

Enhancing Fluorescence Microscopy Resolution Beyond the Diffraction Limit via Cascaded Up- and Down-Sampling Networks

Haziique Aetesam¹[0000–0003–3297–9949]

Birla Institute of Technology Mesra, Patna Campus, 800014, India
haziique@bitmesra.ac.in

Abstract. Fluorescence microscopy enables high-contrast visualization of subcellular structures through the use of synthetic fluorescent markers, though it is limited by lower spatial resolution and the risk of photo-bleaching. Super-resolution (SR) fluorescence microscopy overcomes the diffraction barrier to resolve finer details, but comes with trade-offs such as decreased temporal resolution, the need for sophisticated equipment, optical sectioning requirements, and constraints on the sample’s exposure tolerance.

In this study, we introduce a deep learning model that employs up-sampling and downsampling to achieve super-resolution. By connecting up-projection and down-projection blocks in a cascaded arrangement, our model maintains the important relationships between the input low-resolution (LR) images and the target high-resolution (HR) labels. To extract dependencies across various depths, feature maps produced at intermediate upsampling stages are combined. Additionally, external self-attention blocks are integrated after these intermediate up-projection blocks to assess pixel-level affinities. Unlike traditional self-attention mechanisms that focus on correlations within a single sample, our approach uses external attention to capture inter-sample correlations across the entire dataset, offering the added benefit of reduced computational demands.

Code Availability: <https://github.com/SRFM-UpDownSamp.git>.

Keywords: Fluorescence microscopy · Super-resolution · Deep learning · Uncertainty quantification · External attention.

1 Introduction

Fluorescence microscopy enables the study of sub-cellular particles at the nanoscale range [6]. Specifically, it can identify the presence, absence, and density of these particles even when they are not directly visible as distinct sub-cellular structures. This capability is achieved by utilizing photosensitive and wavelength-specific fluorophores, which attach to particular sub-cellular targets and emit fluorescence when excited by light of a specific wavelength. However, the spatial resolution achievable with fluorescence microscopy is constrained by the

diffraction limit of visible light. Additionally, both signal-dependent and signal-independent noise [14] further complicate image acquisition, creating a trade-off among allowable sample exposure time, signal-to-noise ratio (SNR), and temporal resolution. Exceeding the sample’s exposure tolerance can result in photobleaching and phototoxic effects [12].

Super-resolution microscopy (SRM) [23] techniques, such as structured illumination microscopy (SIM) [11], stimulated emission depletion (STED) [30], and localization-based approaches like stochastic optical reconstruction microscopy (STORM) [29], have revolutionized imaging by enabling spatial resolution beyond the classical diffraction limit. In SR-SIM, the sample is illuminated with patterned (typically sinusoidal) light, which can be systematically shifted and rotated to generate a series of images. The interaction between the structured illumination and the specimen produces moiré fringes—lower-frequency interference patterns that encode otherwise inaccessible high-frequency information about the sample. In contrast, STED microscopy employs a two-laser system: one laser excites the fluorescent molecules, while a second, the STED beam, selectively depletes fluorescence in a controlled fashion, thereby refining the effective point spread function. Despite their transformative capabilities, these super-resolution modalities demand advanced instrumentation and sophisticated computational algorithms for both image acquisition and reconstruction. They often require the collection of multiple frames and are susceptible to challenges such as phototoxicity and photobleaching, primarily due to the high-intensity laser illumination involved.

