Calibration-free single-frame super-resolution fluorescence microscopy

Anežka Dostálová, Dominik Vašinka, Robert Stárek, and Miroslav Ježek
Department of Optics, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 77900 Olomouc, Czechia

Molecular fluorescence microscopy is a leading approach to super-resolution and nanoscale imaging in life and material sciences. However, super-resolution fluorescence microscopy is often bottlenecked by system-specific calibrations and long acquisitions of sparsely blinking molecules. We present a deep-learning approach that reconstructs super-resolved images directly from a single diffraction-limited camera frame. The model is trained exclusively on synthetic data encompassing a wide range of optical and sample parameters, enabling robust generalization across microscopes and experimental conditions. Applied to dense terrylene samples with 150 ms acquisition time, our method significantly reduces reconstruction error compared to Richardson-Lucy deconvolution and ThunderSTORM multi-emitter fitting. The results confirm the ability to resolve emitters separated by 35 nm at 580 nm wavelength, corresponding to sevenfold resolution improvement beyond the Rayleigh criterion. By delivering unprecedented details from a single short camera exposure without prior information and calibration, our approach enables plug-and-play super-resolution imaging of fast, dense, or light-sensitive samples on standard wide-field setups.

I. INTRODUCTION

Optical imaging of single fluorescent molecules results in diffraction-limited point spread functions (PSFs). In modern biomedical imaging and microscopy, resolving structures at the molecular scale is essential, and therefore, many approaches seek to overcome this limit. The ability to resolve, localize, and count individual molecules is vital for understanding molecular processes, cellular dynamics, and material properties [1, 2]. Super-resolution imaging techniques have revolutionized biological and materials research, allowing the study of molecular structures and interactions with unprecedented detail. In cell biology, precise molecular localization is crucial for understanding subcellular architecture [3], tracking protein dynamics [4, 5], and mapping molecular pathways involved in processes such as intracellular transport [6]. In neuroscience, resolving individual synaptic proteins provides insights into synaptic plasticity and neurodegenerative diseases [7, 8]. In materials science and nanotechnology, the ability to detect and count single molecules aids in characterizing nanoscale heterogeneities and defects [9]. When molecules are sparsely distributed and their PSFs are isolated, localization methods can enhance the resolution significantly, down to 10 nm [10, 11]. However, this becomes notably more challenging at higher molecule densities, as overlapping PSFs hinder accurate localization. This challenge can be addressed by two main approaches, image reconstruction and stochastic molecule localization methods.

Image reconstruction methods are based on deconvolution using either direct inverse or statistical methods, such as maximum likelihood estimation. By computationally restoring high-frequency details lost during imaging, these techniques enhance resolution and are widely employed in super-resolution imaging [12–16]. In fluorescence microscopy, the observed image is formed as a convolution of the true fluorophore distribution with the PSF, along with added noise. Deconvolution seeks to recover the underlying distribution by estimating and reversing the blurring effects of the PSF. However, these methods rely on precise system-specific calibration and prior knowledge of the PSF. Their effectiveness diminishes with noise and may introduce artifacts, necessitating careful parameter tuning for each specific case.

Single-molecule localization microscopy (SMLM) offers an alternative solution by utilizing stochastically blinking emitters and accumulating data across many frames [17]. By using photoswitchable fluorophores, SMLM techniques enable the precise localization of individual molecules in successive frames [18–20]. The final super-resolved image is composed of the localizations from individual sparse frames. This allows for molecule localization even in dense samples but requires extended acquisition times, which restrains the application of these methods to live or moving samples. While SMLM techniques offer exceptional spatial resolution, they rely on the stochastic activation and deactivation of fluorescent emitters, and consequently require carefully optimized fluorophore selection, sample preparation, and precise control of the excitation [21, 22]. Their applicability is therefore restricted in delicate biological environments. Furthermore, localization needs thorough system-specific calibration and tuning to accurately extract molecule positions and intensities from noisy fluorescence images [23, 24]. The number of needed sequence frames can be reduced by multi-emitter fitting algorithms [25–27]. However, these methods are limited to certain molecule densities, struggle in the presence of high background noise, and are computationally demanding.

In recent years, integrating deep learning techniques into SMLM has significantly advanced the field, offering powerful tools for improving localization accuracy, accelerating image reconstruction, and addressing challenges associated with high emitter densities and low signal-to-noise ratios [28–30]. An example of such a technique is DeepSTORM [31], a convolutional neural network (CNN)-based approach that accelerates the image reconstruction process. However, the training of the model needs to be fine-tuned for particular experimental conditions separately, and the method is limited to blinking datasets. Another example is localization-based DECODE [32], a probabilistic model that predicts emitter positions and their corresponding uncertainties. As with all deep learning methods, DECODE performance decreases if the experimental conditions deviate from those in the training data. A step toward less calibration-dependent solution is the UniFMIR model [33], a foundation model that focuses on image restoration from noisy data. It is pre-trained on diverse available datasets, making it more generalizable to various microscopy modalities and biological structures. Fine-tuning is still needed for different imaging tasks. In general, achieving high-resolution imaging of individual molecules remains a fundamental challenge, particularly in dynamic or live samples, where time constraints preclude the acquisition of the thousands of frames required by SMLM techniques.

Here, we present a completely calibration-free deep learning model designed for super-resolving image reconstruction in samples with high emitter concentrations from a single frame, without any prior knowledge about the optical system, such as the shape of the point spread function or background noise distribution. Our model bypasses the drawbacks of localization-based methods by performing direct image reconstruction. It requires only a single dense image as input, eliminating the need for specialized datasets with stochastically blinking fluorophores and broadening applicability to a wide range of fluorescent markers. Furthermore, while conventional methods typically rely on acquiring thousands of frames—which can take several minutes per sample—our approach enables rapid measurement times. This significant reduction in acquisition time minimizes phototoxicity and photobleaching, which is particularly essential for imaging sensitive biological samples or when using sensitive organic fluorescent dyes.

Under realistic experimental conditions, parameters of the optical setup vary in time and are non-uniform over the field of view. Unlike the majority of state-of-the-art localization and reconstruction methods, which require extensive prior calibration information or frequent retraining, our model operates effectively without such demand, substantially simplifying experimental workflows and making it particularly well-suited for imaging dynamic samples. Additionally, also without the need for retraining or recalibration, our device-independent model is capable of handling variations in parameters such as different shapes or sizes of the PSFs, caused, e.g., by aberrations, non-uniform background distribution, etc., within the field of view. Our approach offers a fast, flexible solution for super-resolution imaging, enhancing accessibility and minimizing setup requirements across various microscopy platforms.

