exp_noise_rate" and "precision" were calculated individually per cell. Remaining parameters were kept as default. More details on parameter definition and determination can be found in the *SPTAnalyser* manual at https://github.com/HeilemannLab/SPTAnalyser. *Swift* divides trajectories into segments if the diffusion behavior of the particle changes.

8.2.1.4 Diffusion coefficient and diffusion state determination

MSD-based analysis was performed in *SPTAnalyser* and diffusion coefficients and diffusion states were determined per segment. Diffusion coefficients were calculated with eq. (63), with dof = degrees of freedom (4 for 2D dynamics) and Δt = time step between measurements or camera integration time (Δt = 0.02 s). Solely the first four time steps were taken into account for the calculation of the diffusion coefficient. Segments with a diffusion coefficient below 0 were discarded.

$$D = \frac{MSD(\Delta t)}{dof \cdot \Delta t} \tag{63}$$

Segments with a minimum length of 20 frames (400 ms) were classified in the diffusion states immobile, confined or free (Catapano et al., 2023; Harwardt et al., 2017; Orré et al., 2021; Rossier et al., 2012). In a first step, the segments were separated into immobile and mobile diffusion. A threshold was derived from the dynamic localization error (eq. 65) calculated per cell. The third quartile of the dynamic localization errors of the cells was used for the calculation of D_{min}. The parameter n was set to four, as four points were regarded in the calculation of the diffusion coefficients. The result was a D_{min} threshold of 0.0028 µm^2/s. Segments with a diffusion coefficient $< D_{min}$ were labeled as immobile and with a diffusion coefficient $\geqslant D_{min}$ as mobile.

$$D_{min} = \frac{\sigma_{dyn}^2}{dof \cdot n \cdot \Delta t} \tag{64}$$

$$\sigma_{dyn} = \sqrt{\frac{\langle MSD(0) \rangle + \frac{dof}{3} \langle D \rangle \cdot \Delta t}{dof}} \tag{65}$$

Mobile segments were classified as confined or freely diffusing molecules by fitting 60% of the MSD values with eq. 66, where r_c is the radius of confined diffusion and τ is a time constant. The τ values of the segments were compared to a $\tau_{threshold}$ value (eq. 67), with a = 0.6 being the fraction of MSD values used for the fit in eq. 66, l_{min} = 20 frames of the minimum length of segments to be classified into diffusion states, and a factor of 0.5. Segments with $\tau \geqslant \tau_{threshold}$ are classified as freely diffusing molecules, and segments with lower values as confined diffusing molecules.

$$MSD(\Delta t) = \frac{4}{3} r_c^2 \cdot (1 - e^{-\Delta t/\tau}) \quad \rightarrow \quad \tau = \frac{r_c^2}{3D_c} \tag{66}$$

$$\tau_{threshold} = \Delta t \cdot a \cdot l_{min} \cdot 0.5 \tag{67}$$

8.2.1.5 Transition counts

The transition count analysis was conducted using *SPTAnalyser* and applied to trajectories with a minimum of two segments. Within this analysis, three distinct diffusion states were discerned: immobile (i), confined (c), and free (f). Consequently, nine distinct transition types were identified: i-i, i-c, i-f, c-i, c-c, c-f, f-i, f-c, and f-f. Transitions occurring between a segment characterized by a diffusion state and a segment lacking such characterization were excluded. Segments with a duration less than 20 frames were masked, and transitions around the masked segment counted. The mask value was synchronized with the threshold for the MSD-based classification of diffusion states. To facilitate comparisons of transitions across cells, the transition counts were normalized to sum to one per cell. Alternatively, for the purpose of comparing the transitions outgoing from the three different diffusion states, transition counts were normalized per diffusion state within each cell.

8.2.1.6 Hidden Markov modeling

Hidden Markov modeling was conducted using *SPTAnalyser* using the *pyErmine* library (https://github.com/HeilemannLab/pyErmine) (Malkusch et al., 2022). The model requires initial parameters, including the number of states, their respective diffusion coefficients, the initial probabilities associated with each state, and the transition probabilities between these states. These initial parameters were determined by fitting a jump distance mixture model to the jump distance distribution of the cells. The resulting mixture model provided weights and diffusion coefficients that served as initial estimates. For the initial transition probabilities, a probability of 0.9 was assigned to remain in the same state. Multiple models were generated with different numbers of states, ranging from 1 to 5. The Bayesian information criterion (BIC) consistently pointed to a three-state model as the most suitable choice for all cells and experimental conditions. To check if a diffusion state may be regarded as immobile, a limit of detection (LOD) was defined based on the average localization error σ of the localizations (eq. 68). All states with a diffusion coefficient below the limit of detection were categorized as immobile and their diffusion coefficients were set to 0.0 $\mu m^2/s$. To account for static and dynamic errors, the apparent diffusion coefficients of the mobile states were corrected with eq. 69, where ϵ corresponds to the static localization error.

$$LOD = 4\sigma^2 \tag{68}$$

$$D = \frac{r^2 - 4\epsilon^2}{\frac{8}{3}dt} \tag{69}$$

8.2.1.7 Statistical analysis

Average values and standard errors of the mean (SEMs) were calculated per cell and globally averaged. Nonparametric tests were chosen to test the significance as tests for normal distributions were partly rejected (tested with the SHAPIRO-WILK test for normality, significance level 0.05). MANN-WHITNEY U tests were used to compare distributions from two treatment groups. Levels of significance were classified as follows: $p > 0.05$ no significant difference (n.s.), $p < 0.05$ significant difference (*), $p < 0.01$ very significant difference (**), $p < 0.001$ highly significant difference (***).

8.2.2 Temporal response of HER2 in ligand treated cells

8.2.2.1 Data acquisition

The HER2 dataset was recorded in the study of Catapano et al., 2023 and made publicly available, where more experimental details can be found (https://www.ebi.ac.uk/biostudies/bioimages/studies/S-BIAD597).

Essentially, the azide-modified 2Rs15d nanobody targeting HER2 that carried a Cy3B-PEG6-DBCO label (Teodori et al., 2022) was added with 2 nM concentration 10 min prior to SPT measurements. Measurements were conducted between 21 and 23 °C in fresh Live Cell Imaging Solution (Invitrogen, Waltham, MA, USA). The SPT movies were recorded in TIRF mode on a commercial widefield microscope (N-STORM; Nikon, Düsseldorf, Germany) with an oil-immersion objective (100× Apo TIRF oil; NA 1.49) and a 1.5x magnification lens in the detection beam path. A 561-nm laser with an intensity of 6.3 W/cm^2 was used to excite the fluorophores. SPT movies with 256 x 256 px^2 were acquired with a pixel size of 105 nm. One movie consisted of 1000 frames with a camera integration time of 20 ms.

8.2.2.2 Time-series analysis

For the time-series analysis, five HeLa cells were recorded in resting condition. 20 nM ligand solution of either epidermal growth factor (EGF) (#AF-100-15), transforming growth factor alpha (TGFα) (#100-16A), neuregulin beta 1 (NRGβ1) (#100-03), or epiregulin (EREG) (#100-04) (all from PeproTech, Waltham, MA, USA) was added 5 min after measurement start and SPT movies were continued to be acquired for 30 min per coverslip. Per ligand condition, 8 coverslips with 25 cells from at least 6 different measurement days were recorded. Localization and tracking analysis was executed analogously to the MET dataset. Diffusion states were calculated, averaged per cell and pooled into 5 min time intervals.

8.2.2.3 Statistical analysis

Wilcoxon signed-rank tests were used to validate data from the different time intervals compared to the cells in the resting condition ($p \geqslant 0.05$ no significant difference (not labeled), $p < 0.05$ significant difference (*), $p < 0.01$ very significant difference (**), $p < 0.001$ highly significant difference (***)). Most populations did not conform to the assumption of normal distribution (Shapiro-Wilk test, $\alpha = 0.05$), which is why non-parametric tests were chosen.

9 Spatial organization of membrane receptors

Cellular proteins constitute a highly intricate machinery that governs vital biological processes. Membrane receptors and their ligands play a central role, as their interactions initiate complex signaling cascades regulating diverse cellular activities. The spatial proximity of the involved proteins is determinant in understanding how multiple proteins organize into functional assemblies. Single-molecule localization microscopy (SMLM) offers the required resolution to examine the spatial organization of protein assemblies. Moreover, it enables simultaneous multi-color imaging, facilitating the study of multiple targets in a cell.
To investigate the spatial organization of proteins, a data pipeline was established which extracts the density and distribution in multi-target SMLM images. The pipeline was employed to study the fibroblast growth factor receptor (FGFR) network, which comprises four receptors (FGFR1-4), various isoforms, and 22 ligands (FGF1-14 and FGF 16-23) (Raju et al., 2014). The FGFR network is involved in numerous intracellular pathways, including cell growth, differentiation, and survival. Despite their significance in cellular processes and medical contexts, the molecular underpinning of these mechanisms remain incompletely understood. The focus of the study was on how the universal ligand FGF1 influences the spatial arrangement of the FGFR network on the cell membrane (Schröder et al., 2021).

9.1 Results and discussion

9.1.1 Analysis pipeline to unravel the spatial organization of proteins

In a single-molecule localization experiment a movie of spatially separated fluorescent events is recorded and the events are localized with high precision. In a post-processing step, the super-resolved signals are combined to a single image and rendered. Positions of emitters that are bound to a molecular target can be combined to form molecule clusters, as their localization coordinates are very close to each other. The centroid of such a cluster serves as a representative location of a protein assembly in a cell.

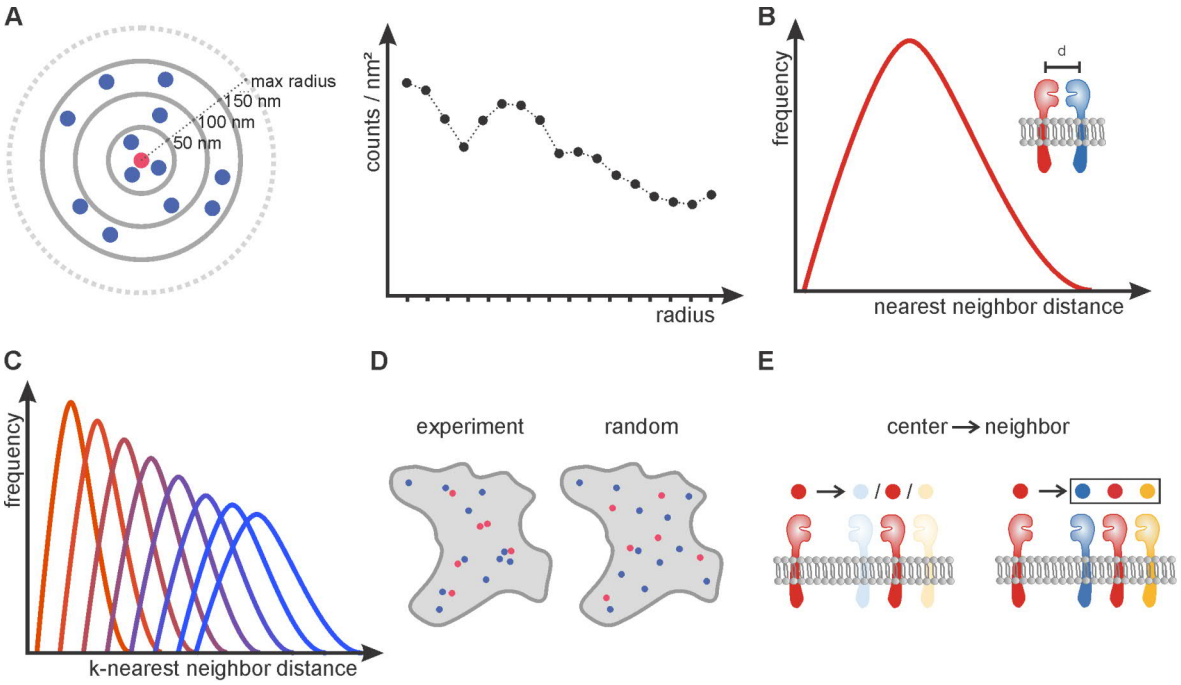

Figure 59: Distance distribution analysis. **A** The number of neighbors is counted within increasing circle increments around a center coordinate. This is repeated for all center coordinates and the number of counts displayed per circle radius. **B** Nearest neighbor distances analysis between two receptor types **C** k-nearest neighbor distance analysis. **D** The distributions can be compared to randomly distributed emitters within the cell area. **E** For multiple targets, centers and neighbors can be selected. In the left case, the red receptor is the center and being compared to only red receptors as neighbors. In the right case, the red receptor is the center and compared to all other receptors (Schröder et al., 2021).

The here presented analysis facilitates the characterization of spatial short- and long-ranged distance distributions of receptor clusters using various methodologies (fig. 59). To explore potential clustering, an area around a cluster coordinate is segmented with circles (fig. 59A). The number of neighbors is counted within the circle increments and the analysis is repeated for all cluster coordinates. This allows to observe patterns of clustering at different spatial scales. K-nearest neighbor distances provide further information on how FGFR1-4 acts as a protein network in response to ligand stimulation (fig. 59B, C). The obtained distribu-

tions can be compared to distributions of randomly distributed clusters within the cell area (fig. 59D). For multi-target experiments, the pipeline offers flexibility in designating which channels serve as center and which as neighbor coordinates (fig. 59E).

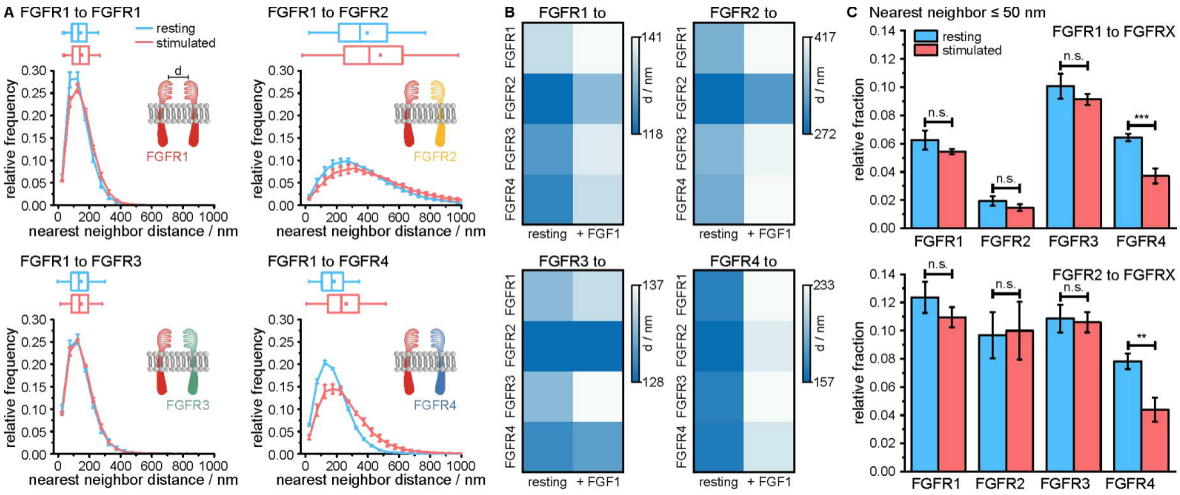

Figure 60: Distance analysis of FGFR1-4 within untreated and FGF1-treated U-2 OS cells. A Nearest neighbor distances are displayed between FGFR1 and FGFR1-4 for untreated cells (blue) and treated cells (red). The boxplots display the 1st, 2nd and 3rd quartile, as well as the mean values. The whiskers depict standard deviations. **B** Median nearest neighbor distances of the receptors in resting and FGF1-treated condition. **C** Fractions of receptors within a short-range distance of $\leq$ 50 nm. The nearest neighbor distances of FGFR1 (top) and FGFR2 (bottom) to FGFR1-4 is shown, N = 7 (resting), N = 5 (stimulated), n.s.: no significant, **: very significant, ***: highly significant difference, (Schröder et al., 2021).

9.1.2 Spatial organization of the FGFR network

The four receptors FGFR1-4 were analyzed and receptor densities and distributions compared between untreated and treated U-2 OS bone osteosarcoma cells. For treatment, FGF1 and heparin were added for 5 min prior to fixation. FGF1 is known to interact with all four FGFRs and induces receptor dimerization and signal transduction (Wheeler et al., 2015). Heparin is a co-factor required for ligand-receptor complex formation (Yayon et al., 1991). In order to extract how FGFR1-4 rearrange upon stimulation, nearest neighbor distances were calculated between FGFR1 and FGFR1-4 for both conditions (fig. 60A) (Schröder et al., 2021). Median distances are shown for all receptor combinations (fig. 60B). Upon ligand treatment, the receptor densities of FGFR1, FGFR3, and FGFR4 decreased, while there was no significant change in receptor densities for FGFR2 (fig. S14). The FGFR1 distance to neighboring FGFR2 and FGFR4 increased. Minute changes were observed for distances between FGFR1 to FGFR1 and FGFR3. To specifically look at short-distance interactions, the fraction of nearest neighbor distances $\leq$ 50 nm were quantified (fig. 60C). From that analysis it was found that FGFR1-3 in relation to neighboring FGFR4 revealed a significant reduction in the relative fraction of nearest neighbors within a 50 nm radius upon ligand

stimulation. Conversely, other receptor combinations exhibited minor decreases.
The overall increasing distances after ligand treatment can be explained by the decreasing number of receptors on the cell surface (fig. S14). To reduce FGFR signaling, the cell internalizes the activated receptors and either degrades them or recycles them as monomers back to the cell membrane (Aucilleo et al., 2013). The distance distributions of FGFR1 to FGFR1 and FGFR3 exhibit minimal shifts toward higher distances upon ligand stimulation, despite a significant reduction in their receptor densities (Schröder et al., 2021). This might be due to increased interactions between the receptors, which counteracts the impact of decreased receptor densities induced by ligand treatment. The reduced number of short-ranged FGFR4 neighbors hints at fast activity regulation of FGFR4 via internalization. It is known that FGFR1 is degraded in cells, whereas FGFR4 is recycled to the cell surface and replaced by monomers when stimulated (Haugsten et al., 2005).

RTKs are internalized via a clathrin-mediated pathway or through caveolae. It is established that FGFRs are recycled primarily through clathrin-mediated endocytosis (Aucilleo et al., 2013). To investigate alterations in receptor density within membrane regions where clathrin is present, circular count analysis was conducted. FGFR1 clusters were regarded as neighbors and clathrin clusters as centers. FGF1-treated and untreated conditions were compared to check how the receptor density changes after ligand treatment in membrane areas where clathrin is present (fig. 61). More receptors were near clathrin clusters upon activation. However, no significant difference was observed and more experiments might be required to analyze the FGFR1 distribution along clathrin.

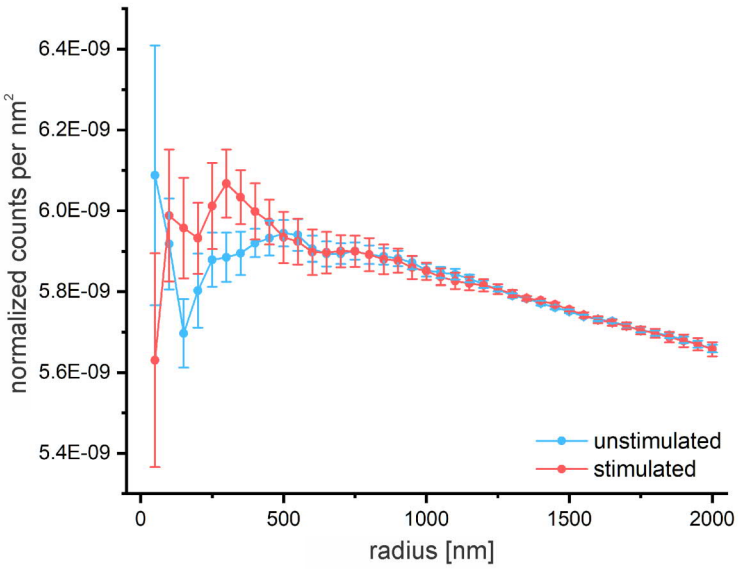

Figure 61: Circular counts of clathrin and FGFR1 in untreated and FGF1-treated U-2 OS cells. The unstimulated condition is shown in blue (N = 10) and the stimulated condition in red (N = 9). The error bars depict the SEM.

To investigate possible co-organization over larger distances up to 1 µm, k-nearest neighbor distances of the first 15 neighbors were determined (Schröder, 2021). The distances between FGFR1 to FGFR1-4 were calculated for experimental data and compared to simulated random cluster distributions. Especially for the first four neighbors, the experimental data exhibited higher frequencies, suggesting the possibility of non-random co-organization at short distances. Thereafter the curve is similar to simulations and decreases faster for long-ranged distances. This behavior was most pronounced for FGFR1-FGFR1, followed by FGFR1-FGFR2 and FGFR1-FGFR3. For FGFR1-FGFR4, the experimental curve runs above the simulated curve over a distance up to ∼800 nm.
The higher receptor count at close proximity up to 200 nm could be attributed to the potential clustering of receptors along cellular structures like heparan sulfate proteoglycans (HSPGs) (Ori et al., 2008; Sarrazin et al., 2011) or clathrin-coated pits (Li et al., 2016), as the spatial dimensions align with the know dimensions of these structures. To investigate this, additional experiments of simultaneous labeling of these cellular structures and the receptor network would be necessary.

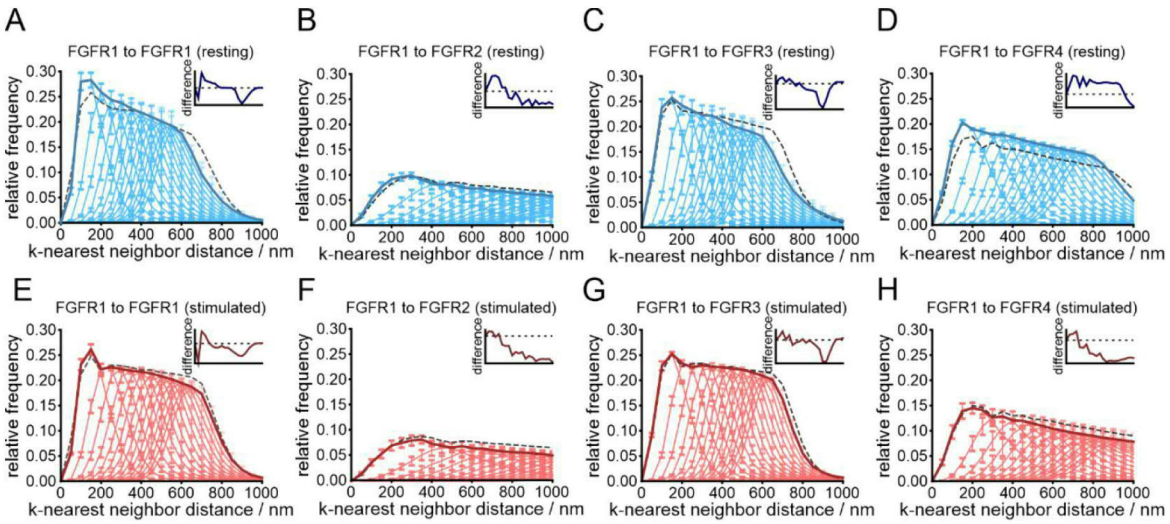

Figure 62: k-nearest neighbor distances of FGFR1 to FGFR1-4 within untreated and FGF1-treated U-2 OS cells. A-H the first 15 nearest neighbor distributions are shown for untreated cells (blue) and treated cells (red) with the error bars depicting SEMs. The hull curve marks the maxima of the nearest neighbor distributions. The inserts show the differences of the experimental to a simulated distribution, which assumes randomly distributed emitters, N = 7 (resting), N = 5 (stimulated), (Schröder, 2021).

The here demonstrated analyses describe the FGFR network with molecular resolution on the membrane of U-2 OS cells. The co-organization from a nanometer to a micrometer scale was revealed and is a step towards understanding protein interactions. The analysis package can be applied straight forward to other protein networks. The comparison to randomly placed coordinates within the cell areas emphasizes the non-random effects in the distributions, also shown in other studies (Barth et al., 2020). Here, the authors described

the k-nearest neighbor distances between segmented blobs of chromatin, revealing multiple weak maxima in the distance analysis compared to the simulated distribution. This hints towards higher degrees of spatial organization within the chromatin. The analysis of the FGFR network extends this to multiple targets. An alternative approach in describing protein assemblies is the extraction of cluster features, such as the density, overlap and distance to other protein clusters, followed by dimensionality reduction and clustering. Unterauer et al., 2024 demonstrated that approach on 30 labeled proteins in the synapse and were able to distinguish between inhibitory, mixed and excitatory synapses in the dimensionality reduced space. This is useful if different protein assemblies are expected within a condition. As the membrane is a heterogeneous surface, studying the organization in context of membrane constitution is highly relevant (Aucilleo et al., 2013; Haugsten et al., 2016). Different proteins of interest could be labeled, for example the HSPGs as mentioned above. Another approach of visualizing the constitution is electron microscopy, which allows segmentation of specific regions within cells or tissues. Simultaneous super-resolution microscopy of the same region allows the determination of receptor cluster densities across various cellular domains (Alfonzo-Méndez et al., 2022; Boer et al., 2015).

9.2 Methods

The spatial organization of the receptor tyrosin kinases FGFR1-4 were studied in U-2 OS cells using DNA-PAINT. A detailed description of the methods of the FGFR1-4 measurements can be found in Schröder et al., 2021. The FGFR1 and clathrin dataset was prepared analogously. Differences are highlighted in the respective sections.

9.2.1 Cell preparation and labeling

FGFR1-4

FGFR1-4 were labeled via direct immunofluorescence. In brief, primary antibodies against FGFR1-4 were conjugated to DNA-docking strands (anti-FGFR1 antibody (Abcam, Germany, #ab829) with the docking strand P9 (5' TTTAATAAGGT 3'), anti-FGFR2 (Abnova, Taiwan, #PAB3041) with P3* (5' GAAGTAATG 3'), anti-FGFR3 (Atlas antibodies AB, Sweden, #HPA067204) with P5 (5' TTTCAATGTAT 3'), and anti-FGFR4 (Abcam, #ab5481) with P1 (5' TTATACATCTA 3')). U-2 OS cells (CLS Cell Lines Service GmbH, Germany) were fixed with 4% methanol-free formaldehyde (Sigma-Aldrich), 0.1% glutaraldehyde (Sigma-Aldrich), and 0.4 M sucrose in 1x PBS. For receptor activation, 100 nM FGF1 (PeproTech, Germany, #AF-100-17A) and 15 µM heparin (Sigma-Aldrich, #H3149) were added for 5 min prior to fixation. 100 µg/mL anti-FGFR1-P9, 50 µg/mL anti-FGFR2-P3*, 12 µg/mL anti-FGFR3-P5, and 20 µg/mL anti-FGFR4-P1 were incubated for 2 hours in blocking buffer (2.5% IgG free bovine serum albumin (BSA) (Sigma-Aldrich) in PBS). The samples were post-fixated with 4% formaldehyde in PBS for 10 min. The samples were supplemented with 125 nm gold beads (Nanopartz, USA) serving as fiducial markers. Shortly before the DNA-PAINT measurements, the imager strands P1, P3*, P5, or P9, labeled with the organic fluorophore ATTO 655, were diluted to a final concentration of 4 nM in an imaging buffer composed of 500 mM NaCl (Sigma-Aldrich) and 75 mM $MgCl_2$ (Sigma-Aldrich) in PBS.

FGFR1 and clathrin

For imaging the spatial organization of FGFR1 and clathrin, clathrin was transfected as fusion protein with mEos2. PALM measurements were executed with a 568 nm and 405 nm laser. Mouse anti-FGFR1 primary antibody (Abcam, Germany, #ab829) and goat anti-mouse-P1 secondary antibody was used for FGFR1 labeling. 3 nM imager strand (P1-ATTO 655) was added for the DNA-PAINT measurements as described above, as well as the fixation protocol and stimulation with FGF1 and heparin.

9.2.2 DNA-PAINT data acquisition

FGFR1-4

The measurements were conducted using an N-STORM microscope (Nikon, Japan). ATTO 655, conjugated to imager strands, was excited with a 647 nm laser with an intensity of 0.6 kW/cm^2. The microscope was operated with an oil immersion objective (Apo, 100x, NA 1.49) in TIRF mode. Images of 256x256 px^2 with a pixel size of 158 nm were recorded. Movies were captured with an EMCCD camera (DU-897U-CS0-#BV, Andor Technology, Ireland) using an EM gain of 150. The camera integration time was set to 150 ms and 20,000 frames per movie were acquired. All four FGF receptors were sequentially recorded for the same cell. Between each measurement, the chamber underwent ten washing steps with PBS to ensure the complete removal of previously used imager strands. 7 cells were recorded for the resting and 5 cells for the stimulated condition.

FGFR1 and clathrin

30,000 frames were recorded for the PALM measurements of clathrin with a camera integration time of 100 ms, an EM gain of 200. The DNA-PAINT imaging of FGFR1 was analogous to the FGFR1-4 measurements. 10 cells were acquired for the resting and 9 cells for the stimulated condition. The mEos fluorophore was excited with a 561 nm and 405 nm laser.

9.2.3 Localization and clustering

FGFR1-4

DNA-PAINT movies were processed using *Picasso* (Schnitzbauer et al., 2017). Single-molecule localizations were extracted with the following settings: baseline = 175 photons, sensitivity = 4.78, quantum efficiency = 0.9, and integrated Gaussian MLE fit. The minimum net gradient was set to 53,000, 50,000, 53,000, and 52,000 for FGFR1, FGFR2, FGFR3, and FGFR4, respectively. The movies were drift corrected and channels aligned. Signal outside the cell was excluded. Localizations were filtered, only keeping localizations with a standard deviation of 0.6 to 1.2 pixels, to reduce signal from other planes. Localizations were linked within a radius of four times NeNA (Endesfelder et al., 2014) with a maximum of 20 dark frames. Subsequently, DBSCAN (Ester et al., 1996) was applied with a radius set to two times the NeNA value and a minimum density of 7 localizations. To minimize artifacts and background signals, clusters were filtered based on their mean frame time (between 1500 frames and 8000 frames and within $\pm 2 \cdot$std).

FGFR1 and clathrin

The minimum net gradient was set to 50,000 for FGFR1 and 30,000 for clathrin. DBSCAN parameters were analogous to FGFR1-4 for FGFR1. The minimum density was set to 15 localizations for clathrin.

9.2.4 Distance analysis

Distance analysis was executed with *SpOmKit*[B] (Schröder et al., 2021). The x- and y-coordinates of the clusters from DBSCAN analysis in *Picasso* (Schnitzbauer et al., 2017) were used as positions in the distance analysis. The relative frequencies of nearest neighbor distances between FGFR1-4 as central points and FGFR1-4 as neighboring points were calculated, resulting in 16 distance distributions per measurement. The relative fractions of nearest neighbor distances between centers and neighbors that are less than or equal to 50 nm were determined. k-nearest neighbor analysis was applied with up to 15 neighbors. A difference plot between experimental and simulated curves was generated by subtracting the experimental from the simulated curve. Circular counts were executed with a maximum radius of 2000 nm and intervals of 50 nm. All analysis were repeated with simulated random cluster distributions, using the same number of clusters and cell shapes as the experimental data. 10 simulations were averaged to generate one simulated curve.

9.2.5 Statistical analysis

The distributions were tested against a normal distribution using the Kolmogorov-Smirnov test ($\alpha = 0.05$), and all were found to be normally distributed. Differences in distributions between untreated and treated cells were tested with two-sampled t-tests in *OriginPro* (Origin Lab) with levels of significance were classified as: $p \geqslant 0.05$ no significance, $p < 0.05$ significant difference (*), $p < 0.01$ very significant difference (**), $p < 0.001$ highly significant difference (***).

[B] https://github.com/HeilemannLab/SpOmKit

10 Conclusion and outlook

In this study the application of deep learning to accelerate SMLM experiments, enable fast and long-time STED imaging, and automated fluorescence staining of wide-field images was demonstrated. Furthermore, methods for characterizing mobility patterns and the spatial organizations of membrane receptors were extended and demonstrated on various targets. The following section briefly summarizes the key goals of each project and gives an outlook. A global conclusion is drawn with regards to deep learning in the field of super-resolution microscopy.

10.1 Accelerated DNA-PAINT imaging with a neural network

DNA-PAINT imaging (Jungmann et al., 2010) has been substantially accelerated with *DeepSTORM* (Nehme et al., 2018), a neural network based approach that obtains super-resolved images under the challenging condition of high emitter density (Narayanasamy et al., 2022). A workflow for acquiring DNA-PAINT training and test data from experimental measurements was established. Emitter localizations were retrieved with a post-processing step. The results were scrutinized with various quality control metrics. This enabled imaging of α-tubulin and TOM20 in MNTB tissue within one minute instead of 25 minutes per target protein.

The model exhibited enhanced accuracy in predicting 1D structures compared to 2D structures. Even though DNA-PAINT offers a constant labeling density due to the exchange of labels on the target structure, local density variations due to the structure of the object may occur. The capability of the network in handling different local densities could be improved by including a larger variety of densities within the training data. A more complex approach would entail the training of multiple models for specific densities and applying them to local patches based on the expected density of the patch. A summation of intensities in the diffraction-limited patch through time could estimate the density of emitters within the patch. Depending on the biological question, 3D information might be required, which would necessitate the use of 3D capable models such as *DeepSTORM3D* (Nehme et al., 2020) or *DECODE* (Speiser et al., 2021).

The here demonstrated workflow was successfully applied to measurements using exchangeable HaloTag ligands (xHTLs) (Kompa et al., 2023), further extending the application capability of *DeepSTORM*. Jang et al., 2023 processed high density SMLM measurements with *DeepSTORM* (Nehme et al., 2018) followed by *DBlink* (Saguy et al., 2023), a neural network generating continuous structures from patchy localization movies. This enabled the observation of the endoplasmic reticulum in a living cell for 12 minutes with a temporal resolution of 1 second. To further increase the temporal resolution, SMLM movies with rapid camera integration could be acquired, which would result in low SNR images. Local-

ization of emitters with low SNRs is challenging, also for deep learning based approaches (Speiser et al., 2021). As a step prior to localization, the signals could be recovered by a denoising approach and then processed by *DeepSTORM* and *DBlink* to generate a movie with continuous structure. This would increase the temporal resolution, crucial for capturing the dynamics in living cells.

10.2 Fast and long-time STED imaging of ER nano-structural dynamics in living cells

In this project, STED images of the ER were denoised with a UNet-RCAN network (Ebrahimi et al., 2023), an approach especially suitable for preserving the high frequency features of the images. This enabled the observation of the ER for several hours with a high spatiotemporal resolution of 90 nm and a pixel dwell time of 0.5 µs. The segmentation of ER structures utilizing *ERnet* (Lu et al., 2023), as well as additional descriptors, allowed for the characterization of drug-induced effects influencing the ER-phagy machinery (Rahm et al., 2024).

Although the quality of the predicted images suffices for quantifying the drug effects, further quality enhancements of the predictions might be required for other research inquiries. Incorporating more information for the denoising task, such as the inclusion of adjacent frames in a live cell movie to denoise the central frame, would potentially boost the performance. Employing perceptual based loss functions can further improve the denoising capability (Liu et al., 2023; Zhao et al., 2015). Additionally, refining the training target through deconvolution can sharpen STED images.
To draw a more holistic picture of the cell, simultaneous imaging of multiple organelles at high spatiotemporal resolutions over long periods of time would enable the observation of dynamic interactions and interdependencies between cellular compartments. For this, the organelles would be recorded in parallel with orthogonal dyes and low laser intensity. The channels would be denoised with models trained for the specific organelle, similarly as in Weigert et al., 2018. To increase the multiplexing capability, organelles that are discernible in a bright-field image could be virtually stained, by models trained to translate label-free images to fluorescence images. Given the rapid acquisition of bright-field images, virtual staining ensures the preservation of high temporal resolution. Organelles like the nucleus, cell membrane, and potentially mitochondria could be suitable targets for virtual staining. Visualizing multiple organelles would allow to study inter-organelle contact sites that have diverse functions. One prominent example are contact sites between the ER and mitochondria (Sassano et al., 2022) that serve as hotspots for fission and fusion events of mitochondria (Abrisch et al., 2020).

10.3 Virtual staining of cellular landmarks

A model has been trained with robust capabilities for virtually staining the nucleus and membrane (label-free to fluorescence) across various experimental conditions (Liu et al., 2023). The key aspect of this generalizability was data augmentation that covered the sample space. Furthermore, the nuclei were instance segmented with *cellpose* (Stringer et al., 2021). As segmentation networks are often not compatible with quantitative phase images, taking an indirect route of translating the channel to fluorescence and then segmenting the image prevents the necessity of training a model from scratch, where vast amounts of hand-labeled annotations would have been required to achieve instance segmentation.

As virtual staining is not occupying fluorescence channels, the multiplexing capabilities are enhanced. Furthermore, the segmentation of cell nuclei and membranes facilitates the extraction of cell properties such as the volume and shape, enabling quantification of organelle properties in high-throughput experiments.
In further work, the model performance has been optimized by incorporating 3D information and using a hybrid loss function of an MAE mixed with the perceptually based MS-SSIM (Liu et al., 2023). The model demonstrated transferability to a different type of microscope and was utilized to virtually label the membrane and nucleus (still remaining in the same cell line) (Ivanov et al., 2023). Jo et al., 2021 demonstrated virtual labeling of various organelles across different cell types and the segmentation tool *cellpose* also generalizes well regarding cell types and microscopes. An essential factor contributing to such generalization besides data augmentation is the pooling of vast amounts of images. In the case of *cellpose*, training data was pooled from diverse sources, even including segmented images of stones and shells. Robustly segmenting the nucleus and membrane will enable the automated identification of cell instances and the quantification of their properties across various microscopes and conditions.

10.4 A pipeline for the analysis of single-particle tracking data

A comprehensive framework was developed to enhance the efficiency of analyzing SPT data and a new method of analysis was introduced. It comprises mean-squared displacement analysis to calculate diffusion coefficients of the single molecules and allows distinguishing between immobile, confined, and freely diffusing molecules. Occurrences of transitions between different diffusion modes within a trajectory are characterized with transition counting and hidden MARKOV modeling. The methodology was applied to single-molecule trajectories of the receptor tyrosine kinases MET (Rahm et al., 2021) and HER2 (Catapano et al., 2023) to study their responses to ligands in HeLa cells.

Extending the pipeline to multi-color SPT (Subach et al., 2010; You et al., 2016) would allow the detection of receptor complexes, as RTKs are known to dimerize and heterodimer-

ize (Harwardt et al., 2020; Paul et al., 2019). These formations could relate to changes in diffusion characteristics.
Furthermore, tracking the target(s) over an extended window of time would allow to observe ligand responses or effects of drug treatments within a single cell. Typical response times of RTKs upon ligand treatment are several minutes (Catapano et al., 2023). Denoising could easily extend the measurable duration of a SPT movie while preventing photobleaching of the label and preventing phototoxicity for the cell. The analysis of spatiotemporal hotspots are a further exciting development that might point to dynamic protein interactions as nanoclusters on the cell membrane (Wallis et al., 2023).

10.5 Spatial organization of membrane receptors

To investigate the spatial organization of membrane receptors, a data pipeline was established, facilitating the extraction of receptor densities and distributions in multi-target SMLM images. The pipeline was employed to study the FGFR network. Influences of the FGF1 ligand on the spatial arrangement of the FGFR network on the cell membrane were studied (Schröder et al., 2021).

Given the inherent heterogeneity of the cell membrane, studying the organization in context of membrane constitution is highly relevant (Aucilleo et al., 2013; Haugsten et al., 2016). Clustering of receptors could be related to structures such as the HSPGs, which modulates FGF-FGFR complex formation (Ori et al., 2008; Sarrazin et al., 2011), and could be visualized along the FGF receptors. Another approach of visualizing the constitution is electron microscopy, which allows segmentation of specific regions within cells or tissues. Simultaneous super-resolution microscopy of the same region allows the determination of receptor cluster densities across various cellular domains (Alfonzo-Méndez et al., 2022; Boer et al., 2015).
The acquisition time of the experiments could be sped up by increasing the density and using deep learning based localization approaches (Narayanasamy et al., 2022; Nehme et al., 2018). This is especially relevant, if many targets should be imaged as they all add to the duration of the experiment.

10.6 Deep learning for super-resolution microscopy in the future

The techniques explored herein address three critical facets in advanced microscopy, namely the reduction in image acquisition time, saving photon budget during measurements, and increasing the multiplexing capability. Beyond the solutions presented, a multitude of alternative approaches are available. Reducing the photon budget can be achieved through cross-modality translation (Bouchard et al., 2023; Chen et al., 2021), by interpolating imaging frames (Priessner et al., 2024), or event triggered microscopy (Alvelid et al., 2022). The

multiplexing capability can be enlarged with spectral demixing (Jiang et al., 2023; Kim et al., 2019b) or potentially by classifying the intensity-time traces stemming from different imager strands in DNA-PAINT measurements. Choosing the appropriate tools is contingent upon on the available data and the specific biological question at hand. Potential relies in synergistically combining multiple neural network tools to further advance the frontiers of super-resolution microscopy, all while establishing reliable methods for quality control checks.

To generate robust models across various experimental conditions, microscopes, and samples, substantial data availability is needed. Moreover, the publication of trained models is imperative to ensure the reproducibility of experiments (Editorial, 2021). Platforms such as `https://zenodo.org/` and `https://bioimage.io/` (Ouyang et al., 2022) simplify the sharing and finding of both data and models. Additionally, pre-trained models may suffice for specific tasks or require minimal retraining, thus mitigating the demands on data and computational resources.

The energy-intensive nature of neural network training raises concerns about its environmental impact (Strubell et al., 2023; Verdecchia et al., 2023). Techniques such as network pruning, which removes non-essential parameters without significantly affecting performance, are an active field of research in the computer science community (Frankle et al., 2019; Li et al., 2017), warranting increased attention for bioimage analysis (Zhou et al., 2024).

The realization of a smart microscope, capable of autonomous decision-making regarding when, where, what, and how to image, was never that tangible (Daetwyler et al., 2023). Current efforts incorporate event-based detections for switching imaging modes and the automated optimization of imaging parameters (Alvelid et al., 2022; Mahecic et al., 2022; McDole et al., 2018). With the emergence of large language models (LLMs), the ability to issue voice commands to the microscope, such as the instruction "N measurements under a specific condition", could be just around the corner.

11 Author contributions

Except where stated otherwise by reference or acknowledgment, the work presented was generated by myself under the supervision of my advisors during my doctoral studies. All contributions from colleagues are explicitly referenced in the thesis. The sections below clarify the collaborative research.

Accelerated DNA-PAINT imaging with a neural network
DNA-PAINT data was acquired by K. K. Narayanasamy. The workflow of how to process experimental DNA-PAINT data for *DeepSTORM* and generating appropriate training and test data was proposed by me, along with the code *ImageSumming*, which was executed by K. K. Narayanasamy. Models were trained by me, K. K. Narayanasamy and S. Tourani. The final SEM model was trained by me and the final MEM model by S. Tourani. Postprocessing and parameter optimization was executed by me. Model performance was evaluated by both K. K. Narayanasamy and me. Figure 30, 33, 35, 34, and 36 were created by K. K. Narayanasamy and adapted by me. Exchangeable HaloTag data acquisition and analysis was executed by S. Jang.

Fast and long-time STED imaging of ER nano-structural dynamics in living cells
STED images were acquired by A. Balakrishnan and drugs proposed to investigate ER-phagy. Data preprocessing was executed by A. Balakrishnan and me. UNet-RCAN code adjustments and the grid search was enabled by me, as well as the hyperparameter importances and detection of hallucination events. *ERNet* results were gathered by me. Fig. 38, 39, 40, S3, S4, and S5 were created by A. Balakrishnan and adapted by me.

Virtual staining of cellular landmarks
The image acquisition, data preprocessing, model training, and evaluation was were performed by me in the robust virtual staining project. The endoplasmic reticulum data was measured by M. Glogger and analyzed by me.

A pipeline for the analysis of single-particle tracking data
The MET dataset was acquired by M.-L. I. E. Harwardt and the HER2 dataset by C. Catapano. The analysis pipeline was established by me. S. Malkusch wrote the hidden MARKOV modeling code, which was integrated into *SPTAnalyser* by me. All MET analyses were conduced by me. Figures about the MET receptor were created by M. S. Dietz and me. C. Catapano executed the analysis for HER2 and created figure 58.

Spatial organization of membrane receptors
The FGFR dataset was acquired by M. Schröder. The analysis pipeline was realized by me and execute by M. Schröder. Fig. 60, S14, and 62 were created by M. Schröder.

12 References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, et al. (2016). "TensorFlow: Large-scale machine learning on heterogeneous distributed systems". In: *arXiv*. DOI: 10.48550/arXiv.1603.04467.

Abbe, E. (1873). "Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung". In: *Archiv f. mikrosk. Anatomie* 9, pp. 413–418. DOI: 10.1007/BF02956173.