II. RESULTS AND DISCUSSION

We applied our calibration-free convolutional neural network (CFCNN) to fluorescence intensity images of terrylene molecules at various concentrations embedded in a polyvinyl alcohol (PVA) layer. These images were acquired using a fluorescence microscope, schematically shown in Fig. 1. The sample is illuminated by an excitation laser through a high numerical aperture oil-immersion objective lens, and the emitted fluorescence is collected by the objective and detected by an sCMOS camera. Due to the high molecule density and overlap

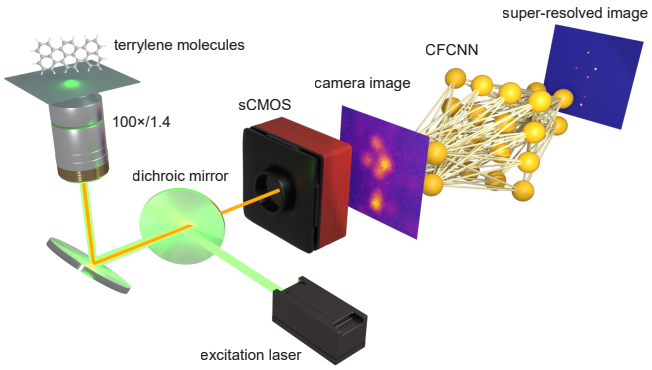

FIG. 1. Schematic illustration of the experimental fluorescence microscope setup and the CFCNN application to the camera image. Terrylene molecules in the sample emit fluorescent light upon laser excitation. The fluorescence is collected by the microscope objective, separated from the excitation light on the dichroic mirror, and detected by an sCMOS camera. The resolution-limited camera image is directly processed by the CFCNN without requiring any prior information or calibration. The CFCNN reconstructs a super-resolved image, revealing individual molecules with significantly enhanced resolution.

of their PSFs, obtaining a super-resolved counterpart for the single dense camera image poses a significant challenge for traditional super-resolution methods. For details on the experiment, please refer to subsection A of the Methods.

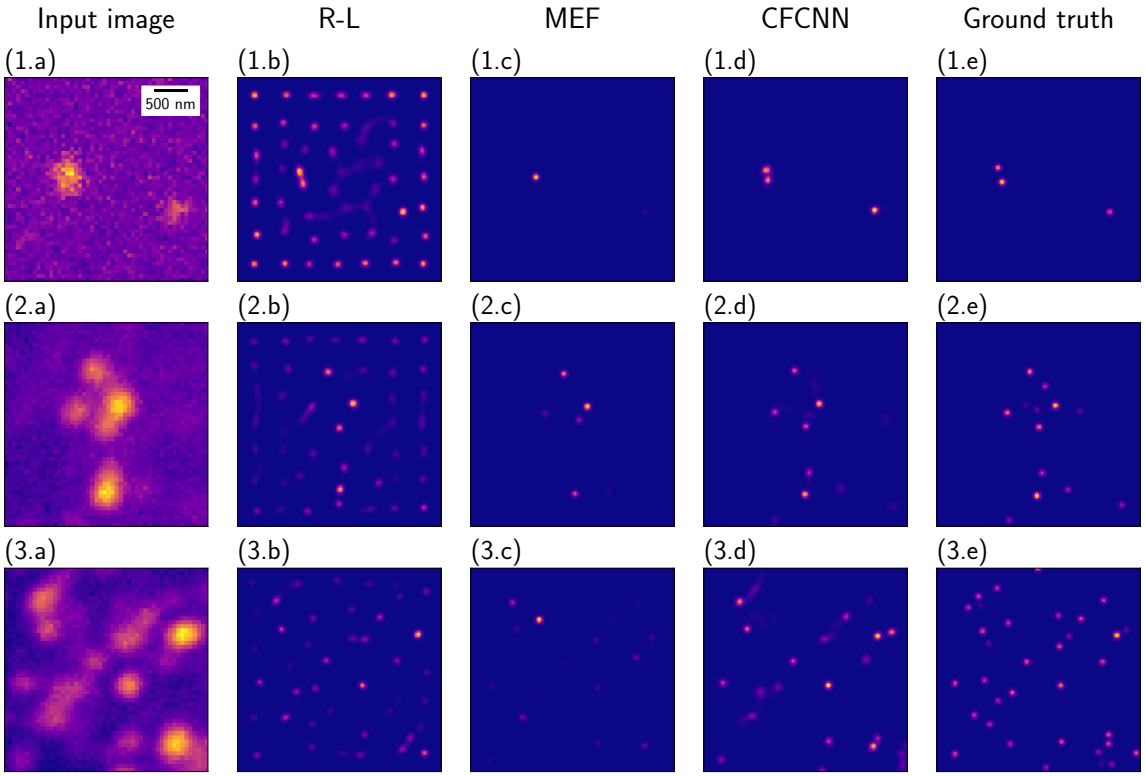

FIG. 2. A visual comparison of the results from two state-of-the-art methods for super-resolution imaging and localization, and our calibration-free neural network (CFCNN). All resulting images in panels (b)–(e) are convolved with a narrow Gaussian filter for better visualisation. Panels (a) show the input intensity images of organic terrylene molecules from a fluorescence microscope acquired by an sCMOS camera, with different molecule powers and densities. (b) Results of the reconstruction by the Richardson-Lucy deconvolution algorithm provided with a specific PSF. (c) The outputs of multi-emitter fitting using ThunderSTORM with specified camera settings. (d) Super-resolved images reconstructed by our CFCNN using no prior information or calibration. (e) An estimation of the ground truth using temporal information beyond the single intensity image.

Following the experimental setup, the camera image is processed by our CFCNN model. This deep-learning approach is calibration-free due to its unique methodology. The input to our model is a single intensity image, and the output is an upsampled, high-resolution reconstruction. The model is based on a convolutional neural network trained exclusively on synthetic data. This allows for the model to learn across a vast range of optical system parameters, covering configurations that would be impossible to obtain experimentally at such a scale, and eliminating systematic errors inherent in experimental measurements. Additionally, our network is not constrained to a single specific optical configuration. Instead, it is trained on diverse combinations of noise levels, PSF profiles, numerical apertures, and other system parameters. This broad training scope enables the model to generalize effectively across various experimental conditions, making it robust both to variations in optical setup parameters across different systems and to variations within a single field of view, including aberrations, non-uniform background distributions, and other spatially varying changes. For further details on the CFCNN architecture and its training, please refer to the Methods, subsection B.

To validate and benchmark the performance and versatility of our CFCNN, we also employed two commonly used state-of-the-art methods on the same input images, specifically, the Richardson-Lucy (R-L) deconvolution algorithm as a reconstruction technique [34, 35] and multi-emitter fitting (MEF) with ThunderSTORM (an ImageJ plugin) as a localization method [23]. All three approaches were provided exclusively with a single resolution-limited image. The amount of additional information and parameter settings required by each method differs, as well as the necessary post-processing steps to produce the final super-resolved image. R-L algorithm iteratively applies Bayes theorem to update an estimate of the true image by maximizing the likelihood of the observed image. The algorithm can achieve significant resolution enhancement. However, it relies on

		R-L	MEF	CFCNN
Image 1	MAE [10^{-5}]	4.64(8)	3.99*	2.4(5)
	KLD	6.0(2)	12.6*	1(2)
Image 2	MAE [10^{-5}]	3.98(7)	3.6(1)	2.27(8)
	KLD	7.8(7)	17.0(2)	2.1(4)
Image 3	MAE [10^{-5}]	3.52(3)	4.01(2)	3.34(4)
	KLD	10.0(3)	19.0(4)	5.2(2)

TABLE I. Quantitative similarity metrics calculated between the super-resolved experimental fluorescence microscopy images relative to the ground truth for the Richardson-Lucy deconvolution, multi-emitter fitting using ThunderSTORM, and the CFCNN. The digit in parentheses represents one standard deviation, e.g. 4.64(4) = 4.64±0.04. Asterisk denotes cases where the standard deviation is lower by at least 3 orders of magnitude than the measured value.

accurate prior knowledge of the PSF profile, which necessitates either acquiring a calibration image of a sparse sample or the presence of an isolated molecule within the field of view. Besides being a time-consuming extra step, this also introduces a simplification, as the PSF shape may vary across the field of view. ThunderSTORM offers extensive customization options tailored to the user's specific application. However, this also means the results are highly dependent on a large number of input parameters, such as the camera settings, image filtering, fitting algorithm method and parameters, etc. Their values need to be optimized for each given setup and measurement. The optimization involves balancing multiple trade-offs, requiring user expertise and adding complexity to the process. The output of the method is a table containing the emitter coordinates and intensities. Subsequent image rendering may lead to ambiguity since there are multiple ways to perform this step, and requires additional decisions in the processing pipeline. Details on the settings for each method are given in part C of the Methods section. In contrast, our deep-learning approach requires no additional information beyond the input image neither before nor after applying the CFCNN, and the resulting super-resolved image outperforms the other methods across all evaluated metrics.

The visual comparison between the input intensity image of fluorescent terrylene molecules, the reconstructed outputs from Richardson-Lucy deconvolution, multi-emitter fitting using ThunderSTORM, and our CFCNN, is presented in Fig. 2. The output images are accompanied by the ground truth, obtained by leveraging the limited photostability of the molecules in PVA and their gradual localization from a recorded time-lapse of bleaching events (see the Methods section, part D, for details). We show the results for three independent experimental intensity images with different molecule concentrations and levels of background noise. The input image size is 50×50 pixels and all outputs are upscaled to 200×200 pixels to enhance the resolution and detail of the reconstructed image. The multi-emitter fitting and our manual localization technique for obtaining the ground truth both produce a list of localized emitters, which are then rendered into the finer grid. The situation is different for R-L deconvolution, which directly returns an image of the original size. In this case, the input image was first upscaled via bicubic interpolation before applying the reconstruction algorithm. Finally, the CFCNN has upsampling layers incorporated in its architecture. Since the reconstructions consist of 200×200 pixels containing small structures, they are convolved with a narrow Gaussian filter (with $\sigma = 2$ px) for visualisation purposes in the figure. For further details on the methods employed and the process used to obtain the ground truth, please refer to the Methods section.

The performance of R-L deconvolution, shown in panels (b), is highly sensitive to the noise level in the data, and the algorithm accuracy deteriorates as the SNR decreases. It introduces artifacts by placing emitters at approximately equidistant positions, effectively filling noisy regions with emitters not present in the original data. This leads to a loss of accuracy and the appearance of artificial structures in the reconstructed image. This effect is most pronounced in Fig. 2(1.b) and is also clearly visible in (2.b).

The resulting images from the MEF regime of ThunderSTORM are shown in panels (c) of Fig. 2. Without temporal information to distinguish overlapping emitters, the algorithm tends to produce a single emitter at the center of mass of a diffraction-limited spot, as clearly seen in Fig. 2(1.c). It cannot determine the actual number of molecules, leading to localization errors. It is evident for all three images that MEF missed a great number