Abdullah-Al-Mamun, M., V. Tyagi, and H. Zhao (2021). "A New Full-Reference Image Quality Metric for Motion Blur Profile Characterization". In: *IEEE Access* 9, pp. 156361–156371. DOI: 10.1109/ACCESS.2021.3130177.

Abrisch, R. G., S. C. Gumbin, B. T. Wisniewski, L. L. Lackner, and G. K. Voeltz (2020). "Fission and fusion machineries converge at ER contact sites to regulate mitochondrial morphology". In: *J. Cell. Biol.* 4, e201911122. DOI: 10.1083/jcb.201911122.

Agasti, S. S, Y. Wang, F. Schueder, A. Sukumar, R. Jungmann, et al. (2017). "DNA-barcoded labeling probes for highly multiplexed Exchange-PAINT imaging". In: *Chem. Sci.* 8, p. 3080. DOI: 10.1039/C6SC05420J.

Airy, G. B. (1895). "On the Diffraction of an Object-glass with Circular Aperture ". In: *Trans. Cambridge Philos. Soc* 5, p. 283.

Alfonzo-Méndez, M. A., K. A. Sochacki, M.-P. Strub, and J. W. Taraska (2022). "Dual clathrin and integrin signaling systems regulate growth factor receptor activation". In: *Nat. Comm.* 13, p. 905. DOI: 10.1038/s41467-022-28373-x.

Alvelid, J., M. Damenti, C. Sgattoni, and I. Testa (2022). "Event-triggered STED imaging". In: *Nat. Methods* 19, pp. 1268–1275. DOI: 10.1038/s41592-022-01588-y.

Ambrose, E. J. (1956). "A Surface Contact Microscope for the study of Cell Movements". In: *Nature* 178, p. 1194. DOI: 10.1038/1781194a0.

Andronov, L., I. Orlov, Y. Lutz, J.-L. Vonesch, and B. P. Klaholz (2016). "ClusterViSu, a method for clustering of protein complexes by Voronoi tessellation in super-resolution microscopy". In: *Sci. Reports* 6, p. 24084. DOI: 10.1038/srep24084.

Aucilleo, G., D. L. Cunningham, T. Tatar, J. K. Heath, and J. Z. Rappoport (2013). "Regulation of fibroblast growth factor receptor signalling and trafficking by Src and Eps8". In: *J. Cell Sci.* 126, pp. 613–624. DOI: 110.1242/jcs.116228.

Axelrod, D. (1981). "Cell-substrate contacts illuminated by total internal reflection fluorescence". In: *J. Cell Biol.* 89, pp. 141–145. DOI: 10.1083/jcb.89.1.141.

Bag, N., S. Huang, and T. Wohland (2015). "Plasma Membrane Organization of Epidermal Growth Factor Receptor in Resting and Ligand-Bound States". In: *Biophys. J.* 109, pp. 1925–1936. DOI: 10.1016/j.bpj.2015.09.007.

Balestriero, B., L. Bottou, and Y. LeCun (2022). "The Effects of Regularization and Data Augmentation are Class Dependent". In: *arXiv*. DOI: 10.48550/arXiv.2204.03632.

Balzarotti, F., Y. Eilers, K. C. Gwosch, A. H. Gynnå, V. Westphal, et al. (2016). "Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes". In: *Science* 355, pp. 606–612. DOI: 10.1126/science.aak991.

Banwell, C. N. (1983). *Fundamentals of Molecular Spectroscopy*. 4th ed. Berkshire: McGraw-Hill International. DOI: 10.1002/ange.19901020643.

Barer, R. (1952). "Interference microscopy and mass determination". In: *Nature* 169, pp. 366–367. DOI: 10.1038/169366b0.

Barnard, T. J., X. Yu, N. Noinaj, and J. W. Taraska (2014). "Crystal Structure of Green Fluorescent Protein". In: available online.

Barth, R., K. Bystricky, and H. A. Shaban (2020). "Coupling chromatin structure and dynamics by live super-resolution imaging". In: *Sci. Adv.* 6, eaaz2196. DOI: 10.1126/sciadv.aaz2196.

Beer, A. (1852). "Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten". In: *Annalen der Physik* 162, pp. 78–88. DOI: 10.1002/andp.18521620505.

Belthangady, C. and L. A. Royer (2019). "Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction". In: *Nat. Methods* 16, pp. 1215–1255. DOI: 10.1038/s41592-019-0458-z.

Berkane, R., H. Ho-Xuan, M. Glogger, P. Sanz-Marinez, L. Brunello, et al. (2023). "The function of ER-phagy receptors is regulated through phosphorylation-dependent ubiquitination pathways". In: *Nat. Comm.* 14, p. 8364. DOI: 10.1038/s41467-023-44101-5.

Berning, S., K. I. Willig, H. Steffens, P. Dibaj, and S. W. Hell (2012). "Nanoscopy in a Living Mouse Brain". In: *Science* 335, p. 551. DOI: 10.1126/science.1215369.

Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, et al. (2006). "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution". In: *Science* 313, pp. 1642–1645. DOI: 10.1126/science.1127344.

Bilmes, J., K. Asanovic, C.-W. Chin, and J. Demmel (1997). "Using PHiPAC to speed error back-propagation learning". In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process* 5, pp. 4153–4156. DOI: 10.1109/ICASSP.1997.604861.

Boer, P. de, J. P. Hoogenboom, and B. N. G. Giepmans (2015). "Correlated light and electron microscopy: ultrastructure lights up!" In: *Nat. Methods* 12, pp. 503–513. DOI: 10.1038/nMeth.3400.

Bouchard, C., T. Wiesner, A. Deschênes, A. Bilodeau, B. Turcotte, et al. (2023). "Resolution enhancement with a task-assisted GAN to guide optical nanoscopy image analysis and acquisition". In: *Na. Mach. Intell.* 5, pp. 830–844. DOI: 10.1038/s42256-023-00689-3.

Boulanger, J., C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, et al. (2010). "Patch-Based Nonlocal Functional for Denoising Fluorescence Microscopy Image Sequences". In: *IEEE Trans. Med. Imaging* 29, pp. 442–454. DOI: 10.1109/TMI.2009.2033991.

Boyd, N., E. Jonas, H. Babcockl, and B. Recht (2018). "DeepLoco: Fast 3D Localization Microscopy Using Neural Networks". In: *bioRxiv*. DOI: 10.1101/267096.

Buades, A., B. Coll, and J.-M. Morel (2005). "A non-local algorithm for image denoising". In: *CVPR* 2, pp. 60–65. DOI: 10.1109/CVPR.2005.38.

Buchholz, T.-O., M. Prakash, A. Krull, and F. Jug (2020). "DenoiSeg: Joint Denoising and Segmentation". In: *arXiv*. DOI: 10.48550/arXiv.2005.02987.

Catapano, C., J. V. Rahm, M. Omer, L. Teodori, J. Kjems, et al. (2023). "Biased activation of the receptor tyrosine kinase HER2". In: *Cell Mol. Life Sci.* 80, p. 158. DOI: 10.1007/s00018-023-04806-8.

Chames, P. and U. Rothbauer (2020). "Special Issue: Nanobody". In: *Antibodies* 9, p. 6. DOI: 10.3390/antib9010006.

Chamier, L. von, R. F. Laine, J. Jukkala, C. Spahn, D. Krentzel, et al. (2021). "Democratising deep learning for microscopy with ZeroCostDL4Mic". In: *Nat. Comm.* 12, p. 2276. DOI: 10.1038/s41467-021-22518-0.

Chapman, K. B., F. Filipsky, N. Peschke, M. Gelléri, V. Weinhardt, et al. (2023). "A comprehensive method to study the DNA's association with lamin and chromatin compaction in intact cell nuclei at super resolution". In: *Nanoscale* 15, pp. 742–756. DOI: 10.1039/d2nr02684h.

Charbonnier, P., L. Blanc-Feraud, G. Aubert, and M. Barlaud (1994). "Two deterministic half-quadratic regularization algorithms for computed imaging". In: *ICIP*.

Chen, J., H. Sasaki, H. Lai, Y. Su, J. Liu, et al. (2021). "Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes". In: *Nat. Methods* 18, pp. 678–687. DOI: 10.1038/s41592-021-01155-x.

Cho, Nathan H., Keith C. Cheveralls, Andreas-David Brunner, Kibeom Kim, André C. Michaelis, et al. (2022). "OpenCell: Endogenous tagging for the cartography of human cellular organization". In: *Science* 375, eabi6983. DOI: 10.1126/science.abi6983.

Chou, Y.-Y., N. S. Heaton, Q. Gao, P. Palese, R. Singer, et al. (2013). "Colocalization of Different Influenza Viral RNA Segments in the Cytoplasm before Viral Budding as Shown by Single-molecule Sensitivity FISH Analysis". In: *PLOS Pathog.* 9, pp. 226–231. DOI: 10.1371/journal.ppat.1003358.

Christiansen, E. M., S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, et al. (2018). "In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images". In: *Cell* 173, pp. 792–803. DOI: 10.1016/j.cell.2018.03.040.

Chung, I., R. Akita, R. Vandlen, D. Toomre, J. Schlessinger, et al. (2010). "Spatial control of EGF receptor activation by reversible dimerization on living cells". In: *Nature* 464, pp. 783–787. DOI: 10.1038/nature08827.

Civitci, F., J. Shangguan, T. Zheng, K. Tao, M. Rames, et al. (2020). "Fast and multiplexed superresolution imaging with DNA-PAINT-ERS". In: *Nat. Comm.* 11, p. 4339. DOI: 10.1038/s41467-020-18181-6.

Condon, E. (1926). "A Theory of Intensity Distribution in Band Systems". In: *Phys. Rev.* 28, pp. 1182–1201. DOI: 10.1103/PhysRev.28.1182.

Coons, A. H., H. J. Creech, and R. N. Jones (1941). "Immunological Properties of an Antibody Containing a Fluorescent Group". In: *Proc. Soc. Exp. Biol. Med.* 47, pp. 200–202. DOI: 10.3181/00379727-47-13084p.

Costello, I. and S. Cox (2021). "Analysing errors in single-molecule localisation microscopy". In: *Int. J. Biochem. Cell Biol.* 134, p. 105931. DOI: 10.1016/j.biocel.2021.105931.

Culley, S., D. Albrecht, C. Jacobs, P. Matso Pereira, C. Leterrier, et al. (2018). "Quantitative mapping and minimization of super-resolution optical imaging artifacts". In: *Nat. Methods* 15, pp. 263–266. DOI: 10.1038/nmeth.4605.

Curry, H. B. (1944). "The Method of Steepest Descent for Non-linear Minimization Problems". In: *Quart. Appl. Math.* 2, pp. 258–261. DOI: 10.1090/qam/10667.

Daetwyler, S. and R. P. Fiolka (2023). "Light-sheets and smart microscopy, an exciting future is dawning". In: *Commun. Biol.* 6, p. 502. DOI: 10.1038/s42003-023-04857-4.

Dauphin, Y. N., R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, et al. (2014). "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization". In: *arXiv*. DOI: 10.48550/arXiv.1406.2572.

Dempsey, D. R., H. Jiang, J. H. Kalin, Z. Chen, and P. A. Cole (2018). "Site-Specific Protein Labeling with N-Hydroxysuccinimide-Esters and the Analysis of Ubiquitin Ligase Mechanisms". In: *J. Am. Chem. Soc.* 140, pp. 9374–9378. DOI: 10.1021/jacs.8b05098.

Dertinger, T., R. Colyer, G. Iyer, S. Weiss, and J. Enderlein (2009). "Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI)". In: *Proc. Natl. Acad. Sci. USA* 106, pp. 22287–22292. DOI: 10.1073/pnas.0907866106.

Descloux, A., K. S. Grußmayer, and A. Radenovic (2019). "Parameter-free image resolution estimation based on decorrelation analysis". In: *Nat. Methods* 16, pp. 918–924. DOI: 10.1038/s41592-019-0515-7.

Dice, L. R. (1945). "Measures of the Amount of Ecologic Association Between Species". In: *Ecology* 26, pp. 297–302. DOI: 10.2307/1932409.

Dietz, M. S., D. Haße, D. M. Ferraris, A. Göhler, H. H. Niemann, et al. (2013). "Single-molecule photobleaching reveals increased MET receptor dimerization upon ligand binding in intact cells". In: *BMC Biophys.* 6, p. 6. DOI: 10.1186/2046-1682-6-6.

Dietz, M. S. and M. Heilemann (2019a). "Optical super-resolution microscopy unravels the molecular composition of functional protein complexes". In: *Nanoscale* 11, pp. 17981–17991. DOI: 10.1039/c9nr06364a.

Dietz, M. S., S. S. Wehrheim, M.-L. I. E. Harwardt, H. Niemann, and M. Heilemann (2019b). "Competitive Binding Study Revealing the Influence of Fluorophore Labels on Biomolecular Interactions ". In: 19, pp. 8245–8249. DOI: 10.1021/acs.nanolett.9b03736.

Díez, L. T., C. Bönsch, S. Malkusch, Z. Truan, M. Munteanu, et al. (2014). "Coordinate-based co-localization-mediated analysis of arrestin clustering upon stimulation of the C-C chemokine receptor 5 with RANTES/CCL5 analogues". In: *Histochem. Cell Biol.* 1, p. 142. DOI: 10.1007/s00418-014-1206-1.

Dozat, T. (2016). "Incorporating Nesterov Momentum into Adam". In: *ICLR Workshop* 1, pp. 2013–2016.

Drozdzal, M., E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal (2016). "The Importance of Skip Connections in Biomedical Image Segmentation". In: *arXiv*. DOI: 10.48550/arXiv.1608.04117.

Dubey, S. R., S. K. Singh, and B. B. Chaudhuri (2021). "Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark". In: *arXiv*. DOI: 10.48550/arXiv.2109.14545.

Duchi, J., E. Hazan, and Y. Singer (2011). "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization". In: *J. Mach. Learn. Res.* 12, pp. 2121–2159. DOI: 10.5555/1953048.2021068.

Dumoulin, V. and F. Visin (2016). "A guide to convolution arithmetic for deep learning". In: *arXiv*. DOI: 10.48550/arXiv.1603.07285.

Dunbrack, R. L. (1986). "Calculation of Franck-Condon factors for undergraduate quantum chemistry". In: 63, p. 953. DOI: 10.1021/ed063p953.

Ebrahimi, V., T. Stephan, J. Kim, P. Carravilla, C. Eggeling, et al. (2023). "Deep learning enables fast, gentle STED microscopy". In: *Commun. Biol.* 6, p. 674. DOI: 10.1038/s42003-023-05054-z.

Edelstein, Arthur, Nenad Amodaj, Karl Hoover, Ron Vale, and Nico Stuurman (2010). "Computer Control of Microscopes Using μManager". In: *Curr. Protoc. Mol. Biol* 92, pp. 14.20.1–14.20.17. DOI: 10.1002/0471142727.mb1420s92.

Editorial, E. (2021). "Moving towards reproducible machine learning". In: *Nat. Comput. Sci.* 1, p. 1608.08710. DOI: 10.1038/s43588-021-00152-6.

Einstein, A. (1905). "Investigations on the Theory of the Brownian Movement". In: *New York: Dover Publications*. DOI: 10.1088/0031-9112/7/10/012.

Endesfelder, M., C. Schießl, B. Turkowyd, T. Lechner, and U. Endesfelder (n.d.). "swift – fast, probabilistic tracking for dense, highly dynamic single-molecule data". In: *Manuscript in preparation* ().

Endesfelder, U., S. Malkusch, F. Fricke, and M. Heilemann (2014). "A simple method to estimate the average localization precision of a single-molecule localization microscopy experiment". In: *Histochem Cell Biol.* 141, pp. 629–638. DOI: 10.1007/s00418-014-1192-3.

Ester, M., H. P. Kriegel, J. Sander, and X. Xu (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise". In: *KDD-96 Proceedings*, pp. 226–231. DOI: 10.5555/3001460.3001507.

Fick, A. (1855). "Ueber Diffusion". In: *Annalen der Physik* 94, pp. 59–86. DOI: 10.1002/andp.18551700105.

Fischer, L. S., C. Klingner, T. Schlichthaerle, M. T. Strauss, R. Böttcher, et al. (2021). "Quantitative single-protein imaging reveals molecular complex formation of integrin, talin, and kindlin during cell adhesion". In: *Nat. Comm.* 12, p. 2919. DOI: 10.1038/s41467-021-21142-2.

Franck, J. and E. G. Dymond (1926). "Elementary processes of photochemical reactions". In: *Trans. Faraday Soc.* 21, pp. 536–542. DOI: 10.1039/TF9262100536.

Frankle, J. and M. Carbin (2019). "The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks". In: *arXiv*. DOI: 10.48550/arXiv.1803.03635.

Fricke, F., J. Beaudouin, R. Eils, and M. Heilemann (2015). "One, two or three? Probing the stoichiometry of membrane proteins by single-molecule localization microscopy". In: *Sci Rep.* 11, p. 14072. DOI: 10.1038/srep14072.

Fukushima, K. (1980). "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position". In: *Biol. Cybernetics* 36, pp. 193–202. DOI: 10.1007/BF00344251.

Gautier, A., A. Juillerat, C. Heinis, I. R. Corrêa, M. Kindermann, et al. (2008). "An Engineered Protein Tag for Multiprotein Labeling in Living Cells". In: *Chem. Biol.* 15, pp. 128–136. DOI: 10.1016/j.chembiol.2008.01.007.

Gazagnes, S., E. Soubies, and L. Blanc-Feraud (2017). "High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation". In: *IEEE Int. Symp. Biomed. Imaging*. DOI: 10.1109/ISBI.2017.7950460.

Geman, S., E. Bienenstock, and R. Doursat (1992). "Neural networks and the bias/variance dilemma". In: *Neural Comput.* 4, pp. 1–58. DOI: 10.1162/neco.1992.4.1.1.

Giannone, G., E. Hosy, F. Levet, A. Constals, K. Schulze, et al. (2010). "Dynamic Superresolution Imaging of Endogenous Proteins on Living Cells at Ultra-high Density". In: *Biophys. J.* 99, pp. 1303–1310. DOI: 10.1016/j.bpj.2010.06.005.

Gibson, T. J., M. Seiler, and R. A. Veitia (2013). "The transience of transient overexpression". In: *Nat. Methods* 10, pp. 715–721. DOI: 10.1038/nmeth.2534.

Glogger, M., D. Wang, J. Kompa, A. Balakrishnan, J. Hiblot, et al. (2022). "Synergizing Exchangeable Fluorophore Labels for Multitarget STED Microscopy". In: *ACS Nano* 16, pp. 17991–17997. DOI: 10.1021/acsnano.2c07212.

Goncharova, A. S., A. Honigmann, F. jug, and A. Krull (2020). "Improving Blind Spot Denoising for Microscopy". In: *arXiv*. DOI: 10.48550/arXiv.2008.08414.

González, A., A. Covarrubias-Pinto, R. M. Bhaskara, M. Glogger, S. K. Kuncha, et al. (2023). "Ubiquitination regulates ER-phagy and remodelling of endoplasmic reticulum". In: *Nature* 681, pp. 394–401. DOI: 10.1038/s41586-023-06089-2.

Guo, S.-M., L.-H. Yeh, J. Folkesson, I. E. Ivanov, A. P. Krishnan, et al. (2020). "Revealing architectural order with quantitative label-free imaging and deep learning". In: *eLife* 9, e55502. DOI: 10.7554/eLife.55502.

Gustafsson, M. G. (2000). "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy". In: *J. Microsc.* 198, pp. 82–87. DOI: 10.1046/j.1365-2818.2000.00710.x.

Hajiabadi, H., I. Mamontova, R. Prizak, A. Pancholi, A. Koziolek, et al. (2022). "Deep-learning microscopy image reconstruction with quality control reveals second-scale rearrangements in RNA polymerase II clusters". In: *PNAS Nexus* 1, pp. 1–11. DOI: 10.1093/pnasnexus/pgac065.

Harke, B., P. Bianchini, G. Vicidomini, S. Galiani, and A. Diaspro (2013). "Stimulated Emission Depletion (STED) Microscopy". In: *Encyclopedia of Biophysics*, Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-16712-6_835.

Harwardt, M.-L. I. E., M. S. Schröder, Y. Li, S. Malkusch, P. Freund, et al. (2020). "Single-Molecule Super-Resolution Microscopy Reveals Heteromeric Complexes of MET and EGFR upon Ligand Activation ". In: *Int. J. Mol. Sci* 21, p. 2803. DOI: 10.3390/ijms21082803.

Harwardt, M.-L. I. E., P. Young, W. M. Bleymüller, T. Meyer, C. Karathanasis, et al. (2017). "Membrane dynamics of resting and internalin B-bound MET receptor tyrosine kinase studied by single-molecule tracking". In: *FEBS Open Bio* 7, pp. 1422–1440. DOI: 10.1002/2211-5463.12285.

Hasan, M. and M. R. El-Sakka (2018). "Improved BM3D image denoising using SSIM-optimized Wiener filter". In: *EURASIP J. Img. Vid. Proc.* 25. DOI: 10.1186/s13640-018-0264-z.

Hass, H., K. Masson an dS. Wohlgemuth, V. Paragas, J. E. Allen, M. Sevecka, et al. (2017). "Predicting ligand-dependent tumors from multi-dimensional signaling features". In: *NPJ Syst. Biol. Appl.* 3, p. 27. DOI: 10.1038/s41540-017-0030-3.

Haugsten, E. M., V. Sørensen, M. K. Bosakova, G. A. de Souza, P. Krejci, et al. (2016). "Proximity Labeling Reveals Molecular Determinants of FGFR4 Endosomal Transport". In: *J. Proteome Res.* 15, pp. 3841–3855. DOI: 10.1021/acs.jproteome.6b00652.

Haugsten, E. M., V. Sørensen, A. Brech, S. Olsnen, and J. Wesche (2005). "Different intracellular trafficking of FGF1 endocytosed by the four homologous FGF receptors". In: *J. Cell Sci.* 118, p. 17. DOI: 10.1242/jcs.02509.

He, K., X. Zhang, S. Ren, and J. Sun (2015). "Deep Residual Learning for Image Recognition". In: *IEEE CVPR*. DOI: 10.1109/CVPR.2016.906.

Heel, M. van, W. Keegstra, W. Schutter, and E. F. J. van Bruggen (1982). "Arthropod hemocyanin studies by image analysis". In: *Life Chem. Rep. Suppl.* 1, pp. 69–73.

Heilemann, M. (2010). "Fluorescence microscopy beyond the diffraction limit". In: *J. Biotechnol.* 149, pp. 243–251. DOI: 10.1016/j.jbiotec.2010.03.012.