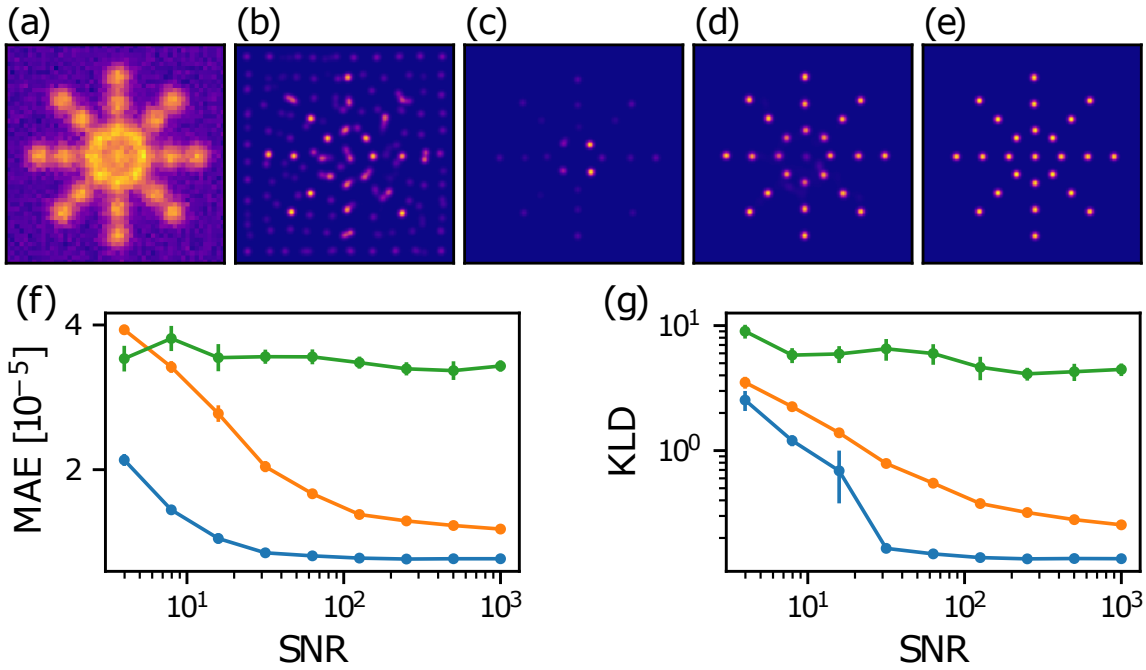

FIG. 3. Synthetic data with emitters arranged in the shape of a star. Images in panels (b)–(e) are convolved with a narrow Gaussian filter for better visualisation. (a) An example of the input image with SNR = 15.8 obtained as the ground truth convolved with a Gaussian PSF with the width of 2 px on the 50×50 pixel grid. (b) The image reconstruction by the Richardson-Lucy deconvolution algorithm. (c) The output of multi-emitter fitting using ThunderSTORM. (d) The reconstructed image by the CFCNN. (e) The real generated emitter distribution. (f) Mean absolute error between the outputs and the ground truth as a function of SNR for R-L (orange), MEF (green), and the CFCNN (blue). (g) Kullback–Leibler divergence of the outputs from the ground truth as a function of SNR.

of emitters, and some were localized incorrectly.

In contrast, our CFCNN demonstrates superior performance across the field of view regardless of background noise level. It accurately reconstructs individual molecules directly from the single intensity frame, without any prior information about the input image origin. It surpasses the Rayleigh criterion limit and achieves a level of super-resolution that typically requires large datasets of sparsely distributed emitters. Panels (d) in Fig. 2 clearly illustrate the ability of the CFCNN to distinguish emitters even in dense regions of the sample.

For quantitative performance evaluation, we compared the output of each method against the ground truth utilizing the mean absolute error (MAE) and Kullback–Leibler divergence (KLD). MAE measures pixel-wise accuracy, while KLD captures the distance of probability distributions. Together, they provide a comprehensive assessment of both localization precision and distributional similarity. These metrics are evaluated relative to the ground truth of each sample image and summarized in Table I with statistical uncertainties. Given the range of factors affecting the absolute values, we present three independent experimental images (shown in Fig. 2) to highlight the trends of the methods and provide a comparative analysis of their performance. It is evident that the CFCNN consistently outperforms the other methods, even with no requirement of prior information and calibration. This underscores the robustness and effectiveness of our approach, especially in challenging scenarios with high emitter concentrations, low SNR, and variable experimental conditions.

To further investigate the methods performance beyond the experimental results, we conducted additional tests on synthetic datasets with controlled parameters and precisely known ground truth, which allows for accurate and reliable evaluation. Additionally, it enables a systematic study of the performance dependence on the intensity noise level. Throughout this work, we define the SNR as the difference between the maximum intensity and the average noise level, divided by the standard deviation of background noise. In Fig. 3(a)–(e), we show a particular example of emitters arranged in the shape of a star. Panel (a) represents the input for the tested methods with the SNR = 15.8, while panel (e) visualizes the generated ground truth with applied Gaussian filter. For the outputs of each studied method, also convolved

with the narrow Gaussian filter, see panels (b)–(d). It is immediately apparent that the CFCNN reconstruction is virtually perfect, with all the emitters correctly positioned and their intensities accurately estimated. On the contrary, the R-L reconstruction contains numerous artificial emitters, not only in the vicinity of the true emitter positions but across the entire field of view. This makes it misleading and significantly complicates distinguishing genuine emitters from artifacts. The multi-emitter fitting struggles in dense regions and inaccurately estimates the emitter intensities, resulting in a compromised image. Some of the emitters were not resolved at all, leading to a loss of valuable information. We also evaluated MAE and KLD between the outputs and the well-defined ground truth. Fig. 3 (f) and (g) depict the dependence of these metrics on the noise level expressed by SNR for all three methods. Our CFCNN consistently outperforms both the R-L algorithm and the MEF method across all SNR values.

Futhermore, we generated a standalone dataset consisting of a pair of emitters with progressively increasing separation to assess the level of super-resolution achievable by our CFCNN. The followed procedure is outlined in the Methods section, part E. For the minimum resolvable separation between two emitters reconstructed by the CFCNN as a function of SNR, see Fig. 4. This separation δ_R is expressed in terms of the Rayleigh distance. It is evident that a significant level of super-resolution is attained even at the lowest SNR values. For higher SNR values, the achieved spatial resolution saturates towards $\delta_R = 0.138$, corresponding to more than sevenfold enhancement. This saturation limit is imposed by the finite pixel size, and we anticipate that finer image sampling would further enhance the resolution. For the experimental images of terrylene molecules with an emission peak at 580 nm, taken with a microscope objective with numerical aperture 1.4, the CFCNN reaches the resolution of 35 nm. Furthermore, while SMLM typically achieves up to tenfold resolution improvement, it requires a long sequence of sparse images and meticulous calibration. By applying the CFCNN to each frame, we could reach the same level of super-resolution while significantly reducing the number of images needed, making the process considerably faster.

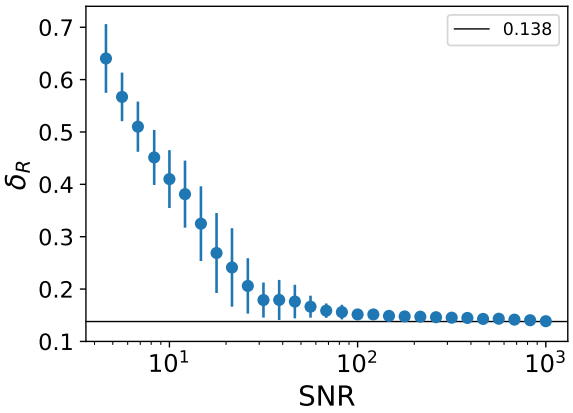

FIG. 4. The level of super-resolution achieved by the CFCNN depending on the signal-to-noise ratio using a simulated dataset. The resolvable separation δ_R between two identical emitters is expressed in terms of the Rayleigh distance. A significant spatial resolution enhancement relative to the Rayleigh criterion is achieved across the entire SNR range, exceeding a sevenfold improvement for higher SNR values and attained solely from a single camera image.

III. CONCLUSION

We presented a calibration-free convolutional neural network for single-frame super-resolution fluorescence microscopy. Unlike conventional methods, this CFCNN operates without any calibration, retraining, or prior information about the imaging system, making it highly versatile and universally applicable. Trained on a vast synthetic dataset, the model has the ability to generalize across different experimental conditions, including variations in noise levels, point spread functions, and optical configurations. We demonstrated its effectiveness on experimental fluorescence microscopy images of terrylene molecules. We compared the super-resolved image reconstructions with the state-of-the-art Richardson-Lucy algorithm provided with a precise PSF, and the ThunderSTORM multi-emitter fitting with numerous tuned input parameters and specification of the camera properties. We showed that the calibration-free CFCNN consistently outperforms the other methods, on both the experimental images and synthetic data, across all SNR values. The graphical visualization of the super-resolved images clearly demonstrates the superiority of the CFCNN, producing clean and credible reconstructions, while the other methods introduce artifacts and information loss. This visual advantage is supported by quantitative evaluation. Namely, the mean absolute error and Kullback–Leibler divergence consistently demonstrate the superior performance of the CFCNN compared to the system-specific Richardson-Lucy algorithm and multi-emitter fitting methods.

Furthermore, we quantified the super-resolving ability of the CFCNN, achieving up to sevenfold enhancement relative to the standard Rayleigh limit. In our experiment with terrylene, with an emission wavelength of 580 nm, this corresponds to resolving fluorophores separated by 35 nm using a single camera frame with expo-

sure time of 150 ms. The single-frame operation reduces measurement times down to tens of milliseconds, allowing real-time super-resolution imaging of dynamic and live samples that would be inaccessible by slower multi-frame techniques, and phototoxic specimens. Moreover, it relaxes the sparsity condition of blinking emitters, significantly broadening its applicability to a wider range of fluorescent markers. When combined with SMLM approaches, our CFCNN could notably shorten the overall acquisition time by substantially reducing the number of frames, while simultaneously improving accuracy.

The presented work demonstrates the potential of a calibration-free deep learning approach for rapid super-resolution fluorescence microscopy. Future improvements, such as incorporating additional upsampling layers for greater image magnification, introducing residual connections for enhancing the propagation of information [36], and using deeper network architectures, could further enhance the versatility and performance of our approach. With these further advancements, we envision our network becoming a completely universal tool for high-precision microscopy, broadly applicable across numerous domains ranging from biomedical imaging to material sciences.

IV. METHODS

A. Experiment

The primary imaging setup consists of a fluorescence microscope operating in a wide-field imaging configuration, equipped with an sCMOS camera (Andor Marana 4.2B-6) with 6.5 μm pixel size. This system enables diffraction-limited image acquisition of fluorescent samples. A 532 nm excitation laser is directed towards the sample through a high numerical aperture oil-immersion objective lens (OLYMPUS 100×/1.4), allowing optimal excitation and collection of the emitted fluorescence. The Stokes-shifted fluorescence is separated from the excitation light path on a dichroic mirror and detected by the camera, onto which it is focused using an achromatic doublet with a focal length of 200 mm as a tube lens. In combination with the objective of 100× magnification, this results in the overall system magnification of $M \approx 111$. A laser-line filter is used to suppress unwanted red light from the diode-pumped solid state excitation laser. Two 550 nm long-pass filters in the collection path prevent any residual laser light from reaching the camera sensor.

The samples contain terrylene molecules embedded in polyvinyl alcohol (PVA) and are prepared by spin coating. The PVA powder was dissolved in 4 ml of warm deionized water to form a 5% w/v solution. We added 100 μl of 2×10^{-5} % w/v terrylene-in-toluene solution and mixed the resulting solution by sonication. Then we spin-coated (30 seconds at 1500 rpm) a 30-μl drop of the mixture on an clean coverslip. This technique creates a thin layer of PVA containing the fluorescent molecules across the coverslip. The emission spectrum of terrylene has a peak at 580 nm.

The PVA forms an environment where terrylene exhibits very limited photostability, with some of the individual molecules undergoing several blinking cycles before irreversibly photobleaching. We captured a sequence of images from the same location within the sample, with progressively fewer active emitters. The acquisition time of each frame was 150 ms. Only the first and densest frame was used as input for all the studied methods–our CFCNN, R-L deconvolution, and MEF–while the subsequent frames served solely as reference data and were utilized to generate the ground truth, providing a high-resolution benchmark against which the super-resolved outputs of all three methods were compared.

B. Calibration-free convolutional neural network

We designed a calibration-free deep learning model for super-resolution imaging. This fully convolutional neural network reconstructs a high-resolution image from a single intensity frame. Consisting of 25 hidden convolutional layers with 50 filters each, the model performs a nonlinear transformation on the input. Information propagates through 5×5 convolutional kernels, performing localized feature extraction across layers. To implement nonlinearity and prevent the dying ReLU problem [37], we employ a LeakyReLU activation function with a negative slope of 0.3. The final layer uses softmax activation [38] to enforce the normalization and non-negativity of the output. Each convolutional layer includes dropout regularization, randomly setting 1% of input unit values to zero during training to mitigate overfitting [39]. Additionally, we integrate upsampling layers to incorporate interpolative resizing into the network architecture. Positioned after the 5th and 10th convolutional layer, each upsampling doubles the input dimensions, resulting in a fourfold image magnification. Furthermore, due to the fully convolutional-based design, the model can process images of arbitrary shapes and sizes, making it highly adaptable for various imaging applications.