Heilemann, M., S. van de Linde, M. Schüttpelz, R. Kasper, B. Seefeldt, et al. (2008). "Subdiffraction-Resolution Fluorescence Imaging with Conventional Fluorescent Probes". In: *Angew. Chem. Int. Ed.* 47, pp. 6172–6176. DOI: 10.1002/anie.200802376.

Heilemann, M., E. Margeat, R. Kasper, M. Sauer, and P. Tinnefeld (2005). "Carbocyanine Dyes as Efficient Reversible Single-Molecule Optical Switch". In: *J. Am. Chem. Soc.* 127, pp. 3801–3806. DOI: 10.1021/ja044686x.

Heilker, R., L. Zemanova, M. J. Valler, and G. U. Nienhaus (2005). "Confocal fluorescence microscopy for high-throughput screening of G-protein coupled receptors". In: *Curr. Med. Chem.* 12, pp. 2551–2559. DOI: 10.2174/092986705774370637.

Hein, B., K. I. Willig, and S. W. Hell (2008). "Stimulated emission depletion (STED) nanoscopy of a fluorescent protein-labeled organelle inside a living cell". In: *Proc. Natl. Acad. Sci.* 105, pp. 14271–14276. DOI: 10.1073/pnas.0807705105.

Hell, S. W. (2007). "Far-Field Optical Nanoscopy". In: *Science* 316, pp. 1153–1158. DOI: 10.1126/science.1137395.

Hell, S. W. and J. Wichmann (1994). "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy". In: *Opt. Lett.* 19, pp. 780–782. DOI: 10.1364/OL.19.000780.

Hershko, E., L. E. Weiss, T. Michaeli, and Y. Shechtman (2019). "Multicolor localization microscopy and point-spread-function engineering by deep learning". In: *Opt. Express* 27, p. 6158. DOI: 10.1364/OE.27.006158.

Hinton, G., N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2012). "Improving neural networks by preventing co-adaptation of feature detectors". In: *arXiv*. DOI: 10.48550/arXiv.1207.0580.

Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory". In: *Neural Comput.* 9, pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

Hornik, K., M. Stinchcombe, and H. White (1989). "Multilayer Feedforward Networks are Universal Approximators". In: *Neural netw.* 2, pp. 359–366. DOI: 10.1016/0893-6080(89)90020-8.

Hotta, J.-I., E. Fron, P. Dedecker, K. P. F. Janssen, C. Li, et al. (2010). "Spectroscopic Rationale for Efficient Stimulated-Emission Depletion Microscopy Fluorophores". In: *J. Am. Chem. Soc.* 132, pp. 5021–5023. DOI: 10.1021/ja100079w.

Huang, F., S. L. Schwartz, J. M. Byras, and K. A. Lidke (2011). "Simultaneous multiple-emitter fitting for single molecule super-resolution imaging". In: *Biomed. Opt. Express* 2, pp. 1377–1393. DOI: 10.1364/BOE.2.001377.

Hubicka, K. and J. Janczura (2020). "Time-dependent classification of protein diffusion types: A statistical detection of mean-squared-displacement exponent transitions". In: *Phys. Rev.* 101, p. 022107. DOI: 10.1103/PhysRevE.101.022107.

Ioffe, S. and C. Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *arXiv*. DOI: 10.48550/arXiv.1502.03167.

Ivanov, I. E., E. Hirata-Miyasaki, T. Chandler, R. C. Kovilakam, Z. Liu, et al. (2023). "Mantis: high-throughput 4D imaging and analysis of the molecular and physical architecture of cells". In: *bioRxiv*. DOI: 10.1101/2023.12.19.572435.

Jaccard, P. (1912). "The Distribution of the Flora in the Alpine Zone". In: *New Phytol.* 11, pp. 37–50. DOI: 10.1111/j.1469-8137.1912.tb05611.x.

Jang, S., K. N. Narayanasamy, J. V. Rahm, A. Saguy, J. Kompa, et al. (2023). "Neural network-assisted single-molecule localization microscopy with a weak-affinity protein tag". In: *Biophys. Reports* 3, p. 100123. DOI: 10.1016/j.bpr.2023.100123.

Jaqaman, K., D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, et al. (2008). "Robust single-particle tracking in live-cell time-lapse sequences". In: *Nat. Methods* 5, pp. 695–702. DOI: 10.1038/NMETH.1237.

Jiang, Y., H. Sha, S. Liu, P. Qin, and Y. Zhang (2023). "AutoUnmix: an autoencoder-based spectral unmixing method for multi-color fluorescence microscopy imaging". In: *bioRxiv*. DOI: 10.1101/2023.05.30.542836.

Jin, L., B. Liu, F. Zhao, S. Hahn, B. Dong, et al. (2020). "Deep learning enables structured illumination microscopy with low light levels and enhanced speed". In: *Nat. Comm.* 11, p. 1934. DOI: 10.1038/s41467-020-15784-x.

Jo, Y., H. Cho, W. S. Park, G. Kim, D. Ryu, et al. (2021). "Label-free multiplexed microtomography of endogenous subcellular dynamics using generalizable deep learning". In: *Nat. Cell Biol.* 23, pp. 1329–1337. DOI: 10.1038/s41556-021-00802-x.

Jungmann, R., M. S. Avendaño, J. B. Woehrstein, M. Dai, W. M. Shih, et al. (2014). "Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT". In: *Nat. Methods* 11, pp. 313–317. DOI: 10.1038/nmeth.2835.

Jungmann, R., C. Steinhauer, M. Scheible, A. Kuzyk, P. Tinnefeld, et al. (2010). "Single-Molecule Kinetics andSuper-Resolution Microscopy by FluorescenceImaging of Transient Binding on DNA Origami". In: *Nano Lett.* 10, pp. 4756–4761. DOI: 10.1021/nl103427w.

Karslake, J. D., E. D. Donarski, S. A. Shelby, L. M. Demey, V. J. RiRita, et al. (2021). "SMAUG: Analyzing single-molecule tracks with nonparametric Bayesian statistics". In: *Methods* 193, pp. 16–26. DOI: 10.1016/j.ymeth.2020.03.008.

Kasha, M. (1950). "Characterization of electronic transitions in complex molecules". In: *Discuss. Faraday Soc.* 9, p. 14. DOI: 10.1039/DF9500900014.

Kefer, P., F. Iqbal, M. Locatelli, J. Lawrimore, M. Zhang, et al. (2021). "Performance of deep learning restoration methods for the extraction of particle dynamics in noisy microscopy image sequences". In: *Mol. Biol. Cell* 32, pp. 903–914. DOI: 10.1091/mbc.E20-11-0689.

Keppler, A., M. Kindermann, S. Gendreizig, H. Pick, H. Vogel, et al. (2004). "Labeling of fusion proteins of O6 -alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro". In: *Methods* 32, pp. 437–444. DOI: 10.1016/j.ymeth.2003.10.007.

Khater, I. M., I. R. Nabi, and G. Hamarneh (2020). "A Review of Super-Resolution Single-Molecule Localization Microscopy Cluster Analysis and Quantification Method". In: *Patterns* 1, p. 100038. DOI: 10.1016/j.patter.2020.100038.

Kim, B., M. Han, H. Shim, and J. Baek (2019a). "A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose CT images". In: *Med. Phys.* 46, pp. 3906–3923. DOI: 10.1002/mp.13713.

Kim, T., S. Moon, and K. Xu (2019b). "Information-rich localization microscopy through machine learning". In: *Nat. Comm.* 10, p. 1996. DOI: 10.1038/s41467-019-10036-z.

Kim, Y. C. and K.-L. Guan (2015). "mTOR: a pharmacologic target for autophagy regulation". In: *J. Clin. Invest.* 125, pp. 25–32. DOI: 10.1172/JCI73939.

Kingma, D. P. and J. Ba (2014). "Adam: A Method for Stochastic Optimization". In: *arXiv*. DOI: 10.48550/arXiv.1412.6980.

Klar, T. A., S. Jakobs, M. Dyba, A. Egner, and S. W. Hell (2000). "Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission". In: *Proc. Natl. Acad. Sci.* 97, pp. 8206–8210. DOI: 10.1073/pnas.97.15.8206.

Klein, J. S., P. N. P. Gnanapragasam, R. P. Galimidi, C. P. Foglesong, A. P. West, et al. (2009). "Examination of the contributions of size and avidityto the neutralization mechanisms of the anti-HIV antibodies b12 and 4E10". In: *Proc. Soc. Exp. Biol. Med.* 106, pp. 7385–7390. DOI: 10.1073/pnas.0811427106.

Knuth, D. E. (1968). *The Art of Computer Programming*. 1st ed. available online. Boston: Addison-Wesley.

Köhler, A. (1893). "Ein neues Beleuchtungsverfahren für mikrophotographische Zwecke". In: *Z. Wiss. Mikrosk.* 10, pp. 433–440.

Kolb, H. C., M. G. Finn, and K. B. Sharpless (2001). "Click Chemistry: Diverse Chemical Function from a Few Good Reactions". In: *Angew. Chem. Int. Ed. Engl.* 40, pp. 2004–2021. DOI: 10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5.

Kölln, L. S., O. Salem, J. Valli, C. G. Hansen, and G. McConnell (2022). "Label2label: Training a neural network to selectively restore cellular structures in fluorescence microscopy". In: *J. Cell Sci.* 135, jcs258994. DOI: 10.1242/jcs.258994.

Kompa, J., J. Bruins, M. Glogger, J. Wilhelm, M. S. Frei, et al. (2023). "Exchangeable HaloTag Ligands for Super-Resolution Fluorescence Microscopy". In: *J. Am. Chem. Soc.* 145, pp. 3075–3083. DOI: 10.1021/jacs.2c11969.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). *ImageNet Classification with Deep Convolutional Neural Networks*. Red Hook: Curran Associates, Inc.

Krogh, A. and J. A. Hertz (1991). "A simple weight decay can improve generalization". In: *NeurIPS* 8, pp. 950–957. DOI: 10.5555/2986916.2987033.

Krull, A., T.-O. Buchholz, and F. Jug (2018). "Noise2Void - Learning Denoising from Single Noisy Images". In: *arXiv*. DOI: 10.48550/arXiv.1811.10980.

Kuhn, H. (1949). "A Quantum-Mechanical Theory of Light Absorption of Organic Dyes and Similar Compounds". In: *J. Chem. Phys.* 17, pp. 1192–1212. DOI: 10.1063/1.1747143.

Kusimi, A. and Y. Sako (1996). "Cell surface organization by the membrane skeleton". In: *Curr. Opin. Cell Biol.* 8, pp. 566–574. DOI: 10.1016/s0955-0674(96)80036-6.

Laine, R. F., G. Jacquemet, and A. Krull (2021). "Imaging in focus: An introduction to denoising bioimages in the era of deep learning". In: *Int. J. Biochem. Cell Bio.* 140, p. 106077. DOI: 10.1016/j.biocel.2021.106077.

Laine, R. F., K. L. Tosheva, N. Gustafsson, R. D. M. Gray, P. Almada, et al. (2019). "NanoJ: a high-performance open-source super-resolution microscopy toolbox". In: *J. Phys. D: Appl. Phys.* 52, p. 163001. DOI: 10.1088/1361-6463/ab0261.

Lakowicz, J. (2006). *Principles of fluorescence spectroscopy*. 3rd ed. New York: Springer. DOI: 10.1007/978-0-387-46312-4.

Lambert, J. H. (1852). *Photometria, sive de mensura et gradibus luminis, colorum et umbrae*. Augsburg: Klett. DOI: 10.3931/e-rara-9488.

LeCun, Y., Y. Bengio, and G. Hinton (2015). "Deep learning". In: *Nature* 521, pp. 436–444. DOI: 10.1038/nature14539.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner (1998). "Gradient-Based Learning Applied to Document Recognition". In: *Proc. IEEE* 86, pp. 2278–2324. DOI: 10.1109/5.726791.

Lehtinen, J., J. Munkberg, J. Hasselgren, S. Laine, T. Karras, et al. (2018). "Noise2Noise: Learning Image Restoration without Clean Data". In: *arXiv*. DOI: 10.48550/arXiv.1803.04189.

Lemmon, M. A. and J. Schlessinger (2010). "Cell signaling by receptor-tyrosine kinases". In: *Cell* 141, pp. 1117–1134. DOI: 10.1016/j.cell.2010.06.011.

Lequyer, J., R. Philip, A. Sharma, W.-H. Hsu, and L. Pelletier (2022). "A fast blind zero-shot denoiser". In: *Nat. Mach. Intell.* 4, pp. 953–963. DOI: 10.1038/s42256-022-00547-8.

Levet, F., E. Hosy, A. Kechkar, C. Butler, A. Beghin, et al. (2015). "SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data". In: *Nat. Methods* 12, pp. 1065–1071. DOI: 10.1038/nmeth.3579.

Li, H., A. Kadav, I. Durdanovic, H. Samet, and H. P. Graf (2017). "Pruning Filters for Efficient ConvNets". In: *arXiv*. DOI: 10.48550/arXiv.1608.08710.

Li, N., G.-S. Xiang, H. Dokainish, K. Ireton, and L. A. Elferink (2005). "The Listeria Protein Internalin B Mimics Hepatocyte Growth Factor-Induced Receptor Trafficking". In: *Traffic* 6, pp. 459–473. DOI: 10.1111/j.1600-0854.2005.00290.x.

Li, Y., L. Shang, and G. U. Nienhaus (2016). "Super-resolution imaging-based single particle tracking reveals dynamics of nanoparticle internalization by live cells". In: *Nanoscale* 8, p. 7423. DOI: 10.1039/c6nr01495.

Liu, S., P. Hoess, and J. Ries (2022a). "Super-Resolution Microscopy for Structural Cell Biology". In: *Annu. Rev. Biophys.* 51, pp. 301–326. DOI: 10.1146/annurev-biophys-102521-112912.

Liu, T., J. Kompa, J. Ling, N. Lardon, Y. Zhang, et al. (2024). "Gentle rhodamines for live-cell fluorescence microscopy". In: *bioRxiv*. DOI: 10.1101/2024.02.06.579089.

Liu, T., T. Stephan, P. Chen, J. Keller-Findeisen, J. Chen, et al. (2022b). "Multi-color live-cell STED nanoscopy of mitochondria with a gentle inner membrane stain". In: *Proc. Natl. Acad. Sci. USA* 119, e2215799119. DOI: 10.1073/pnas.2215799119.

Liu, Y.-L., C.-K. Chou, M. Kim, R. Vasiht, Y.-A. Kuo, et al. (2019). "Assessing metastatic potential of breast cancer cells based on EGFR dynamics". In: *Sci. Rep.* 9, p. 3395. DOI: 10.1038/s41598-018-37625-0.

Liu, Z., E. Hirata-Miyasaki, C. Foley, J. Rahm, S. Pradeep, et al. (2023). "Robust virtual staining of cellular landmarks in high-throughput correlative microscopy". In: *ASCB*.

Livingstone, D. J. (2010). *Artificial Neural Networks Methods and Applications*. Totowa: Humana Press. DOI: 10.1007/978-1-60327-101-1.

Lopes, F. B., Š. Bállint, S. Valvo, J. H. Felce, E. M. Hessel, et al. (2017). "Membrane nanoclusters of FcγRI segregate from inhibitory SIRPα upon activation of human macrophages". In: *J. Cell. Biol.* 4, pp. 1123–1141. DOI: 10.1083/jcb.201608094.

Los, G. V., L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, et al. (2008). "HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis". In: *ACS Chem. Biol.* 3, pp. 373–382. DOI: 10.1021/cb800025k.

Loshchilov, I. and F. Hutter (2017). "Decoupled Weight Decay Regularization". In: *arXiv*. DOI: 10.48550/arXiv.1711.05101.

Low-Nam, S. T., K. A. Lidke, P. J. Cutler, R. C. Roovers, P. M. P. van Bergen en Henegouwen, et al. (2011). "Single-Particle Tracking Reveals Switching of the HIV Fusion Peptide between Two Diffusive Modes in Membranes". In: *Na. Struct. Mol. Biol.* 18, pp. 1244–1249. DOI: 10.1038/nsmb.2135.

Lu, M., C. N. Christensen, J. M. Weber, T. Konno, N. F. Läubli, et al. (2023). "ERnet: a tool for the semantic segmentation and quantitative analysis of endoplasmic reticulum topology". In: *Nat. Methods* 20, pp. 569–579. DOI: 10.1038/s41592-023-01815-0.

Mahecic, D., W. L. Stepp, C. Zhang, J. Griffié, M. Weigert, et al. (2022). "Event-driven acquisition for content-enriched microscopy". In: *Nat. Methods* 19, pp. 1262–1267. DOI: 10.1038/s41592-022-01589-x.

Maier-Hein, L., A. Reinke, R. Godau, M. D. Tizabi, F. Buettner, et al. (2024). "Metrics reloaded: recommendations for image analysis validation". In: *Nat. Methods* 21, pp. 195–212. DOI: 10.1038/s41592-023-02151-z.

Majumder, P., K. Roy, S. Bagh, and D. Mukhopadhyay (2019). "Receptor tyrosine kinases (RTKs) consociate in regulatory clusters in Alzheimer's disease and type 2 diabetes". In: *Mol. Cell. Biochem.* 459, pp. 171–182. DOI: 10.1007/s11010-019-03560-5.

Malkusch, S., U. Endesfelder, J. Mondry, M. Gelléri, P. J. Verveer, et al. (2012). "Coordinate-based colocalization analysis of single-molecule localization microscopy data". In: *Histochem. Cell Biol.* 137, pp. 1–10. DOI: 10.1007/s00418-011-0880-5.

Malkusch, S., J. V. Rahm, M. S. Dietz, M. Heilemann, J.-B. Sibarita, et al. (2022). "Receptor tyrosine kinase MET ligand-interaction classified via machine learning from single-particle tracking data". In: *Mol. Biol. Cell* 33, 33(6):ar60. DOI: 10.1091/mbc.E21-10-0496.

Manley, S., J. M. Gillette, G. H. Patterson, H. Shroff, H. F. Hess, et al. (2008). " High-density Mapping of Single-Molecule Trajectories with Photoactivated Localization Microscopy". In: *Nat. Methods* 5, pp. 155–157. DOI: 10.1038/nmeth.1176.

Manzo, C. and M. F. Garcia-Parajo (2015). "A review of progress in single particle tracking: from methods to biophysical insights". In: *Rep. Prog. Phys.* 78, p. 124601. DOI: 10.1088/0034-4885/78/12/124601.

Marchetti, L., A. Callegari, S. Luin, G. Signore, A. Viegi, et al. (2013). "Ligand signature in the membrane dynamics of single TrkA receptor molecules". In: *J. Cell Sci.* 126, pp. 4445–4456. DOI: 10.1242/jcs.129916.

Marsh, R. J., I. Costello, M.-A. Gorey, D. Ma, F. Huang, et al. (2021). "Sub-diffraction error mapping for localisation microscopy images". In: *Nat. Comm.* 12, p. 5611. DOI: 10.1038/s41467-021-25812-z.

Martens, K. J. A., B. Turkowdy, J. Hohlbein, and U. Endesfelder (2024). "Temporal analysis of relative distances (TARDIS) is a robust, parameter-free alternative to single-particle tracking". In: *Nat. Methods*. DOI: 10.1038/s41592-023-02149-7.

Matz, M. V., A. F. Fradkov, Y. A. Labas, A. P. Savitsky, A. G. Zaraisky, et al. (1999). "Fluorescent proteins from nonbioluminescent Anthozoa species". In: *Nat. Biotechnol.* 17, pp. 969–973. DOI: 10.1038/13657.

Mauvezin, C., P. Nagy, G. Juhász, and T. P. Neufeld (2015). "Autophagosome–lysosome fusion is independent of V-ATPase-mediated acidification". In: *Nat. Comm.* 6, p. 7007. DOI: 10.1038/ncomms8007.

McCulloch, W. S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity". In: *Bull. Math. Biol.* 5, pp. 115–133. DOI: 10.1007/BF02478259.

McDole, K., L. Guignard, F. Amata nd A. Berger, G. Malandain, L. A. Royer, et al. (2018). "In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level". In: *Cell* 175, pp. 859–976. DOI: 10.1016/j.cell.2018.09.031.

Mikolov, T. (2012). "Statistical Language Models based on Neural Networks". In: *Brno University of Technology*.

Milletari, F., N. Navab, and S.-A. Ahmadi (2016). "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation". In: *arXiv*. DOI: 10.48550/arXiv.1606.04797.

Min, J., C. Vonesch, H. Kirshner, L. Carlini, N. Olivier, et al. (2015). "FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data". In: *Sci. Rep.* 4, p. 4577. DOI: 10.1038/srep04577.

Möckl, L., D. C. Lamb, and C. Bräuchle (2014). "Super-resolved Fluorescence Microscopy: Nobel Prize in Chemistry 2014 for Eric Betzig, Stefan Hell, and William E. Moerner". In: *Angew. Chem. Int. Ed.* 53, pp. 13972–13977. DOI: 10.1002/anie.201410265.

Möckl, L., A. R. Roy, and W. E. Moerner (2020). "Deep learning in single-molecule microscopy: fundamentals, caveats, and recent developments". In: *Biomed. Opt. Express* 11, pp. 1633–1661. DOI: 10.1038/s41467-021-22518-0.

Mollazade, M., T. Tabarin, P. R. Nicovich, A. Soeriyadi, D. J. Nieves, et al. (2017). "Can single molecule localization microscopy be used to map closely spaced RGD nanodomains?" In: *PLOS ONE* 12, e0180871. DOI: 10.1371/journal.pone.0180871.

Mortensen, K. I., L. S. Churchman, J. A. Spudich, and H. Flyvbjerg (2010). "Optimized localization analysis for single-molecule tracking and super-resolution microscopy". In: *Nat. Methods* 7, pp. 337–381. DOI: 10.1038/nmeth.1447.

Nair, V. and G. E. Hinton (2010). "Rectified Linear Units Improve Restricted Boltzmann Machines". In: *ICML*, pp. 807–814.

Narayanasamy, K. K., J. V. Rahm, S. Tourani, and M. Heilemann (2022). "Fast DNA-PAINT imaging using a deep neural network". In: *Nat. Commun.* 13, p. 5047. DOI: 10.1038/s41467-022-32626-0.

Narayanasamy, K. K., A. Stojic, Y. Li, S. Sass, M. R. Hesse, et al. (2021). "Visualizing Synaptic Multi-Protein Patterns of Neuronal Tissue With DNA-Assisted Single-Molecule Localization Microscopy". In: *Front. Synaptic Neurosci.* 13, p. 671288. DOI: 10.3389/fnsyn.2021.671288.