We train the model using a simulated dataset of low-resolution 50×50 pixel images paired with their corresponding 200×200 high-resolution ground truth representations. The dataset is generated using diverse optical conditions to ensure broad applicability across different imaging scenarios. To create each sample, we randomly

assign emitter intensities and positions within a 200×200 pixel grid. The number of emitters in a sample is randomly selected from the $[1, 25]$ range, while their intensities follow a Poisson distribution with a mean value ranging between $[1, 3000]$ photons. The resulting matrix represents our ground truth image. Next, we simulate the effects of finite resolution by convolving the ground truth matrix with a point spread function. We consider two distinct PSF shapes: A Gaussian function, commonly used to approximate high numerical aperture imaging conditions, and an Airy disc, representing lower numerical aperture scenarios [40]. The full width at half maximum of the PSF varies between approximately $[2.5, 7.5]$ pixels in the 50×50 grid, equivalent to $[10, 30]$ pixels in the 200×200 grid. Moreover, we introduce PSF asymmetry by applying a squeezing transformation along a random axis. Following the convolution, the image is downsampled to 50×50 pixels, and background noise is added, with intensity values randomly selected from the $[1, 100]$ photon range per pixel. Each dataset sample is generated independently, ensuring a unique combination of the number of emitters, intensity, noise level, PSF shape, width, and asymmetry.

To train the model, we employed incremental learning, a dynamic approach where newly generated data pairs gradually replace the existing samples during training [41]. This combination of incremental learning and data simulation is particularly beneficial for developing universal networks capable of handling diverse imaging conditions. Throughout 90 incremental iterations, the model processed nearly 500,000 data pairs, with each iteration consisting of 50 epochs of training on mini-batches of 32 samples. The training was performed using the Adam optimizer, which leverages adaptive moment estimation for efficient gradient descent [42]. The initial learning rate was set to 10^{-4}. Given the high sparsity of the ground truth images, we used an adjusted mean squared error loss function. Firstly, both the ground truth I and the model prediction $\hat{I}$ are convolved with a Gaussian filter g of $\sigma = 2$. Then, we compute the squared difference between the filtered images and average it over a batch of N samples, $\frac{1}{N}\sum_{i=1}^{N} \|I_i * g - \hat{I}_i * g\|_2^2$. Similarly, an adjusted mean absolute error metric, $\frac{1}{N}\sum_{i=1}^{N} \|I_i * g - \hat{I}_i * g\|_1^2$, was used to monitor convergence and evaluate the model performance. The presented hyperparameter configuration results from manual optimization of network architecture and dataset ranges to enhance the performance across the device-independent regime. We expect further improvement by using a larger network architecture. However, the computational resources, namely the required operating memory for its training, are presently beyond our technical capacities.

C. Richardson-Lucy algorithm, ThunderSTORM

We employed the Richardson-Lucy algorithm [34, 35] and multi-emitter fitting using ThunderSTORM [23] on both the experimental fluorescence microscopy images and synthetic data as relevant representatives of state-of-the-art methods for comparison with our deep learning approach.

Richardson-Lucy algorithm recovers the true image by performing deconvolution, i.e., reversing the blurring caused by a known point-spread function. It is an iterative procedure, maximizing the likelihood that the observed blurred image corresponds to the true image convolved with the PSF. Therefore, the precise PSF, an initial guess, and the number of iterations or error threshold must be specified and are user-dependent. In our application, a uniform image was chosen as a starting point, and the iterations were stopped when the MAE between consecutive iterations satisfied $\epsilon \leq 10^{-10}$. To achieve the 200×200 pixel output size, we upscaled the camera image using bicubic interpolation prior to applying the R-L reconstruction. Upscaling before reconstruction ensures that the R-L algorithm operates on the image at the target resolution, preserving the intended precision. Performing the reconstruction first and interpolating the result afterward would handicap the method by introducing artifacts, which could distort the fine details in the deconvolved image. For each analyzed dataset of blinking emitters, only the single densest image was given to the reconstruction algorithm. The rest of the sequence provided a reference where a few isolated emitters were fitted with a Gaussian PSF to set the optimal kernel for the deconvolution.

ThunderSTORM is an open-source ImageJ plugin designed to analyze data obtained through SMLM techniques. The default settings of the software are to produce reasonable results on various datasets. To optimize the results, the users can adapt the analysis to suit their particular data using the included processing and post-processing methods and tools. This customization, however, requires experience and extensive knowledge about the system under study. While users usually have this knowledge, the process is time-consuming and complex, involving trade-offs that must be reconsidered whenever conditions change. Moreover, the same settings may not be optimal across the entire field of view and should always be adjusted for different experimental realizations to yield the best results. In our analysis, we enabled multi-emitter fitting and specified the details of our imaging setup and camera (photoelectrons per A/D count, etc.). When applied, the method returns a table as a result, which contains localized emitter coordinates and intensities. We generated the image by rendering the in-

formation from the table into a 200×200 pixel grid using the nearest pixel method.

It is important to note again that all the studied methods were provided only with the single densest fluorescence image of organic molecules. We recognize that ThunderSTORM would achieve notably better performance with access to the full image sequence. However, our objective is to compare these methods on dense samples without relying on additional data or imposing specific experimental requirements.

D. Ground truth determination

When testing a super-resolution method on experimental data, the fundamental challenge is the lack of an absolute ground truth. Unlike simulated datasets, where ground truth information is inherently available, real fluorescence microscopy data provide no direct reference for evaluating reconstruction accuracy. To address this, we leveraged the limited photostability of terrylene in PVA to evaluate the performance of all methods on the realistic fluorescence microscopy dataset. Sequential images of the sample were captured, with emitters gradually photobleaching over time. This natural reduction in emitter density allowed the application of a standard localization method on the emitters that became isolated and distinguishable throughout the sequence and on the temporal difference of the frames. For the ground truth construction, we began with the final image of the sequence, where only a few distinguishable emitters remained active. These emitters were fitted with Gaussian profiles, enabling precise localization of their positions and intensities. This final frame was subtracted from the preceding one to produce another sparse image, allowing additional emitters to be isolated and localized. This process was iterated over the image sequence, progressively building a localization table, which was then rendered into a 200×200 pixel ground truth image using the nearest pixel method. By starting from the sparsest frame and working backward, this approach minimizes localization errors and ensures consistency across the sequence.

The resulting manual localization serves as a benchmark for evaluating the performance of the studied methods on the realistic fluorescence microscopy dataset. For quantitative comparison, we calculated two metrics between the visulized images $A = A' * g$ and $B = B' * g$, where A', B' denote the reconstructed images and the ground truth, respectively, which are convolved with a narrow Gaussian visualization filter g of $\sigma = 2$. All the presented images are normalized to a unit L_1 norm after subtracting the minimum intensity so that the relative intensity distribution is retained,

$$P_i(A) = \frac{A_i - \min A}{\sum_{j=1}^{N} (A_j - \min A)}, \tag{1}$$

where N is the number of pixels. The mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} |P_i(A) - P_i(B)| \tag{2}$$

quantifies the average absolute difference between corresponding pixel values of two images. The Kullback-Leibler divergence (KLD)

$$\mathrm{KLD} = \sum_{i=1}^{N} P_i(B) \log_2 \left(\frac{P_i(B)}{P_i(A)} \right) \tag{3}$$

evaluates the difference between the probability distributions of the pixel values in the two images. To estimate statistical errors, we used the following approach. Our bootstrapping model consists of Poissonian emitters on top of a normally distributed background. We estimated the parameters of background distribution from data regions without emitters, and we determined the λ value of the Poissonian signal as a distance from the mean value of the background. This model accurately approximates the statistical noise characteristics of our camera. We generated 100 random samples with that model and inserted them into tested reconstruction methods to estimate the standard deviation for each tested metric. Such an error bar mainly determines the numerical stability of the reconstruction method.

E. Synthetic data and super-resolution

To systematically evaluate and compare the performance of the studied methods, we generated synthetic datasets designed to mimic a wider range of experimental conditions, for example the background noise level. These datasets consist of simulated fluorescence images with varying emitter numbers, intensities, spatial distributions, structural complexity, and noise levels.

The emitters were positioned in specified patterns within a defined field of view, serving as the absolute ground truth. The availability of this exact ground truth is a key advantage of the simulated datasets. The synthetic data thus provides a fully reliable reference for method evaluation. To simulate the optical system, we convolved the δ-functions at the emitter positions with a PSF modeled as a two-dimensional Gaussian function with a standard deviation corresponding to the imaging conditions, replicating the diffraction-limited appearance

of molecules in fluorescence microscopy. To assess robustness against experimental imperfections, we incorporated different levels of Gaussian background noise. The SNR was systematically varied to study its impact on reconstruction quality.

Using the synthetic data, we demonstrated the performance of our calibration-free convolutional neural network, R-L deconvolution, and MEF using ThunderSTORM, depending on the amount of background noise in the input image by varying the SNR. The reconstruction accuracy was again quantified by computing MAE and KLD between the output images and the ground truth. Monte Carlo simulations were performed to estimate error bars for each data point.

Furthermore, we assessed the super-resolution achievable by our CFCNN. To this end, we generated a dataset consisting of emitter pairs with identical intensity. In the first frame, the emitters were perfectly overlapping, and their separation was incrementally increased in subsequent frames within the range of $[0, 1]$ of the Rayleigh distance, with a step size of $1/80$ of the Rayleigh distance. Since the Rayleigh criterion is based on the first minimum of an Airy pattern, this distance is approximately equivalent to three standard deviations of Gaussian emitter profiles. Each frame in the sequence was processed by the CFCNN, and the resulting images were analyzed by their marginal intensity distributions. The modulation in these distributions determines the CFCNN's ability to resolve closely spaced emitters. Following the Rayleigh criterion, the emitters are resolved if the modulation dip has a relative depth of 26.3%. The same sequence was analyzed across an SNR range of $[1, 1000]$ with a logarithmic step and 100 Monte Carlo samples per data point.

DATA AVAILABILITY

The scripts and data analyzed in this study are publicly accessible on GitHub [43].

ACKNOWLEDGEMENTS

We acknowledge the use of cluster computing resources provided by the Department of Optics, Palacký University Olomouc. We thank J. Provazník for maintaining the cluster and providing support.

FUNDING

This work was supported by the Czech Science Foundation (project 21-18545S), and the Ministry of Education, Youth, and Sports of the Czech Republic (project OP JAC CZ.02.01.01/00/23 021/0008790). AD and DV acknowledge the support by Palacký University Olomouc (projects IGA-PrF-2024-008 and IGA-PrF-2025-010).

[1] A. Dani, B. Huang, J. Bergan, C. Dulac, and X. Zhuang, Superresolution imaging of chemical synapses in the brain, *Neuron* **68**, 843–856 (2010).

[2] A. Gahlmann and W. E. Moerner, Exploring bacterial cell biology with single-molecule tracking and superresolution imaging, *Nature Reviews Microbiology* **12**, 9–22 (2014).

[3] M. Sauer and M. Heilemann, Single-molecule localization microscopy in eukaryotes, *Chemical Reviews* **117**, 7478–7509 (2017).

[4] Y. Cui, M. Yu, X. Yao, J. Xing, J. Lin, and X. Li, Single-particle tracking for the quantification of membrane protein dynamics in living plant cells, *Molecular Plant* **11**, 1315–1327 (2018).

[5] G. A. Dunn, I. M. Dobbie, J. Monypenny, M. R. Holt, and D. Zicha, Fluorescence localization after photobleaching (FLAP): a new method for studying protein dynamics in living cells, *Journal of Microscopy* **205**, 109–112 (2002).

[6] M. L. Zhang, H. Y. Ti, P. Y. Wang, and H. Li, Intracellular transport dynamics revealed by single-particle tracking, *Biophysics Reports* **7**, 413–427 (2021).

[7] C. Sun, Single-molecule-resolution approaches in synaptic biology, *The Journal of Physical Chemistry B* **128**, 3061–3068 (2024).

[8] P. Padmanabhan, A. Kneynsberg, and J. Götz, Super-resolution microscopy: a closer look at synaptic dysfunction in Alzheimer disease, *Nature Reviews Neuroscience* **22**, 723–740 (2021).

[9] J. Feng, H. Deschout, S. Caneva, S. Hofmann, I. Lončarić, P. Lazić, and A. Radenovic, Imaging of optically active defects with nanometer resolution, *Nano Letters* **18**, 1739–1744 (2018).

[10] D. Aquino, A. Schönle, C. Geisler, C. v Middendorff, C. A. Wurm, Y. Okamura, T. Lang, S. W. Hell, and A. Egner, Two-color nanoscopy of three-dimensional volumes by 4Pi detection of stochastically switched fluorophores, *Nature Methods* **8**, 353–359 (2011).

[11] F. Huang, G. Sirinakis, E. S. Allgeyer, L. K. Schroeder, W. C. Duim, E. B. Kromann, T. Phan, F. E. Rivera-Molina, J. R. Myers, I. Irnov, M. Lessard, Y. Zhang, M. A. Handel, C. Jacobs-Wagner, C. P. Lusk, J. E. Rothman, D. Toomre, M. J. Booth, and J. Bewersdorf, Ultra-high resolution 3D imaging of whole cells, *Cell* **166**, 1028–1040 (2016).