Nehme, E., D. Freedman, R. Gordon, B. Ferdman, L. E. Weiss, et al. (2020). "DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning". In: *Nat. Methods* 17, pp. 734–740. DOI: 10.1038/s41592-020-0853-5.

Nehme, E., L. E. Weiss, T. Michaeli, and Y. Shechtman (2018). "Deep-STORM: super-resolution single-molecule microscopy by deep learning". In: *Optica* 5, pp. 458–464. DOI: 10.1364/OPTICA.5.000458.

Nesterov, Y. (1983). "A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$". In: *Doklady ANSSSR (translated as Soviet. Math. Docl.)* 269, pp. 543–547.

Nguyen, T. L., E. R. Polanco, A. N. Patananan, T. A. Zangle, and M. A. Teitell (2020). "Cell viscoelasticity is linked to fluctuations in cell biomass distributions". In: *Sci. Reports* 10, p. 7403. DOI: 10.1038/s41598-020-64259-y.

Nguyen, T. L., S. Pradeep, R. L. Judson-Torres, J. Reed, M. A. Teitell, et al. (2022). "Quantitative Phase Imaging: Recent Advances and Expanding Potential in Biomedicine". In: *ACS Nano* 16, pp. 11516–11544. DOI: 10.1021/acsnano.1c11507.

Niederauer, C., C. nguyen, M. Wang-Henders, J. Stein, S. Strauss, et al. (2023). "Dual-color DNA-PAINT single-particle tracking enables extended studies of membrane protein interactions". In: *Nat. Comm.* 14, p. 4345. DOI: 10.1038/s41467-023-40065-8.

Nomarski, G. M. (1955). "Differential microinterferometer with polarized waves". In: *J. Phys. Radium* 16, 9S.

Odell, I. D. and D. Cook (2013). "Immunofluorescence Techniques". In: *Proc. Soc. Exp. Biol. Med.* 133, e4. DOI: 10.1038/jid.2012.455.

Okabe, A., B. Boots, K. Sugihara, and S. N. Chiu (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. 2nd ed. Chichester: Wiley.

Ori, A., M. C. Wilkinson, and D. G. Fernig (2008). "The heparanome and regulation of cell function: structures, functions and challenges". In: *Front. Biosci.* 13, pp. 4309–4338. DOI: 10.2741/3007.

Orré, T., A. Joly, Z. Karatas, B. Kastberger, C. Cabriel, et al. (2021). "Molecular motion and tridimensional nanoscale localization of kindlin control integrin activation in focal adhesions". In: *Nat. Comm.* 12, p. 3104. DOI: 10.1038/s41467-021-23372-w.

Ott, M., Y. Shai, and G. Haran (2013). "Single-Particle Tracking Reveals Switching of the HIV Fusion Peptide between Two Diffusive Modes in Membranes". In: *J. Phys. Chem.* 117, pp. 13308–13321. DOI: 10.1021/jp4039418.

Ounkomol, C., S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson (2018). "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy". In: *Nat. Methods* 15, pp. 917–920. DOI: 10.1038/s41592-018-0111-2.

Ouyang, W., A. Aristov, M. Lelek, X. Hao, and C. Zimmer (2018). "Deep learning massively accelerates super-resolution localization microscopy". In: *Nat. Biotechnol.* 36, pp. 460–468. DOI: 10.1038/nbt.4106.

Ouyang, W., F. Beuttenmueller, E. Gómez-de-Mariscal, C. Pape, T. Burke, et al. (2022). "BioImage Model Zoo: A Community-Driven Resource for Accessible Deep Learning in BioImage Analysis". In: *bioRxiv*. DOI: 10.1101/2022.06.07.495102.

Ovesny, M., P. Krizek, J. Borkovec, Z. Svindrych, and G. M. Hagen (2014). "ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging ". In: *Bioinformatics* 30, pp. 2389–2390. DOI: 10.1093/bioinformatics/btu202.

Pageon, S. V., P. R. Nicovich, M. Mollazade, T. Tabarin, and K. Gaus (2016). "Clus-DoC: a combined cluster detection and colocalization analysis for single-molecule localization microscopy data". In: *Mol. Biol. Cell* 27, pp. 3627–3636. DOI: 10.1091/mbc.E16-07-0478.

Park, Y., C. A. Best, K. Badizadegan, R. R. Dasari, M. S. Feld, et al. (2010). "Measurement of red blood cell mechanics during morphological changes". In: *Proc. Natl. Acad. Sci. USA* 107, pp. 6731–6736. DOI: 10.1073/pnas.0909533107.

Park, Y., C. Depeursinge, and G. Popescu (2018). "Quantitative phase imaging in biomedicine". In: *Nat. Photonics* 12, pp. 578–589. DOI: 10.1038/s41566-018-0253-x.

Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, et al. (2019). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Adv. Neural. Inf. Process. Syst.* 32, pp. 8024–8035. DOI: 10.5555/3454287.3455008.

Paul, M. D. and K. Hristova (2019). "The RTK Interactome: Overview and Perspective on RTK Hetero-Interactions". In: *Chem Rev.* 119, pp. 5881–5921. DOI: 10.1021/acs.chemrev.8b00467.

Pearson, K. (1895). "Notes on regression and inheritance in the case of two parents". In: *Proc. R. Soc. Lond.* 58, pp. 240–242.

Pereira, C. F., J. Rossy, D. M. Owen, J. Mak, and K. Gaus (2012). "HIV taken by STORM: Super-resolution fluorescence microscopy of a viral infection". In: *Virol. J.* 9, p. 84. DOI: 10.1186/1743-422X-9-84.

Persson, F., M. Lindén, C. Unoson, and J. Elf (2013). "Extracting intracellular diffusive states and transition rates from single-molecule tracking data". In: *Nat. Methods* 10, pp. 265–269. DOI: 10.1038/nmeth.2367.

Pike, L. J. (2006). "Rafts defined: a report on the Keystone symposium on lipid rafts and cell function". In: *J. Lipid. Res.* 47, pp. 1597–1598. DOI: 10.1194/jlr.E600002-JLR200.

Planck, M. (1901). "Über das Gesetz der Energieverteilung im Normalspectrum". In: *Annalen der Physik* 309 (3), pp. 553–563. DOI: 10.1002/andp.19013090310.

Popescu, G. (2011). *Quantitative Phase Imaging of Cells and Tissues*. 1st ed. available online. New York: McGraw-Hill.

Priessner, M., D. C. A. Gaboriau, A. Sheridan, T. Lenn, C. Garzon-Coral, et al. (2024). "Content-aware frame interpolation (CAFI): deep learning-based temporal super-resolution for fast bioimaging". In: *Nat. Methods* 21, pp. 322–330. DOI: 10.1038/s41592-023-02138-w.

Qian, N. (1999). "On the momentum term in gradient descent learning algorithms". In: *Neural Netw.* 12, pp. 145–151. DOI: 10.1016/S0893-6080(98)00116-6.

Racine, V., A. Hertzog, J. Jouanneau, J. Salamero, C. Kervrann, et al. (2006). "Multiple-target tracking of 3D fluorescent objects based on simulated annealing". In: *IEEE Int. Symp. Biomed. Imaging*, pp. 1020–1023. DOI: 10.1109/ISBI.2006.1625094.

Rahm, J. V., A. Balakrishnan, M. Wehrheim, M. Glogger, M. Kaschube, et al. (2024). "Fast and long-time tracking of ER nano-structural dynamics in living cells". In: *Manuscript in preparation*.

Rahm, J. V., S. Malkusch, U. Endesfelder, M. S. Dietz, and M. Heilemann (2021). "Diffusion State Transitions in Single-Particle Trajectories of MET Receptor Tyrosine Kinase Measured in Live Cells". In: *Front. Comput. Sci.* 3, p. 757653. DOI: 10.3389/fcomp.2021.757653.

– (2022). "Extraction of diffusion state transitions in single-particle tracking data of membrane receptors". In: *SPIE BiOS* 11967, p. 1196705. DOI: 10.1117/12.2609681.

Raju, R., S. M. Palapetta, V. K. Sandhya, A. Sahu, A. Alipoor, et al. (2014). "A Network Map of FGF-1/FGFR Signaling System". In: *J. Signal Transduct.* 2014, p. 962962. DOI: 10.1155/2014/962962.

Ran, F. A., P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, et al. (2013). "Genome engineering using the CRISPR-Cas9 system ". In: *Nat. Protoc.* 8, pp. 2281–2308. DOI: 10.1038/nprot.2013.143.

Rayleigh (1896). "On the theory of optical images, with special reference to the microscope". In: *Lond. Edinb. Dublin philos. mag.* 42, pp. 167–195. DOI: 10.1080/14786449608620902.

Reinhardt, S. C. M., L. A. Masullo, I. Baudrexel, P. R. Steen, R. Kowalewski, et al. (2023). "Ångström-resolution fluorescence microscopy". In: *Nature* 617, pp. 711–716. DOI: 10.1038/s41586-023-05925-9.

Reinke, A., M. D. Tizabi, M. Baumgartner, M. Eisenmann, D. Heckmann-Nötzel, et al. (2024). "Understanding metric-related pitfalls in image analysis validation". In: *Nat. Methods* 21, pp. 182–194. DOI: 10.1038/s41592-023-02150-0.

Requejo-Isidro, J. (2013). "Fluorescence nanoscopy. Methods and applications". In: *J. Chem. Biol.* 6, pp. 97–120. DOI: 10.1007/s12154-013-0096-3.

Requena, B., S. Masó, J. Bertran, M. Lewenstein, C. Manzo, et al. (2023). "Inferring pointwise diffusion properties of single trajectories with deep learning". In: *Biophys. J.* DOI: 10.1016/j.bpj.2023.10.015.

Resch-Genger, U., M. Grabolle, S. Cavaliere-Jaricot, R. Nitschke, and T. Nann (2008). "Quantum dots versus organic dyes as fluorescent labels". In: *Nat. Methods* 5, pp. 763–775. DOI: 10.1038/NMEtH.1248.

Ripley, B. D. (1977). "Modelling Spatial Patterns". In: *J. R. Stat. Soc. Series B Methodol.* 39, pp. 172–192. DOI: 10.1111/j.2517-6161.1977.tb01615.x.

Robbins, H. and S. Monro (1951). "A Stochastic Approximation Method". In: *Ann. Math. Stat.* 22, p. 400. DOI: 10.1214/aoms/1177729586.

Rocha-azevedo, B. da, S. Lee, A. Dasgupta, A. R. Vega, L. R. de Oliveria, et al. (2020). "Heterogeneity in VEGF Receptor-2 Mobility and Organization on the Endothelial Cell Surface Leads to Diverse Models of Activation by VEGF". In: *Cell Rep.* 32, p. 108187. DOI: 10.1016/j.celrep.2020.108187.

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *arXiv*. DOI: 10.48550/arXiv.1505.04597.

Rosenblatt, F. (1957). "The perceptron – A perceiving and recognizing automaton". In: *CAL*, pp. 85-60–1. DOI: 10.1037/h0042519.

– (1958). "The perceptron: A probabilistic model for information storage and organization in the brain". In: *Physiol. Rev.* 65, pp. 386–408. DOI: 10.1037/h0042519.

Rossier, O., V. Octeau, J.-B. Sibarita, C. Leduc, B. Tessier, et al. (2012). "Integrins β1 and β3 exhibit distinct dynamic nanoscale organizations inside focal adhesions". In: *Nat. Cell Biol.* 14, pp. 1057–1067. DOI: 10.1038/ncb2588..

Ruder, S. (2016). "An overview of gradient descent optimization algorithms". In: *arXiv*. DOI: 10.48550/arXiv.1609.04747.

Rumelhart, D. E., G. E. Hinton, and R. J. Wiliams (1986). "Learning representations by back-propagating errors". In: *Nature* 323, pp. 533–536. DOI: 10.1038/323533a0.

Rust, M. J., M. Bates, and X. Zhuang (2006). "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)". In: *Nat. Methods* 3, pp. 793–796. DOI: 10.1038/nmeth929.

Saffman, P. G. and M. Delbrück (1975). "Brownian motion in biological membranes". In: *Proc. Nat. Acad. Sci. USA* 72, pp. 3111–3113. DOI: 10.1073/pnas.72.8.3111.

Sage, D., T.-A. Pham, H. Babcock, T. Lukes, T. Pengo, et al. (2019). "Neural network-assisted single-molecule localization microscopy with a weak-affinity protein tag". In: *Nat. Methods* 16, pp. 387–395. DOI: 10.1038/s41592-019-0364-4.

Saguy, A., O. Alalouf, N. Opatovski, S. Jang, M. Heilemann, et al. (2023). "DBlink: dynamic localization microscopy in super spatiotemporal resolution via deep learning". In: *Nat. Methods* 20, pp. 1939–1948. DOI: 10.1038/s41592-023-01966-0.

Sanderson, M. J., I. Smith, I. Parker, and M. D. Bootman (2014). "Fluorescence Microscopy". In: *Cold Spring Harb. Protoc.* 10, pp. 1042–1065. DOI: 10.1101/pdb.top071795.

Saraiva, B. M., I. M. Cunha, A. D. Brito, G. Follain, R. Portela, et al. (2023). "NanoPyx: super-fast bioimage analysis powered by adaptive machine learning". In: *bioRxiv*. DOI: 10.1101/2023.08.13.553080.

Sarrazin, S., W. C. Lamanna, and J. D. Esko (2011). "Heparan Sulfate Proteoglycans". In: *Cold Spring Harb. Perspect. Biol.* 3, a004952. DOI: 10.1101/cshperspect.a004952.

Sassano, M. L., B. Felipe-Abrio, and P. Agostinis (2022). "ER-mitochondria contact sites; a multifaceted factory for Ca2+ signaling and lipid transport". In: *Front. Cell Dev. Biol.* 10. DOI: 10.3389/fcell.2022.988014.

Sauer, M. and M. Heilemann (2017). "Single-Molecule Localization Microscopy in Eukaryotes". In: *Chem. Rev.* 117, pp. 7478–7509. DOI: 10.1021/acs.chemrev.6b00667.

Saxton, M. J. and K. Jacobson (1997). "SINGLE-PARTICLE TRACKING: Applications to Membrane Dynamics". In: *Annu. Rev. Biophys. Biomol. Struct.* 26, pp. 373–399. DOI: 10.1146/annurev.biophys.26.1.373.

Saxton, W. O. and W. Baumeister (1982). "The correlation averaging of a regularly arranged bacterial cell envelope protein". In: *J. Microsc.* 127, pp. 127–138. DOI: 10.1111/j.1365-2818.1982.tb00405.x.

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, et al. (2012). "Fiji: an open-source platform for biological-image analysis". In: *Nat. Methods* 9, pp. 676–682. DOI: 10.1038/nmeth.2019.

Schmidt, U., M. Weigert, C. Broaddus, and G. Myers (2020). "Cell Detection with Star-convex Polygons". In: *arXiv*. DOI: 10.48550/arXiv.1806.03535.

Schnitzbauer, J., M. T. Strauss, T. Schlichthaerle, F. Schueder, and R. Jungmann (2017). "Super-resolution microscopy with DNA-PAINT". In: *Nat. Protoc.* 12, pp. 1198–1228. DOI: 10.1038/nprot.2017.024.

Schröder, M. (2021). "Thesis: Spatial organization of fibroblast growth factor receptors in the plasma membrane of cells investigated by DNA-assisted single-molecule super-resolution microscopy". In: *Goethe University Frankfurt*.

Schröder, M., M.-L. I. E. Harwardt, J. V. Rahm, Y. Li, P. Freund, et al. (2021). "Imaging the fibroblast growth factor receptor network on the plasma membrane with DNA-assisted single-molecule super-resolution microscopy". In: *Methods* 193, pp. 49–54. DOI: 10.1016/j.ymeth.2020.05.004.

Schueder, F., J. Stein, F. Stehr, A. Auer, B. Sperl, et al. (2019). "An order of magnitude faster DNA-PAINT imaging by optimized sequence design and buffer conditions". In: *Nat. Methods* 16, pp. 1101–1104. DOI: 10.1038/s41592-019-0584-7.

Shah, Z. H., M. Müller, T.-C. Wang, P. M. Scheidig, A. Schneider, et al. (2021). "Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images". In: *Photon. Res.* 9, B168–B181. DOI: 10.1364/PRJ.416437.

Shalev-Shwartz, S., Y. Singer, N. Srebro, and A. Cottor (2010). "Pegasos: Primal Estimated sub-GrAdient SOlver for SVM". In: *Math. Program.* 127, pp. 3–30. DOI: 10.1007/s10107-010-0420-4.

Shaner, N. C., R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, et al. (2004). "Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein". In: *Nat. Biotechnol.* 22, pp. 1567–1572. DOI: 10.1038/nbt1037.

Sharonov, A. and R. M. Hochstrasser (2006). "Wide-field subdiffraction imaging by accumulated binding of diffusing probes". In: *Proc. Natl. Acad. Sci.* 103, pp. 18911–18916. DOI: 10.1073/pnas.0609643104.

Shen, H., L. J. Tauzin, R. Baiyasi, W. Wang, N. Moringo, et al. (2017). "Single Particle Tracking: From Theory to Biophysical Applications". In: *Chem. Rev.* 117, pp. 7331–7376. DOI: 10.1021/acs.chemrev.6b00815.

Shimomura, O., F. H. Johnson, and Y. Saiga (1962). "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan". In: *J. Cell Comp. Physiol.* 59, pp. 223–239. DOI: 10.1002/jcp.1030590302.

Shorten, C. and T. M. Khoshgoftaar (2019). "A survey on Image Data Augmentation for Deep Learning". In: *J. Big Data* 6, p. 60. DOI: 10.1186/s40537-019-0197-0.

Sibarita, J.-B. (2014). "High-density single-particle tracking: quantifying molecule organization and dynamics at the nanoscale". In: *Histonchem. Cell. Biol.* 141, pp. 587–595. DOI: 10.1007/s00418-014-1214-1.

Simonyan, K. and A. Zisserman (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *arXiv*. DOI: 10.48550/arXiv.1409.1556.

Singer, S. J. and G. L. Nicolson (1972). "The fluid mosaic model of the structure of cell membranes". In: *Science* 175, pp. 720–731. DOI: 10.1126/science.175.4023.720.

Smith, L. N. (2017). "Cyclical Learning Rates for Training Neural Networks". In: *arXiv*. DOI: 10.48550/arXiv.1506.01186.

Sograte-Idrissi, S., N. Oleksiievets, S. Isbaner, M. Eggert-Martinez, J. Enderlein, et al. (2019). "Nanobody Detection of Standard Fluorescent Proteins Enables Multi-Target DNA-PAINT with High Resolution and Minimal Displacement Errors". In: *Cells* 8, p. 48. DOI: 10.3390/cells8010048.

Sokolova, M., N. Japkowicz, and S. Szpakowicz (2006). "Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation". In: *Adv. in Art. Intell.* 4304, pp. 1015–1021. DOI: 10.1007/11941439_114.

Sørensen, T. (1948). "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons". In: *K. Dansk. Vidensk. Selsk. Skr.* 5, pp. 1–34.

Soto, J. M., J. A. Rodrigo, and T. Alieva (2017). "Label-free quantitative 3D tomographic imaging for partially coherent light microscopy". In: *Opt. Express* 25, pp. 15699–15712. DOI: 10.1364/OE.25.015699.

Spahn, C., J. B. Grimm, L. D. Lavis, M. Lampe, and M. Heilemann (2019). "Whole-Cell, 3D, and Multicolor STED Imaging with Exchangeable Fluorophores". In: *Nano Lett.* 19, pp. 500–505. DOI: 10.1021/acs.nanolett.8b04385.

Speiser, A., L.-R. Müller, P. Hoess, U. Matti, C. J. Obara, et al. (2021). "Deep learning enables fast and dense single-molecule localization with high accuracy". In: *Nat. Methods* 18, pp. 1082–1090. DOI: 10.1038/s41592-021-01236-x.

Stallinga, S. and B. Rieger (2012). "The effect of background on localization uncertainty in single emitter imaging". In: *IEEE Int. Symp. Biomed. Imaging*, pp. 988–991. DOI: 10.1109/ISBI.2012.6235723.

Steffens, H., W. Wegner, and K. I. Willig (2020). "In vivo STED microscopy: A roadmap to nanoscale imaging in the living mouse". In: *Methods* 174, pp. 42–48. DOI: 10.1016/j.ymeth.2019.05.020.

Stehr, F., J. Stein, J. Bauer, C. Niederauer, R. Jungmann, et al. (2021). "Tracking single particles for hours via continuous DNA-mediated fluorophore exchange". In: *Nat. Comm.* 12, p. 4432. DOI: 10.1038/s41467-021-24223-4.

Stockhammer, A. and F. Bottanelli (2021). "Appreciating the small things in life: STED microscopy in living cells". In: *J. Phys. D: Appl. Phys.* 54, p. 033001. DOI: 10.1088/1361-6463/abac81.

Stokes, G. G. (1852). "On the change of refrangibility of light". In: *Phil. Trans. R. Soc.* 142, pp. 463–562. DOI: 10.1098/rstl.1852.0022.

Stone, M. B., S. A. Shelby, and S. L. Veatch (2017). "Super-Resolution Microscopy: Shedding Light on the Cellular Plasma Membrane". In: *Chem. Rev.* 117, pp. 7457–7477. DOI: 10.1021/acs.chemrev.6b00716.

Strauss, S. and R. Jungmann (2020). "Up to 100-fold speed-up and multiplexing in optimized DNA-PAINT". In: *Nat. Methods* 17, pp. 789–791. DOI: 10.1038/s41592-020-0869-x.

Stringer, C., T. Wang, M. Michaelos, and M. Pachitariu (2021). "Cellpose: a generalist algorithm for cellular segmentation". In: *Nat. Methods* 18, pp. 100–106. DOI: 10.1038/s41592-020-01018-x.

Strubell, E., A. Ganesh, and A. McCallum (2023). "Energy and Policy Considerations for Deep Learning in NLP". In: *arXiv*. DOI: 10.48550/arXiv.1906.02243.

Subach, F. V., G. H. Patterson, M. Renz, J. Lippincott-Schwartz, and V. V. Verkhusha (2010). "Bright Monomeric Photoactivatable Red Fluorescent Protein for Two-Color Super-Resolution sptPALM of Live Cells". In: *J. Am. Chem. Soc.* 132, pp. 6481–6491. DOI: 10.1021/ja100906g.

Sungkaworn, T., M.-L. Jobin, K. Burnecki, A. Weron, M. J. Lohse, et al. (2017). "Single-molecule imaging reveals receptor–G protein interactions at cell surface hot spots". In: *Nature* 550, pp. 543–547. DOI: 10.1038/nature24264.

Sutton, R. S. (1986). "Two problems with backpropagation and other steepest-descent learning procedures for networks". In: *Proceedings of the Eighth Annual Meeting of the Cognitive Science Society*.

Teodori, L., M. Omer, A. Märcher, M. K. Skaanning, V. L. Andersen, et al. (2022). "Site-specific nanobody-oligonucleotide conjugation for super-resolution imaging". In: *J. Biol. Methods* 9, e159. DOI: 10.14440/jbm.2022.381.

Thompson, R. E., D. R. Larson, and W. W. Webb (2002). "Precise nanometer localization analysis for individual fluorescent probes". In: *Biophys. J.* 82, pp. 2775–2783. DOI: 10.1016/S0006-3495(02)75618-X.

Tokunaga, M., N Imamoto, and K. Sakata-Sogawa (2008). "Highly inclined thin illumination enables clear single-molecule imaging in cells". In: *Nat. Methods* 5, pp. 159–161. DOI: 10.1038/nmeth1171.

Tompson, J., R. Goroshin, A. Jain, Y. LeCun, and C. Bregler (2015). "Efficient Object Localization Using Convolutional Networks". In: *arXiv*. DOI: 10.48550/arXiv.1411.4280.

Tonzani, S. (2009). "TIRF: imaging at the cellular edge". In: *Nat. Cell. Biol.* 11, p. 16. DOI: 10.1038/ncb1933.

Tortarolo, G., M. Castello, A. Diaspro, S. Koho, and G. Vicidomini (2018). "Evaluating image resolution in stimulated emission depletion microscopy". In: *Optica* 5, pp. 32–35. DOI: 10.1364/OPTICA.5.000032.

Toseland, C. P. (2013). "Fluorescent labeling and modification of proteins". In: *J. Chem. Biol.* 6, pp. 85–95. DOI: 10.1007/s12154-013-0094-5.

Unterauer, E. M., S. S. Boushehri, K. Jevdokimenko, L. A. Masullo, M. Ganji, et al. (2024). "Spatial proteomics in neurons at single-protein resolution". In: *Cell* 187, pp. 1785–1800. DOI: 10.1016/j.cell.2024.02.045.

Váradi, T., M. Schneider, E. Sevcisk, D. Kiesenhofer, F. Baumgart, et al. (2019). "Homo- and Heteroassociations Drive Activation of ErbB3". In: *Biophys. J.* 117, pp. 1935–1947. DOI: 10.1016/j.bpj.2019.10.001.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, et al. (2017). "Attention Is All You Need". In: *arXiv*. DOI: 10.48550/arXiv.1706.03762.

Verdecchia, R., J. Sallou, and L. Cruz (2023). "A Systematic Review of Green AI". In: *arXiv*. DOI: 10.48550/arXiv.2301.11047.

Verzelli, P., A. Nold, C. Sun, M. Heilemann, E. S. Schuman, et al. (2022). "Unbiased choice of global clustering parameters for single-molecule localization microscopy". In: *Sci. Reports*, p. 22561. DOI: 10.1038/s41598-022-27074-1.

Vicidomini, G., P. Bianchini, and A. Diaspro (2018). "STED super-resolved microscopy". In: *Nat. Methods* 15, pp. 173–182. DOI: 10.1038/nmeth.4593.

Vink, J. N. A., S. J. J. Brouns, and J. Hohlbein (2020). "Extracting Transition Rates in Particle Tracking Using Analytical Diffusion Distribution Analysis". In: *Biophys. J.* 119, pp. 1970–1983. DOI: 10.1016/j.bpj.2020.09.033.

Vinod, V., C. J. Padmakrishnan, B. Vijayan, and S. Gopala (2015). "'How can I halt thee?' The puzzles involved in autophagic inhibition". In: *Pharmacol. Res.* 82, pp. 1–8. DOI: 10.1016/j.phrs.2014.03.005.

Virtanen, P., R. Gommers, T. E. Oliphant, m. Haberland, T. Reddy, et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nat. Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Voulodimos, A., N. Doulamis, A. Doulamis, and E. Protopapadakis (2018). "Deep Learning for Computer Vision: A Brief Review". In: *Comput. Intell. Neurosci.* DOI: 10.1155/2018/7068349.

Wade, O. K., J. B. Woehrstein, P. C. Nickels, S. Strauss, F. Stehr, et al. (2019). "124-Color Super-resolution Imaging by Engineering DNA-PAINT Blinking Kinetics". In: *Nano Lett.* 19, pp. 2641–2646. DOI: 10.1021/acs.nanolett.9b00508.

Wäldchen, S., J. Lehmann, T. Klein, S. van de Linde, and M. Sauer (2015). "Light-induced cell damage in live-cell super-resolution microscopy". In: *Sci. Reports* 5, p. 15348. DOI: 10.1038/srep15348.

Wallis, T. P., A. Jiang, K. Young, H. Hou, K. Kudo, et al. (2023). "Super-resolved trajectory-derived nanoclustering analysis using spatiotemporal indexing". In: *Nat. Comm.* 14, p. 3353. DOI: 10.1038/s41467-023-38866-y.

Walt, S. van der, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, et al. (2014). "Scikit-image: image processing in Python". In: *PeerJ* 2, e453. DOI: 10.7717/peerj.453.

Wang, H., Y. Rivenson, Y. Jin, Z. Wei, R. Gao, et al. (2019). "Deep learning enables cross-modality super-resolution in fluorescence microscopy". In: *Nat. Methods* 16, pp. 103–110. DOI: 10.1038/s41592-018-0239-0.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli (2004). "Image quality assessment: from error visibility to structural similarity". In: *IEEE Trans. Image Process.* 12, pp. 600–612. DOI: 10.1109/TIP.2003.819861.

Wang, Z., E. P. Simoncelli, and A. C. Bovik (2003). "Multiscale structural similarity for image quality assessment". In: *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers* 2, pp. 1398–1402. DOI: 10.1109/ACSSC.2003.1292216.

Weigend, A. S., D. E. Rumelhart, and B. A. Huberman (1990). "Generalization by Weight-Elimination with Application to Forecasting". In: *NeurIPS* 3, pp. 875–882. DOI: 10.5555/2986766.2986885.

Weigert, M., U. Schmidt, T. Boothe, A. Müller, A. Dibrov, et al. (2018). "Content-aware image restoration: pushing the limits of fluorescence microscopy". In: *Nat. Methods* 15, pp. 1090–1097. DOI: 10.1038/s41592-018-0216-7.

Werbos, P. J. (1974). "Thesis: Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences". In: *Harvard University*.

Weron, A., K. Burnecki, E. J. Akin, L. Solé, M. Balcerek, et al. (2017). "Ergodicity breaking on the neuronal surface emerges from random switching between diffusive states". In: *Sci. Rep.* 7, p. 5404. DOI: 10.1016/j.bpj.2023.10.015.

Wheeler, D. L. and Y. Yarden (2015). *Receptor Tyrosine Kinases: Family and Subfamilies*. 1st ed. Basel: Springer Cham. DOI: 10.1007/978-3-319-11888-8.

Willig, K. I., B. Harke, R. Medda, and S. W. Hell (2007). "STED microscopy with continuous wave beams". In: *Nat. Methods* 4, pp. 915–918. DOI: 10.1038/nmeth1108.

Wills, M. F. K., C. B. Alejo, N. Hundt, A. J. Hudson, and I. C. Eperon (2023). "FluoroTensor: identification and tracking of colocalised molecules and their stoichiometries in multi-colour single molecule imaging via deep learning". In: *bioRxiv*. DOI: 10.1101/2023.11.21.567874.

Wolter, S., U. Endesfelder, S. van de Linde, M. Heilemann, and M. Sauer (2011). "Measuring localization performance of super-resolution algorithms on very active samples". In: *Opt. Express* 19, pp. 7020–7033. DOI: 10.1364/OE.19.007020.

Wolter, S., A. Löschberger, T. Holm, S. Aufmkolk, M.-C. Dabauvalle, et al. (2012). "rapidSTORM: accurate, fast open-source software for localization microscopy". In: *Nat. Methods* 9, pp. 1040–1041. DOI: 10.1038/nmeth.2224.

Xu, J., G. Qin, F. Luo, L. Wang, R. Zhao, et al. (2019). "Automated Stoichiometry Analysis of Single-Molecule Fluorescence Imaging Traces via Deep Learning". In: *J. Am. Chem. Soc.* 141, pp. 6976–6985. DOI: 10.1021/jacs.9b00688.

Yang, Y., F. Ma, Z. Liu, Q. Su, Y. Liu, et al. (2018). "The ER-localized Ca2+-binding protein calreticulin couples ER stress to autophagy by associating with microtubule-associated protein 1A/1B light chain 3". In: *Cell. Biol.* 294, pp. 772–782. DOI: 10.1074/jbc.RA118.005166.

Yayon, A., M. Klagsbrun, J. D. Esko, P. Leder, and D. M. Ornitz (1991). "Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor". In: *Cell* 64, pp. 841–848. DOI: 10.1016/0092-8674(91)90512-W.

Yazdanfar, S., K. B. Kenny, K. Tasimi, A. D. Corwin, E. L. Dixon, et al. (2008). "Simple and robust image-based autofocusing for digital microscopy". In: *Opt. Express* 16, pp. 8670–8677. DOI: 10.1364/OE.16.008670.

Yeh, Li-Hao, Ivan E. Ivanov, Janie R. Byrum, Bryant B. Chhun, Syuan-Ming Guo, et al. (2021). "uPTI: uniaxial permittivity tensor imaging of intrinsic density and anisotropy". In: *bioRxiv*. DOI: 10.1101/2020.12.15.422951.

Yoon, B.-J. (2009). "Hidden Markov Models and their Applications in Biological Sequence Analysis". In: *Curr. Genomics* 10, pp. 402–415. DOI: 10.2174/138920209789177575.

You, C., T. T. Marquez-Lago, C. P. Richter, S. Wilmes, I. Moraga, et al. (2016). "Receptor dimer stabilization by hierarchical plasma membrane microcompartments regulates cytokine signaling". In: *Sci. Adv.* 2, e1600452. DOI: 0.1126/sciadv.1600452.

You, K., J. Wang, M. Long, and M. I. Jordan (2019). "How Does Learning Rate Decay Help Modern Neural Networks?" In: *arXiv*. DOI: 10.48550/arXiv.1908.01878.

Yu, D., H. Wang, P. Chen, and Z. Wei (2014). "Mixed Pooling for Convolutional Neural Networks". In: 8818, pp. 364–375. DOI: 10.1007/978-3-319-11740-9_34.

Zamir, S. W., A. Arora, S. Khan, M. Hayat, F. S. Khan, et al. (2021). "Multi-Stage Progressive Image Restoration". In: *arXiv*. DOI: 10.48550/arXiv.2102.02808.

Zangle, T. A., D. Burnes, C. Mathis, O. N. Witte, and M. A. Teitell (2013). "Quantifying Biomass Changes of Single CD8+ T Cells during Antigen Specific Cytotoxicity". In: *PLOS ONE* 9, e68916. DOI: 10.1371/journal.pone.0068916.

Zeiler, M. D. (2012). "ADADELTA: An Adaptive Learning Rate Method". In: *arXiv*. DOI: 10.48550/arXiv.1212.5701.

Zeiler, M. D., G. W. Taylor, and R. Fergus (2011). "Adaptive Deconvolutional Networks for Mid and High Level Feature Learning". In: *ICCV*, pp. 2018–2025. DOI: 10.1109/ICCV.2011.6126474.

Zernike, F. (1955). "How I discovered phase contrast". In: *Science* 121, pp. 345–349. DOI: 10.1126/science.121.3141.345.

Zhang, R., P. Isola, A. A. Efros, E. Shechtman, and O. Wang (2018a). "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric". In: *arXiv*. DOI: 10.48550/arXiv.1801.03924.

Zhang, Y., M. Lara-tejero, J. Bewersdorf, and J. E. Galán (2017). "Visualization and characterization of individual type III protein secretion machines in live bacteria". In: *Proc. Natl. Acad. Sci.* 114, pp. 6098–6103. DOI: 10.1073/pnas.1705823114.

Zhang, Y., K. Li, K. Li, L. Wang, B. Zhong, et al. (2018b). "Image super-resolution using very deep residual channel attention networks". In: *ECCV*, pp. 286–301. DOI: 10.48550/arXiv.1807.02758.

Zhang, Y., K.-Y. Park, K. F. Suazo, and M. D. Distefano (2018c). "Recent progress in enzymatic protein labelling techniques and their applications". In: *Chem. Soc. Rev.* 47, pp. 9106–9136. DOI: 10.1039/c8cs00537k.

Zhao, H., O. Gallo, I. Frosio, and J. Kautz (2015). "Loss Functions for Image Restoration with Neural Networks". In: *arXiv*. DOI: 10.48550/arXiv.1511.08861.

Zhao, W., Z. Huang, J. Yang, G. Qiu, L. Qu, et al. (2023). "Quantitatively mapping local quality of super-resolution microscopy by rolling Fourier ring correlation". In: *Light Sci. Appl.*, p. 298. DOI: 10.1038/s41377-023-01321-0.

Zhou, Y., J. Cao, J. Sonneck, S. Banerjee, S. Dörr, et al. (2024). "EfficientBioAI: making bioimaging AI models efficient in energy and latency". In: *Nat. Methods*. DOI: 10.1038/s41592-024-02167-z.

Zhuge, H., B. Summa, J. Hamm, and J. Q. Brown (2021). "Deep learning 2D and 3D optical sectioning microscopy using cross-modality Pix2Pix cGAN image translation". In: *Biomed. Opt. Express* 12, pp. 7526–7543. DOI: 10.1364/BOE.439894.

Zuo, C., J. Li, J. Sun, Y. Fan, J. Zhang, et al. (2020). "Transport of intensity equation: a tutorial". In: *Science* 135, p. 106187. DOI: 10.1016/j.optlaseng.2020.106187.

13 Appendix

13.1 Fast and long-time STED imaging of ER nano-structural dynamics in living cells

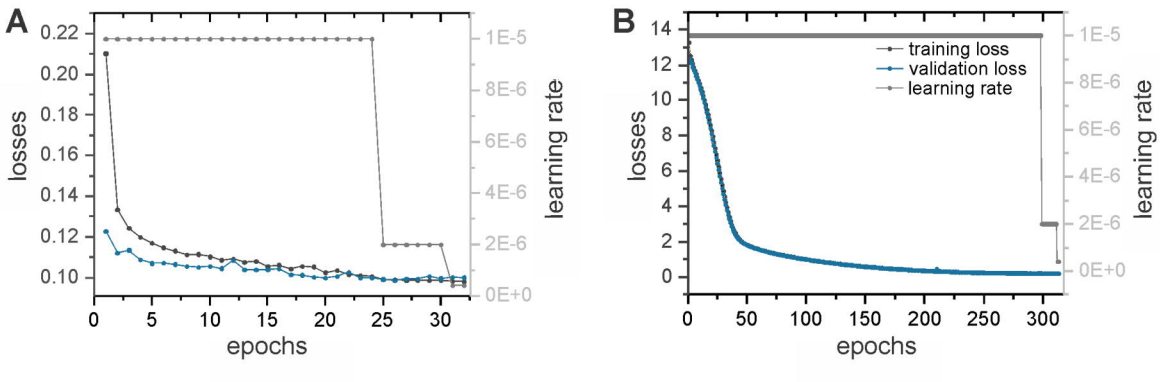

Figure S1: Loss curves and learning rates of the training process. Training-, validation -loss curves and the learning rates for the two models (**A** Resting model; **B** Mixed model) used in this study, which were reduced through LearningRateOnPlateau during training. The training process stopped if the validation loss did not improve for 6 epochs (Rahm et al., 2024).

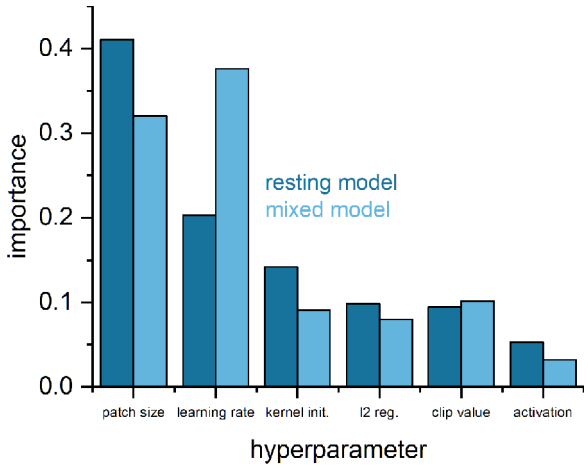

Figure S2: Hyperparameter importance. For both the models (resting model and mixed model) used in this study, the patch size and learning rate are the most important hyperparameters influencing model performance, followed by kernel initialization, l2 regularization, and gradient clip value in the middle field, and lastly the type of activation function (Rahm et al., 2024).

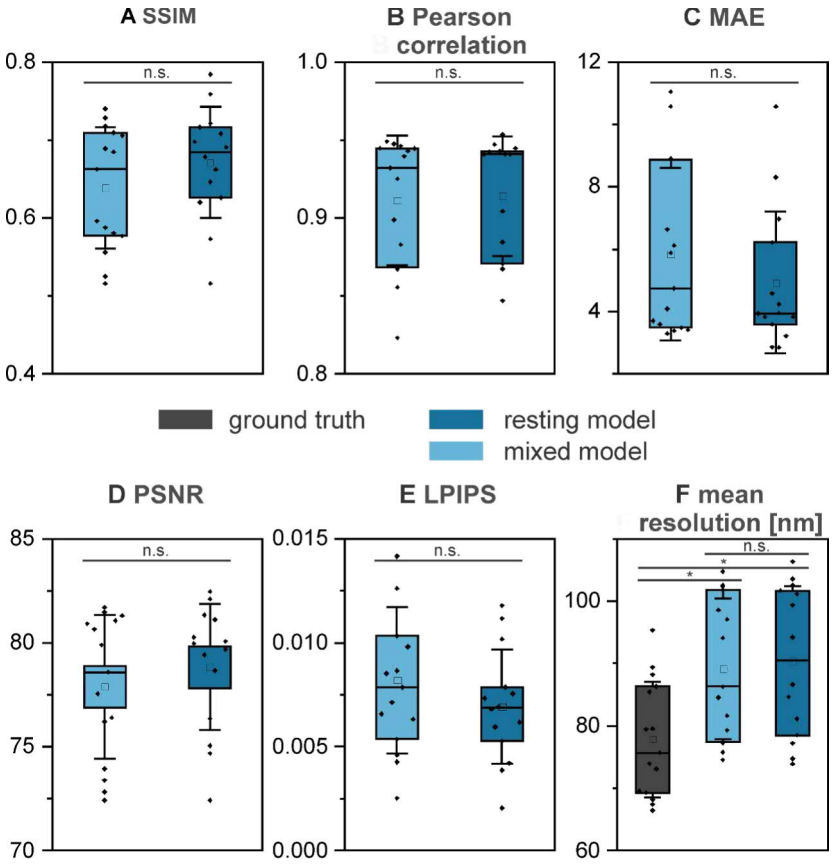

Figure S3: Quality control metrics of prediction from autophagy-induced cell dataset. Quality control metrics **A** SSIM, **B** PEARSON correlation, **C** MAE, **D** PSNR, **E** LPIPS, and **F** mean resolution from rFRC of datasets acquired from stress-induced cells predicted using either the mixed model (light blue) or the resting model (blue), dark gray: mean resolution of ground truth image, N = 15 (Rahm et al., 2024).

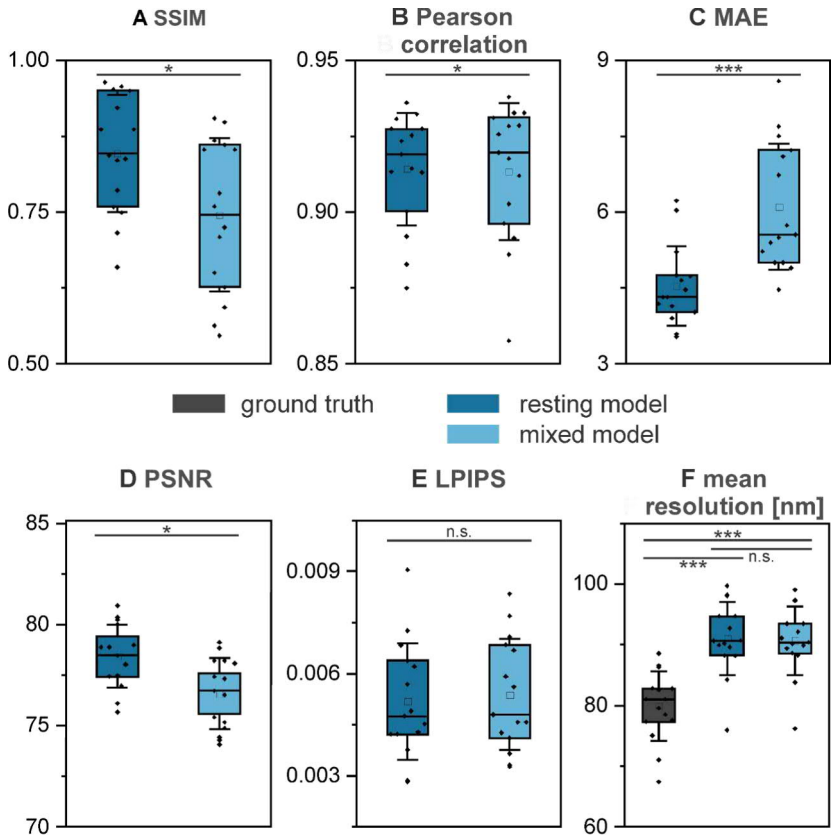

Figure S4: Quality control metrics of prediction from the resting cell dataset. Quality control metrics **A** SSIM, **B** PEARSON correlation, **C** MAE, **D** PSNR, **E** LPIPS, and **F** mean resolution from rFRC of datasets acquired from resting cells predicted using either the resting model (blue) or the mixed model (light blue), dark gray: mean resolution of ground truth image, N = 15 (Rahm et al., 2024).