[12] J. J. de Rooi, C. Ruckebusch, and P. H. C. Eilers, Sparse deconvolution in one and two dimensions: Applications in endocrinology and single-molecule fluorescence imaging, *Analytical Chemistry* **86**, 6291–6298 (2014).

[13] H. Jang, Y. Li, A. A. Fung, P. Bagheri, K. Hoang, D. Skowronska-Krawczyk, X. Chen, J. Y. Wu, B. Bintu, and L. Shi, Super-resolution SRS microscopy with A-PoD, *Nature Methods* **20**, 448–458 (2023).

[14] S. Martínez, M. Toscani, and O. E. Martinez, Superresolution method for a single wide-field image deconvolution by superposition of point sources, *Journal of Microscopy* **275**, 51–65 (2019).

[15] P. Zhang, C. Gao, Z. Zhang, Z. Yuan, Q. Zhang, P. Zhang, S. Du, W. Zhou, Y. Li, and S. Li, Systematic inference of super-resolution cell spatial profiles from histology images, *Nature Communications* **16**, 1838 (2025).

[16] E. A. Mukamel, H. Babcock, , and X. Zhuang, Statistical deconvolution for superresolution fluorescence microscopy, *Biophysical Journal* **102**, 2391–2400 (2012).

[17] M. Lelek, M. T. Gyparaki, G. Beliu, F. Schueder, J. Griffié, S. Manley, R. Jungmann, M. Sauer, M. Lakadamyali, and C. Zimmer, Single-molecule localization microscopy, *Nature Reviews Methods Primers* **1**, 39 (2021).

[18] M. Rust, M. Bates, and X. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM), *Nature Methods* **3**, 793–796 (2006).

[19] U. Endesfelder and M. Heilemann, Direct stochastic optical reconstruction microscopy (dSTORM), *Methods in Molecular Biology* **1251**, 263–276 (2015).

[20] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, Imaging intracellular fluorescent proteins at nanometer resolution, *Science* **313**, 1642–1645 (2006).

[21] G. Crivat and J. W. Taraska, Imaging proteins inside cells with fluorescent tags, *Trends in Biotechnology* **30**, 8–16 (2012).

[22] S. M. Hickey, B. Ung, C. Bader, R. Brooks, J. Lazniewska, I. R. D. Johnson, A. Sorvina, J. Logan, C. Martini, C. R. Moore, L. Karageorgos, M. J. Sweetman, and D. A. Brooks, Fluorescence microscopy-an outline of hardware, biological handling, and fluorophore considerations, *Cells* **11**, 35 (2021).

[23] M. Ovesný, P. Křížek, J. Borkovec, Z. Švindrych, and G. M. Hagen, ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging, *Bioinformatics* **30(16)**, 2389–2390 (2014).

[24] D. Sage, H. Kirshner, T. Pengo, N. Stuurman, J. Min, S. Manley, and M. Unser, Quantitative evaluation of software packages for single-molecule localization microscopy, *Nature Methods* **12**, 717–724 (2015).

[25] F. Huang, S. L. Schwartz, J. M. Byars, and K. A. Lidke, Simultaneous multiple-emitter fitting for single molecule super-resolution imaging, *Biomedical Optics Express* **2(5)**, 1377–1393 (2011).

[26] S. Holden, S. Uphoff, and A. A. Kapanidis, DAOSTORM: an algorithm for high- density super-resolution microscopy, *Nature Methods* **8**, 279–280 (2011).

[27] T. Takeshima, T. Takahashi, J. Yamashita, Y. Okada, and S. Watanabe, A multi-emitter fitting algorithm for potential live cell super-resolution imaging over a wide range of molecular densities, *Journal of Microscopy* **271**, 266–281 (2018).

[28] H. Manko, Y. Mély, and J. Godet, Advancing spectrally-resolved single molecule localization microscopy with deep learning, *Small* **19**, 2300728 (2023).

[29] M. F. Serag, M. Abadi, H. Al-Zarah, O. Ibrahim, and S. Habuchi, Deep Learning and Single-Molecule Localization Microscopy Reveal Nanoscopic Dynamics of DNA Entanglement Loci, *ACS Nano* **19**, 6236–6249 (2025).

[30] J. Mao and H. He, Deep learning in fluorescence imaging and analysis, *Journal of Intelligent Medicine* **1**, 42–62 (2024).

[31] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, Deep-STORM: super-resolution single-molecule microscopy by deep learning, *Optica* **5**, 458–464 (2018).

[32] A. Speiser, L.-R. Müller, P. Hoess, U. Matti, C. J. Obara, W. R. Legant, A. Kreshuk, J. H. Macke, J. Ries, and S. C. Turaga, Deep learning enables fast and dense single-molecule localization with high accuracy, *Nature Methods* **18**, 1082–1090 (2021).

[33] C. Ma, W. Tan, R. He, and B. Yan, Pretraining a foundation model for generalizable fluorescence microscopy-based image restoration, *Nature Methods* **21**, 1558–1567 (2024).

[34] W. H. Richardson, Bayesian-based iterative method of image restoration, *Journal of the Optical Society of America* **62**, 55–59 (1972).

[35] L. B. Lucy, An iterative technique for the rectification of observed distributions, *Astronomical Journal* **79**, 745–754 (1974).

[36] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 770–778 (2016).

[37] A. L. Maas, A. Y. Hannun, and A. Y. Ng, Rectifier nonlinearities improve neural network acoustic models, *Proceedings of the 30th International Conference on Machine Learning* **28**, 6 (2013).

[38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, (2016), `http://www.deeplearningbook.org`.

[39] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research* **15**, 1929–1958 (2014).

[40] S. Stallinga and B. Rieger, Accuracy of the gaussian point spread function model in 2d localization microscopy, *Optics Express* **18**, 24461 (2010).

[41] G. M. V. de Ven, T. Tuytelaars, and A. S. Tolias, Three types of incremental learning, *Nature Machine Intelligence* **4**, 1185–1197 (2022).

[42] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *Proceedings of the 3rd International Conference on Learning Representations* (2015), arXiv:1412.6980.

[43] A. Dostálová, R. Stárek, D. Vašinka, and M. Ježek, Github repository for *Calibration-free single-frame super-resolving fluorescence microscopy* (2025), https://github.com/RobStarek/CFCNN.