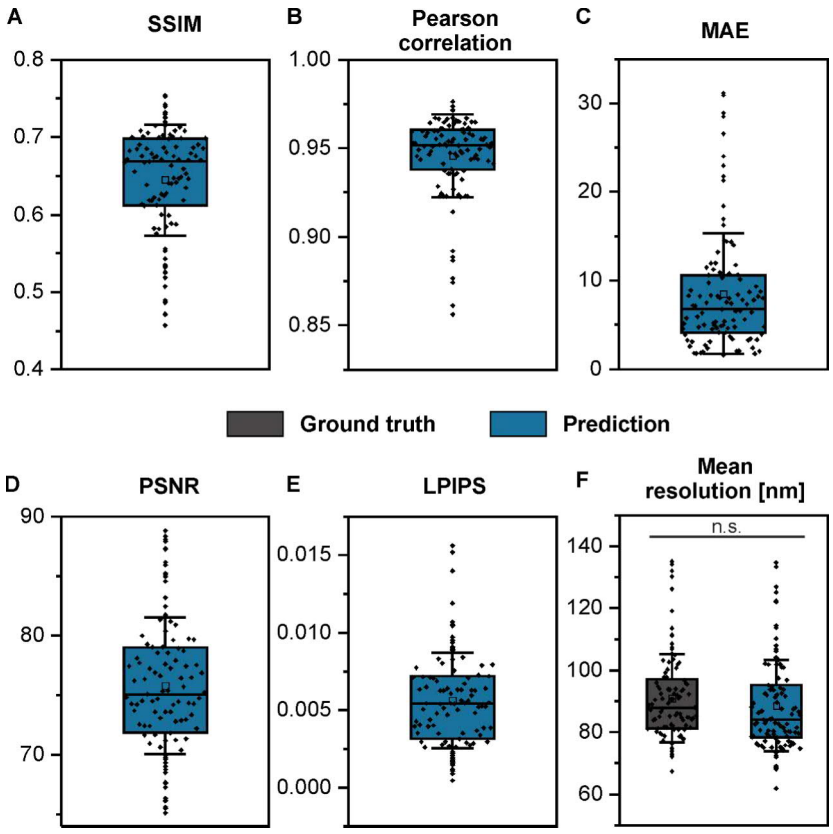

Figure S5: Quality control metrics of 3D controls. QC metrics of predictions from planar images acquired with the same settings as the volumetric imaging including a top-hat PSF for the depletion laser beam ranging from **A** SSIM **B** PEARSON correlation **C** MAE **D** PSNR **E** LPIPS to **F** mean resolution from rFRC, N = 101 (Rahm et al., 2024).

Table S1: Hyperparameter importance of the resting condition. The optimal parameter choices are highlighted in bold.

learning rate	score	patch size	score	L2	score
1E-4	13.00	**304**	**28.28**	**0**	**35.39**
1E-5	**47.97**	259	20.44	0.01	30.65
1E-6	39.03	200	25.75	0.001	33.96
		128	25.52		
activation	**score**	**kernel init.**	**score**	**clip value**	**score**
leaky relu	48.79	**Glorot uniform**	**45.25**	0.1	32.23
tanh	**51.21**	Lecun uniform	29.74	0.01	30.59
		orthogonal	25.01	**0.001**	**37.18**

Table S2: Hyperparameter importance of the mixed condition. The optimal parameter choices are highlighted in bold.

learning rate	score	patch size	score	L2	score
5E-4	4.56	**304**	**25.95**	0	32.73
1E-4	11.81	256	23.26	**0.01**	**35.72**
5E-5	28.29	200	24.98	0.001	31.54
1E-5	**31.08**	128	25.81		
1E-5	24.27				
activation	**score**	**kernel init.**	**score**	**clip value**	**score**
leaky relu	47.93	**Glorot uniform**	**37.51**	0.1	32.69
tanh	52.07	Lecun uniform	35.66	**0.01**	**35.29**
		orthogonal	26.83	0.001	32.02

Table S3: Resting and mixed model parameters. Unstated parameters are set as default values.

learning rate	patch size	batch size	l2	activation	kernel init.	clip value	epochs
resting model							
1E-5	304	6	0	leaky relu	Glorot uniform	0.001	32
mixed model							
1E-5	304	6	0.001	leaky relu	Lecun uniform	0.1	313

13.2 Robust virtual staining for cell instance segmentation

13.2.1 Data preprocessing

Before the 3D-stack of bright-field images and matching nucleus and membrane fluorescence images can be used as training data, several preprocessing steps must be applied. These include the determination of appropriate regularization parameters for bright-field to phase conversion and fluorescence channel deconvolution, registration of the label-free and fluorescence channels in x, y and z, and tiling of the images to make them suitable training inputs.

13.2.1.1 Phase reconstruction and fluorescence deconvolution

The 3D stack of bright-field images was converted to phase images and the fluorescence channels were deconvolved with waveOrder (Yeh et al., 2021). Compared to bright-field images, the phase images have a more uniform background, higher contrast, and greater depth of field (fig. S6 A, B). The fluorescence signal is sharper after deconvolution because the blur around the fluorescent regions stemming from out of focus planes is reduced (fig. S6 C-F).

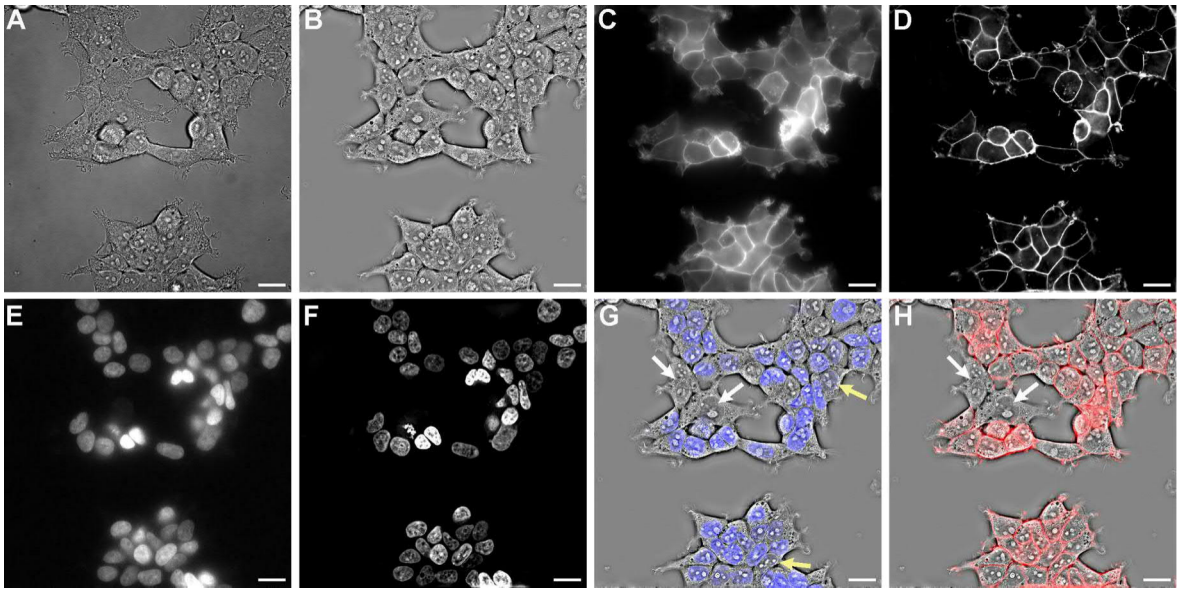

Figure S6: Phase reconstruction and fluorescence deconvolution. **A** Bright-field, **B** phase, **C** membrane fluorescence, **D** deconvolved membrane fluorescence, **E** nucleus fluorescence, and **F** deconvolved nucleus fluorescence image. Deconvolved fluorescence and phase images are overlayed in **G** for the nucleus and **H** for the membrane. White arrows indicate cells without nucleus and membrane stain. Yellow arrows highlight areas with low-intensity fluorescence signal. Scale bars 20 µm.

For both the bright-field to phase conversion and deconvolution, values for TIKHONOV regularization must be chosen. Too small regularization values generate artifacts in the phase image and make the output look overly sharp and incomplete. Too large regularization values make the output look blurry. There was not much of a difference between values of $1 \cdot 10^{-3}$ to $1 \cdot 10^{-1}$ for phase reconstruction, but $1 \cdot 10^{-3}$ had the slightly higher contrast and was therefore chosen. For fluorescence deconvolution, a value of $8 \cdot 10^{-4}$ proofed most suitable. Due to the polyclonal labeling, the intensities in the fluorescence images are heterogeneous. Fig. S6 G and H highlight parts with weak fluorescence (yellow arrows) and absent fluorescence (white arrows) although there are cells present in the corresponding phase image.

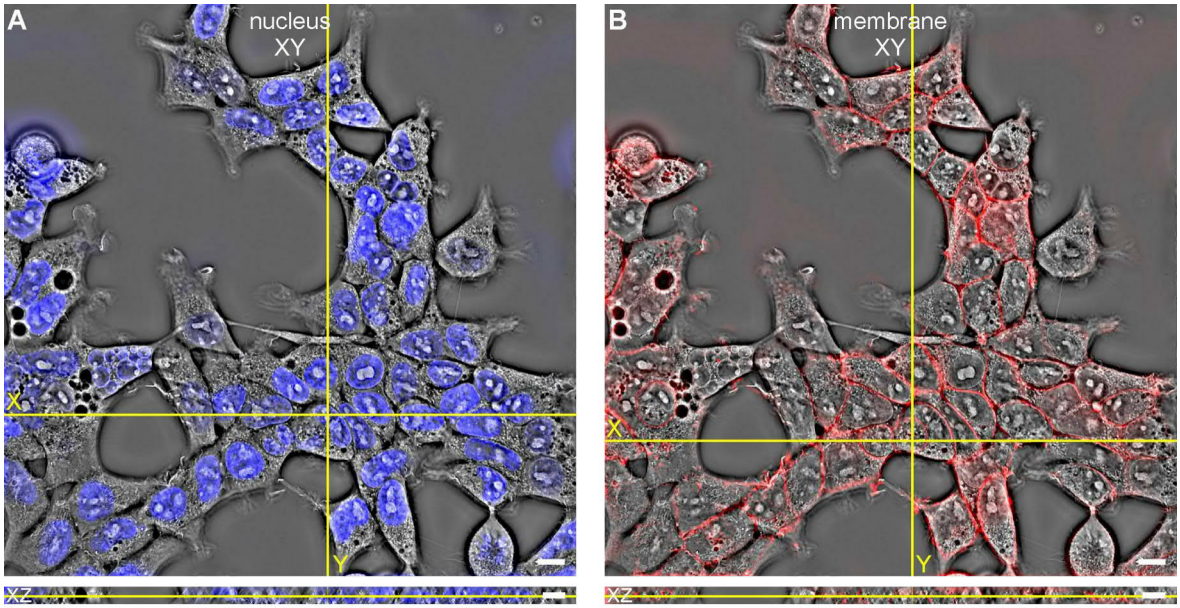

Figure S7: Channel registration. A The nucleus channel in blue and **B** the membrane channel in red are aligned with the phase channel. The yellow lines highlight the position of the XZ sectioning. Scale bars in XY view 10 µm and in XZ view 2 µm.

13.2.1.2 Registration

Images were registered along the z-axis by determining the focus with the BRENNER gradient (Yazdanfar et al., 2008) for each channel and aligning the focal planes across the channels. The focal planes of the phase channel and the membrane channel were aligned and the registration offset applied to the nucleus channel. This is done because the phase channel appeared to be in focus when the membrane edges were displayed sharply. Due to the egg-shaped form of a cell, the focal plane of the nucleus lies above the bright-field and membrane fluorescence focal planes and that offset was maintained during registration. Some positions had very high or low z-focus indices, possibly due to unstable positioning of the microscope in z-direction during acquisition. Positions with a z-focus index in the

membrane channel of < 25 and > 60 were filtered out, as they do not contain enough information for correct phase reconstruction. Registration along x and y was not necessary. The phase and fluorescence channels appear to be registered after z-alignment (fig. S7).

13.2.1.3 Tiling

Images were tiled and kept as training instances if they contained a minimum of 5% foreground pixels. Different thresholding approaches were compared to distinguish between foreground and background pixels (fig. S8). Unimodal thresholding marked fluorescent regions in the central FOV ($z = 25$), but incorrectly marked large amounts of background in FOVs at the border of the z-volume ($z = 65$). Otsu's thresholding was more conservative in marking fluorescence regions but was stable across z-positions. Therefore, Otsu's thresholding was chosen as the masking approach, and tiles with fluorescent regions were selected by adjusting the minimum foreground fraction to a small value of 5%. In the case of models with multiple targets, the masks were merged across different channels before selecting training instances.

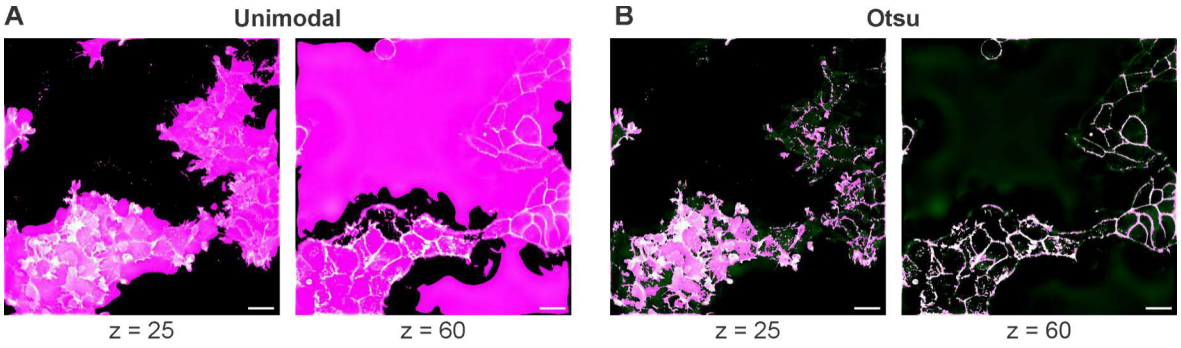

Figure S8: Binary mask parameters. **A** Unimodal thresholding and **B** Otsu's thresholding of the membrane fluorescence channel at different z-depths. Scale bars 20 µm.

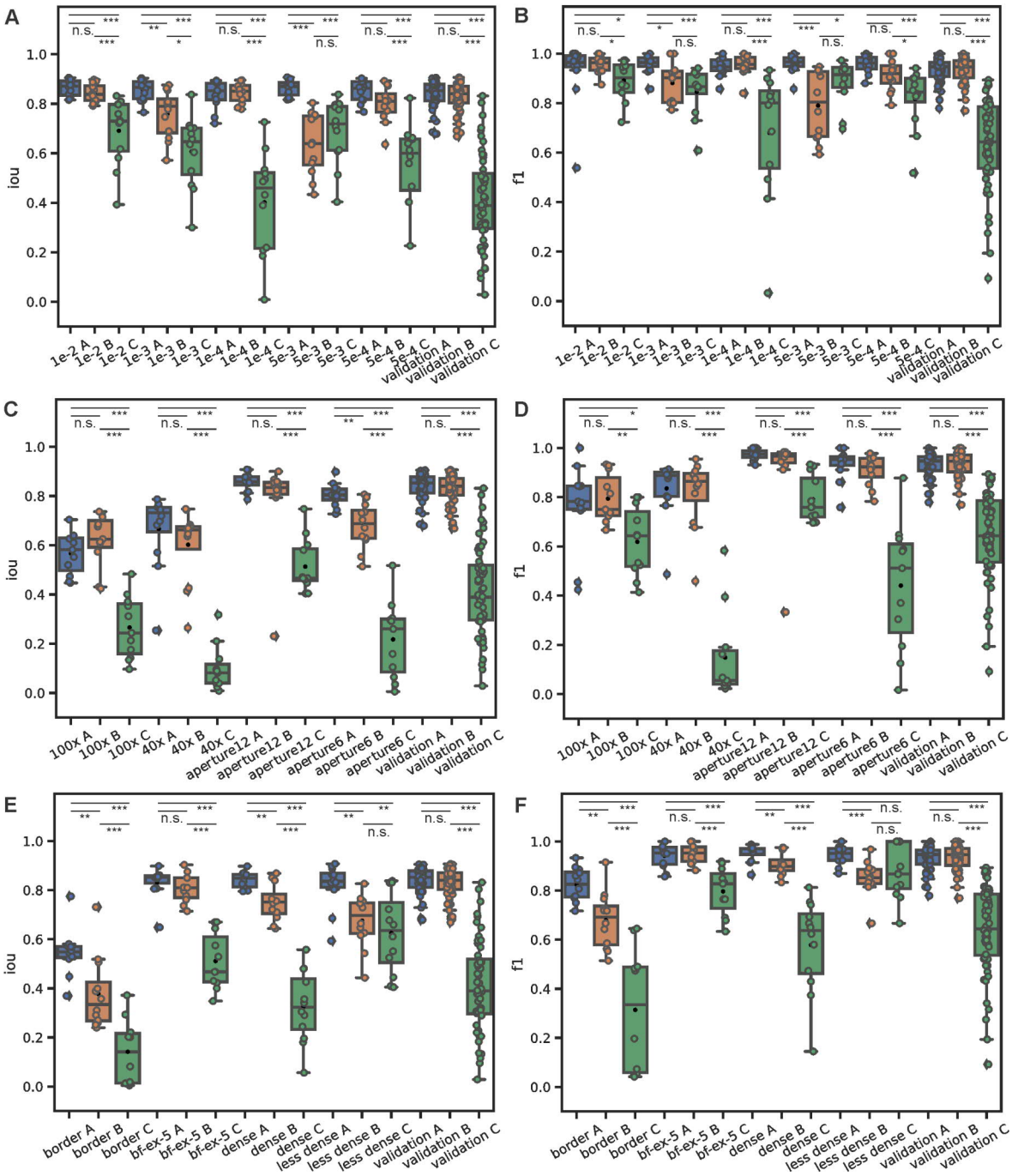

Figure S9: Metric comparison of different augmentation strengths. A, **C**, and **E** show the intersection over union and **B**, **D**, and **F** the f1-score for different regularizations parameters for the phase ($1 \cdot 10^{-2}$, $1 \cdot 10^{-3}$, $1 \cdot 10^{-4}$, $5 \cdot 10^{-3}$, $5 \cdot 10^{-4}$), different magnifications (100x, 40x), condenser aperture openings (12, 6), imaging at the border of the well, half exposure time, FOVs with dense cells, FOVs with less dense cells, and the validation data. Strong augmentation (blue), moderate augmentation (orange), light augmentation (green).

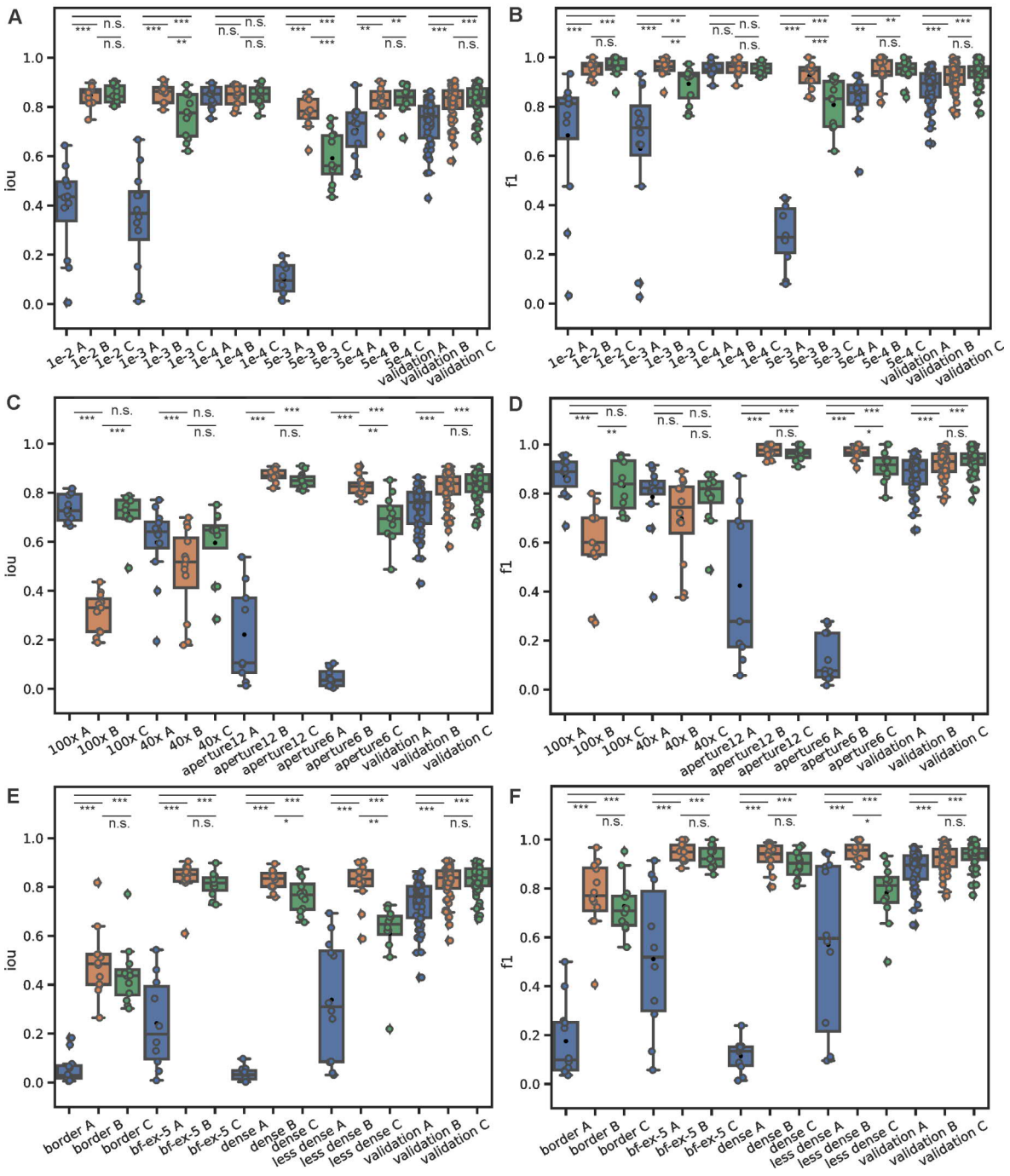

Figure S10: Metric comparison of augmentation without intensity or zoom. A, **C**, and **E** show the intersection over union and **B**, **D**, and F the f1-score for different regularizations parameters for the phase ($1 \cdot 10^{-2}$, $1 \cdot 10^{-3}$, $1 \cdot 10^{-4}$, $5 \cdot 10^{-3}$, $5 \cdot 10^{-4}$), different magnifications (100x, 40x), condenser aperture openings (12, 6), imaging at the border of the well, half exposure time, FOVs with dense cells, FOVs with less dense cells, and the validation data. No intensity (blue), no zoom (orange), moderate augmentation (green).

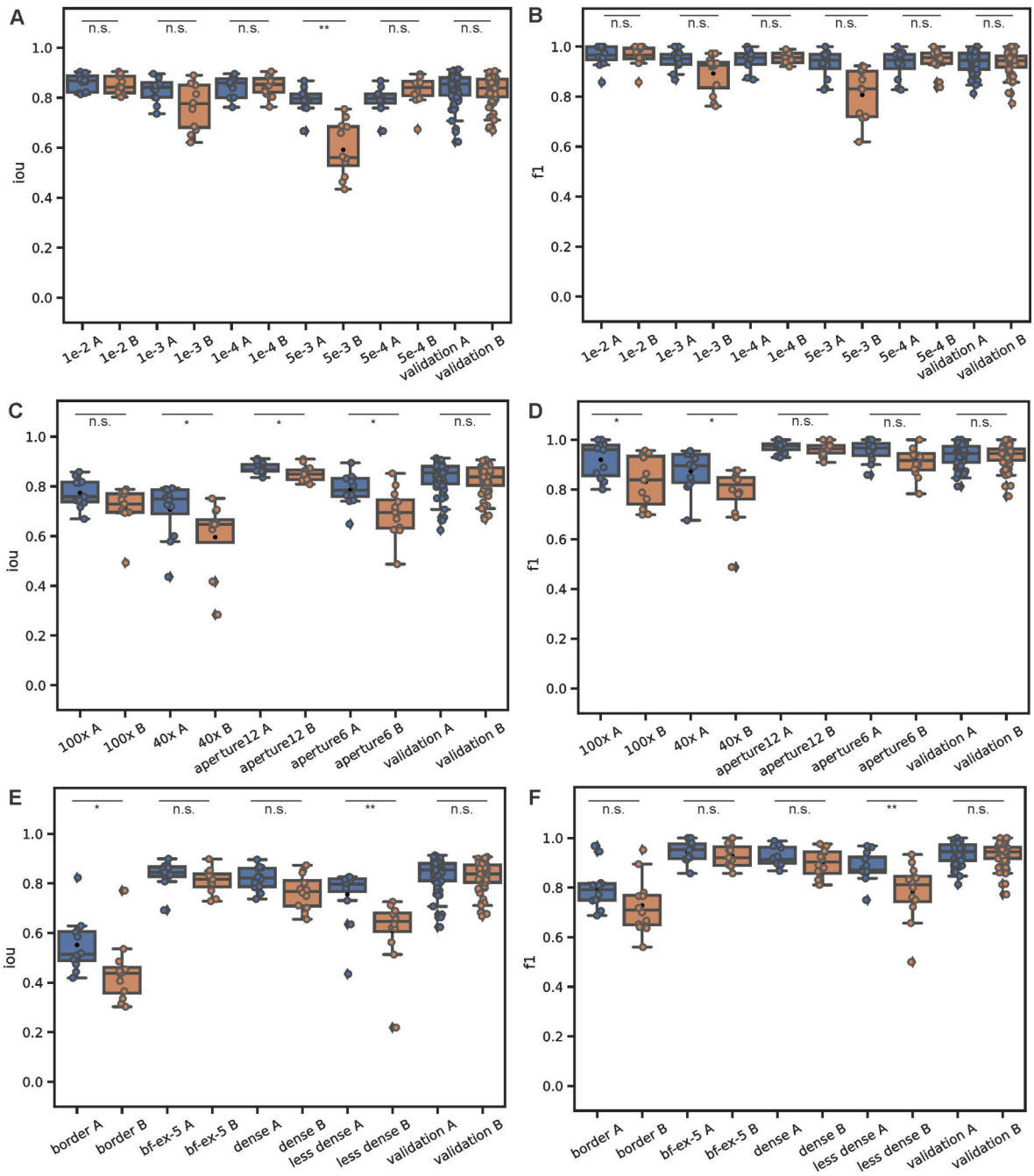

Figure S11: Metric comparison of nucleus as target channel vs nucleus and membrane as target channels. **A**, **C**, and **E** show the intersection over union and **B**, **D**, and F the f1-score for different regularizations parameters for the phase ($1 \cdot 10^{-2}$, $1 \cdot 10^{-3}$, $1 \cdot 10^{-4}$, $5 \cdot 10^{-3}$, $5 \cdot 10^{-4}$), different magnifications (100x, 40x), condenser aperture openings (12, 6), imaging at the border of the well, half exposure time, FOVs with dense cells, FOVs with less dense cells, and the validation data. Nucleus target channel (blue), nucleus and membrane target channels (orange).

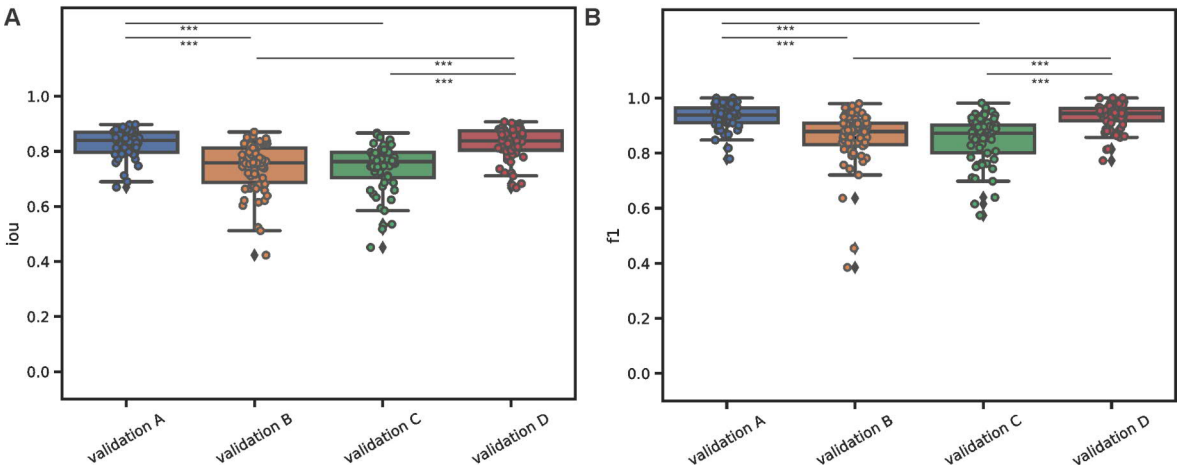

Figure S12: Metric comparison of different imaging modalities. **A** Intersection over union and **B** f1-score of the validation data. Bright-field → deconvolved fluorescence (blue), phase → non-deconvolved fluorescence, bright-field → non-deconvolved fluorescence, phase → deconvolved fluorescence.

13.3 A pipeline for the analysis of single-particle tracking data

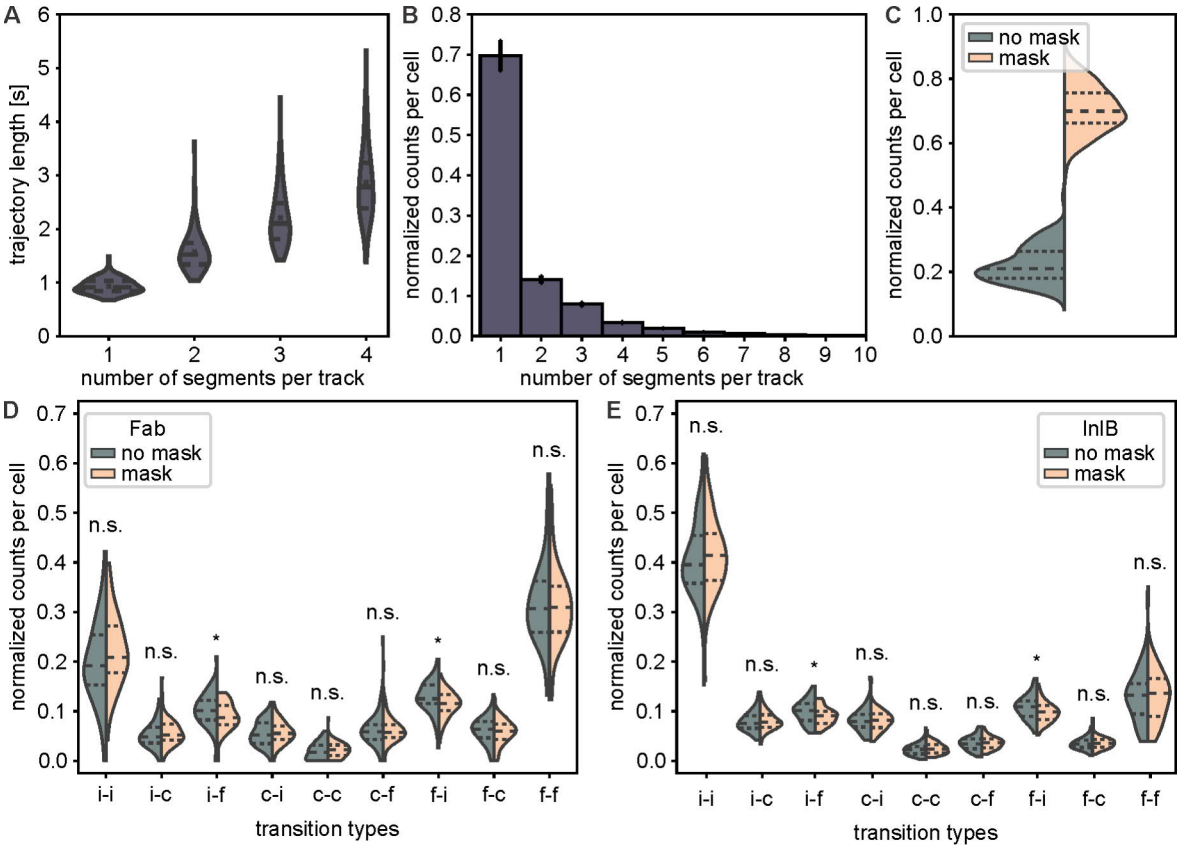

Figure S13: Reasons for low numbers in transition counts and masking effect. **A** dependency of trajectory length and number of segments. Lengths were averaged per cell. 120 cells per violinplot. **B** Number of segments per trajectory. All trajectories of the 120 analyzed cells (60 Fab, 60 InlB) that contain at least one classified segment are included. Error bars depict the SEMs. **C** Violin plot of the relative occurrence of transitions between classified segments without and with masking. Masking of segments below the classification threshold of 20 frames increases the occurrence of transitions between classified segments for trajectories consisting of at least three segments. N = 120 cells per distribution. Violin plots of the nine different transition types between segments within trajectories in **D** Fab- and **E** InlB-treated cells (N per distribution = 60 cells, i = immobile, c = confined, f = free). Left and right sides of the violin plots compare normalized counts without and with masking (masking value = 20 frames). Counts are normalized to one per cell. Transitions between not classified segments are neglected (Rahm et al., 2021, CC-BY).

13.4 Spatial organization of membrane receptors

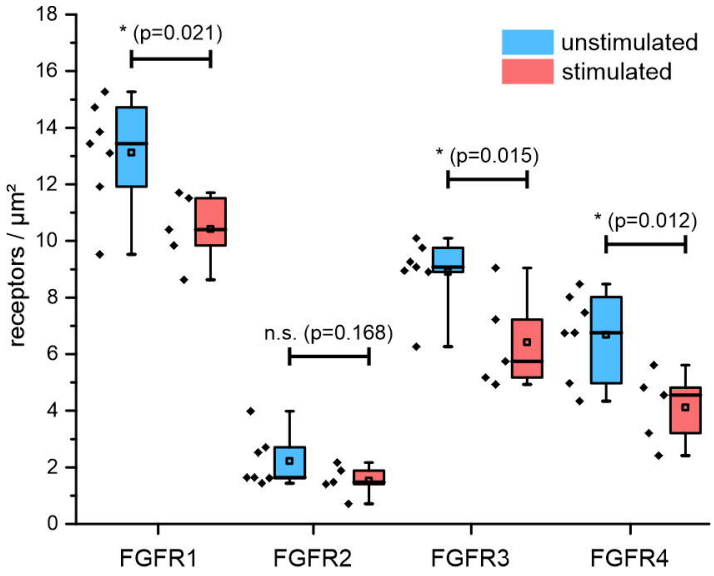

Figure S14: FGFR cluster densities. DBSCAN cluster analysis of DNA-PAINT images was used to assess the membrane receptor cluster density of FGFR1-4 in both unstimulated (blue) and FGF1-stimulated (red) U-2 OS cells, N = 7 (resting), N = 5 (stimulated), (Schröder et al., 2021).

14 Acronyms

AB	antibody
AdaGrad	adaptive gradient algorithm
Adam	adaptive moment estimation
AdamW	Adam with decoupled weight decay
APD	avalanche photo diode
BM3D	block matching and 3D
CAFI	content-aware frame interpolation
CBC	coordinate-based colocalization
CNN	convolutional neural network
DBSCAN	density-based spatial clustering of applications with noise
DECODE	deep context dependent
DIC	differential inference contrast
DNA-PAINT	DNA points accumulation for imaging in nanoscale topography
dSTORM	direct stochastic optical reconstruction microscopy
EMCCD	electron-multiplying charge-coupled device
EGF	epidermal growth factor
ETL	electrically tunable lens
EREG	epiregulin
ExM	expansion microscopy
FGFR	fibroblast growth factor receptor
FN	false negative
FOV	field of view
FP	false positive
FP	fluorescent protein
FRC	FOURIER ring correlation
FWHM	full width half maximum
GPU	graphics processing units
HAWKMAN	HAAR wavelet kernel analysis method for the assessment of nanoscopy
HER2	human epidermal growth factor receptor 2
HILO	highly inclined and laminated optical sheet
HMM	hidden Markov modeling
HOMO	highest occupied molecular orbital
ILSVRC	ImageNet large scale visual recognition challenge
IoU	intersection over union
ISBI	internatonal symposium on biomedical imaging
LLM	large language model
LOD	limit of detection

LR	low-resolution
LS	least squares
LSTM	long short-term memory
LUMO	lowest occupied molecular orbital
LPIPS	learned perceptual image patch similarity
MAE	mean absolute error
MEA	β-mercaptoethylamin
MINFLUX	maximally informative luminescence excitation
MJD	mean jump distance
MLE	maximum likelihood estimation
MLP	multi-layer perceptron
MNTB	medial nucleus of the trapezoid body (MNTB)
MSD	mean squared displacement
MSE	mean squared error
MS-SSIM	multi-scale structural similarity metric
Nadam	Nesterov momentum Adam
NAG	Nesterov accelerated gradient
NeNA	nearest neighbor analysis
NRGβ1	neuregulin
NLM	non-local means
NRMSE	normalized root mean squared error
PAINT	points accumulation for imaging in nanoscale topograph
PALM	photoactivated localization microscopy
PC-ODT	partially coherent optical diffraction tomography
PSF	point spread function
PSNR	peak signal to noise ratio
QPI	quantitative phase imaging
RCAB	residual channel attention block
RCAN	residual channel attention network
ReLU	rectified linear units
RESI	resolution enhancement by sequential imaging
RG	residual group
RIR	residual in residual
rFRC	rolling Fourier ring correlation
RI	refractive index
RMS	root mean square
RSE	resolution scaled error
RSF	resolution scaling function
SBT	sheet based tubes

SD	standard deviation
SED	standard error of the difference
SEM	standard error of the mean
SGD	stochastic gradient descent
SIM	structured illumination microscopy
SMLM	single-molecule localization microscopy
SNR	signal-to-noise ratio
SOFI	super-resolution optical fluctuation imaging
SPT	single-particle tracking
SR	super-resolution
SRRF	super-resolution radial fluctuations
SRM	super-resolution microscopy
SSIM	structural similarity metric
STED	stimulated emission depletion
STORM	stochastic optical reconstruction microscopy
TGFα	transforming growth factor α
TIRF	total internal reflection fluorescence
TN	true negative
TP	true positive
UMAP	uniform manifold approximation and projection
uPAINT	universal points accumulation for imaging in nanoscale topograph

15 List of Figures

1 Examples of deep learning enabled tasks for super-resolution microscopy . . 1
2 JABLONSKI diagram . . . 12
3 FRANCK-CONDON principle . . . 13
4 Exemplary structures of organic fluorophore. . . . 15
5 Crystal structure of GFP from the jellyfish *Aequorea victoria* . . . 15
6 Fluorescence labeling strategies . . . 16
7 Illumination schemes . . . 17
8 Stimulated emission depletion microscopy concept. . . . 21
9 Principle of single-molecule localization microscopy . . . 23
10 DNA-PAINT concept . . . 25
11 Exchange DNA-PAINT concept . . . 26
12 Clustering with DBSCAN . . . 27
13 k-nearest neighbor distances . . . 28
14 uPAINT concept . . . 30
15 Principle of trajectory computation . . . 30
16 Events occurring in single-particle tracking . . . 31
17 MSD calculation and curves of different diffusion states. . . . 32
18 Hidden MARKOV modeling of single-particle tracking data . . . 33
19 Phase shift and interference . . . 35
20 Köhler illumination . . . 35
21 Architecture of a multi-layer perceptron . . . 37
22 Activation function and a linear vs a non-linear classification problem . . . 38
23 Receptive field and the convolutional operation . . . 39
24 Different types of down- and upsampling layers . . . 40
25 Bias and variance tradeoff. . . . 46
26 Learning curve examination to detect bias and variance . . . 47
27 U-Net architecture . . . 48
28 Residual connection . . . 49
29 RCAN architecture . . . 50
30 *DeepSTORM* workflow with experimental DNA-PAINT data . . . 60
31 Optimization of the high density movie . . . 62
32 *DeepSTORM* performance on a low density DNA-PAINT movie . . . 63

33 *DeepSTORM* performance on a high density exchange DNA-PAINT movie 64
34 Number of frame optimization for complete image reconstruction with *DeepSTORM* . . . 65
35 DeepSTORM predicted image of high density DNA-PAINT quality assessment against the ground truth . . . 65
36 Comparison between *DeepSTORM*'s super-resolved image and rendered image from single-molecule localization . . . 66
37 STED denoising workflow . . . 76
38 STED denoising of ER in living cells . . . 77
39 STED denoising of living cells induced for stress . . . 79
40 Prediction of a 3D stack . . . 79
41 Hallucination artifacts . . . 81
42 Development of ER structure after drug treatment . . . 82
43 *ERnet* structure quantification . . . 83
44 Additional descriptors for dynamic ER structure quantification . . . 83
45 Virtual staining and segmentation workflow . . . 90
46 Predictions of fluorescence signal of cell membrane and nucleus with nucleus segmentation . . . 91
47 Generalizability towards different experimental conditions . . . 92
48 Prediction results of different augmentation strengths . . . 93
49 Prediction results of augmentation without intensity or zoom . . . 94
50 Predictions of different imaging modalities . . . 95
51 Prediction of the ER with different input modalities . . . 102
52 SSIM of the ER of different input modalities and over time. . . . 103
53 Workflow of the analysis of single-particle tracking data with SPTAnalyser. 107
54 Analysis pipeline and MSD-based results of single-particle tracking data of the MET membrane receptor. . . . 108
55 False negative rate of freely diffusing receptors that were classified as confined. 109
56 Single-molecule trajectories and quantification of transitions within trajectories. . . . 111
57 State transition diagrams of MET receptor trajectories from hidden MARKOV modeling. . . . 113
58 HER2 temporal responses to ligand treatment . . . 114
59 Distance distribution analysis . . . 122

60 Distance analysis of FGFR1-4 within untreated and FGF1-treated U-2 OS cells . . . 123
61 Circular counts of clathrin and FGFR1 in untreated and FGF1-treated U-2 OS cells . . . 124
62 k-nearest neighbor distances of FGFR1 to FGFR1-4 within untreated and FGF1-treated U-2 OS cells . . . 125
S1 Loss curves and learning rates of the training process . . . 159
S2 Hyperparameter importance . . . 159
S3 Quality control metrics of prediction from autophagy-induced cell dataset . 160
S4 Quality control metrics of prediction from the resting cell dataset . . . 161
S5 Prediction output from different model . . . 162
S6 Phase reconstruction and fluorescence deconvolution . . . 164
S7 Channel registration . . . 165
S8 Binary mask parameters . . . 166
S9 Metric comparison of different augmentation strengths . . . 167
S10 Metric comparison of augmentation without intensity or zoom . . . 168
S11 Metric comparison of nucleus as target channel vs nucleus and membrane as target channels . . . 169
S12 Metric comparison of different imaging modalities . . . 170
S13 Reasons for low numbers in transition counts and masking effect. . . . 171
S14 FGFR cluster densities . . . 172

17 Publications

A. N. Birtasu, K. Wieland, U. H. Ermel, **J. V. Rahm**, M. P. Scheffer, B. Flottmann, M. Heilemann, F. Grahammer, A. S. Frangakis, "The molecular architecture of the kidney slit diaphragm", bioRxiv, 2023.10.27.564405, 2023.

S. Jang, K. K. Narayanasamy, **J. V. Rahm**, A. Saguy, J. Kompa, M. S. Dietz, K. Johnsson, Y. Shechtman, M. Heilemann, "Neural network-assisted single-molecule localization microscopy with a weak-affinity protein tag", Biophys. Rep., 3, 100123, 2023.

C. Catapano, **J. V. Rahm**, M. Omer, L. Teodori, J. Kjems, M. S. Dietz, M. Heilemann, "Biased activation of the receptor tyrosine kinase HER2", Cell. Mol. Life Sci., 80, 158, 2023.

K. K. Narayanasamy, **J. V. Rahm**, S. Tourani, M. Heilemann, "Fast DNA-PAINT imaging using a deep neural network", Nat. Comm., 13, 5047, 2022.

J. V. Rahm, S. Malkusch, U. Endesfelder, M. S. Dietz, M. Heilemann, "Extraction of diffusion state transitions in single-particle tracking data of membrane receptors", Proc. SPIE, 11967, 1196705, 2022.

S. Malkusch, **J. V. Rahm**, M. S. Dietz, M. Heilemann, J.-B. Sibarita, J. Lötsch, "Receptor tyrosine kinase MET ligand-interaction classified via machine learning from single-particle tracking data", Mol. Biol. Cell, 33, 6, 33(6):ar60, 2022.

J. V. Rahm, S. Malkusch, U. Endesfelder, M. S. Dietz, M. Heilemann, "Diffusion State Transitions in Single-Particle Trajectories of MET Receptor Tyrosine Kinase Measured in Live Cells", Front. Comput. Sci., 3, 757643, 2021.

T. N. Baldering, C. Karathanasis, M.-L. I. E. Harwardt, P. Freund, M. Meurer, **J. V. Rahm**, M. Knop, M. S. Dietz, M. Heilemann, "CRISPR/Cas12a-mediated labeling of MET receptor enables quantitative single-molecule imaging of endogenous protein organization and dynamics", iScience, 24, 101895, 2021.

M. Schröder, M.-L. I. E. Harwardt, **J. V. Rahm**, Y. Li, P. Freund, M. S. Dietz, M. Heilemann, "Imaging the fibroblast growth factor receptor network on the plasma membrane with DNA-assisted single-molecule super-resolution microscopy", Methods, 193, 38-45, 2021.