2 Related Works

This section examines model-driven (in this paragraph) and data-driven (in the next paragraph) computational techniques that achieve resolution comparable to high-resolution super-resolution microscopy through post-processing. To address limitations in imaging high-density fluorophores, Holden et al. [13] developed a multi-PSF model enabling resolution at densities of 10 molecules/ μm^2 . Meanwhile, Zhu et al. [32] leveraged compressed sensing principles to achieve unprecedented temporal resolutions of 3 seconds per image. In single-molecule localization microscopy (SMLM), dual-image acquisition under identical conditions is standard. Descloux et al. [8] introduced a spatial resolution estimation method using correlation analysis, requiring non-saturated, band-limited signals with robust spatial sampling. For 3D light-sheet microscopy, Chen et al. [4] proposed super-resolution radial fluctuation (SRRF) to optimize Bayesian analysis of blinking and bleaching events, significantly enhancing optical sectioning. A comprehensive review of SMLM methodologies is provided by Sage et al. [22]. Conventional model-driven approaches demand precise *a priori* knowledge of acquisition parameters, including point spread function (PSF) characteristics and photon-sample interaction dynamics. Inaccuracies in parameter estimation inevitably degrade reconstruction quality, highlighting the critical need for robust calibration.

Data-driven methods overcome the limitations of model-based techniques by eliminating the need to estimate acquisition parameters like PSF or emitter locations [17]. Deep-STORM [18] implements a parameter-free convolutional neural network that achieves ultrafast reconstruction under low-SNR conditions and high emitter densities. Similarly, Ouyang et al. [20] leverage conditional generative adversarial networks (cGANs) with a Pix2Pix framework to recover photo-activated localization microscopy (PALM) images. Li et al. [15] integrate Bayesian inference with deep learning to capture fluorophore distributions through spatiotemporal feature extraction, validated on both simulated and real datasets. Pioneering work by Weigert et al. [28] introduced content-aware image restoration (CARE), a U-Net-based framework applicable to noise reduction, isotropic resolution enhancement, and super-resolution. For structured illumination microscopy (SIM), Shah et al. [24] developed a residual encoder-decoder architecture that jointly denoises and reconstructs images. Addressing the scarcity of realistic training data, Christensen et al. [5] created an Open-SIM simulation pipeline to generate synthetic training pairs, subsequently validating their model on experimental SIM data. Wang et al. [27] trained networks on wave-optics-simulated light-field images for volumetric reconstruction, while Chen et al. [2] achieved $2.5\times$ lateral resolution enhancement in time-lapse volumes using residual channel attention network (RCAN) with STED ground truths. In the frequency domain, Qiao et al. [21] proposed deep Fourier channel attention network (DFCAN) and deep Fourier generative adversarial network (DFGAN) architectures yielding SIM-comparable resolution. Speiser et al. [26] introduced deep context dependent (DECODE) neural network—a stacked dual U-Net framework—enabling high-density single-molecule localization. Chen et al. (2023) [3] attained $30\text{ nm}/10\text{ ms}$ spatiotemporal resolution using sub-pixel edge priors and multi-component optimization, eliminating multi-frame requirements. To democratize access, von Chamier et al. [1] launched *ZeroCostDL4Mic*—a Google Colab-integrated library supporting segmentation, denoising, super-resolution, and image translation for fluorescence microscopy data. The key advantages of the above methods include elimination of parameter estimation for PSF/emitter locations, robustness under challenging conditions (low SNR, high density), versatile architectures (GANs, U-Nets, attention networks), simulation-to-real transfer strategies and cloud-based accessibility for non-specialists. However, they are limited in many ways; that include dependence on large, high-quality training data, simulation-to-real domain gap, interpretability and parameter transparency.

While numerous architectural frameworks and data acquisition strategies have been developed to generate super-resolved fluorescence microscopy images from accessible platforms such as widefield and confocal microscopes—achieving resolutions comparable to advanced modalities like SIM, STED, and STORM—this paper specifically highlights the pivotal influence of model architecture within a single-image super-resolution context. In particular, we explore the synergistic integration of deep back-projection networks and attention mechanisms to enhance image reconstruction.

The remainder of this paper is structured as follows. Section 3 introduces the image formation model and outlines the objectives of our study. Section 4 details the proposed methodology. A comprehensive experimental evaluation is presented in Section 5, and concluding remarks are provided in Section 6.

3 Image Formation Model and Objective

The imaging pipeline for a typical fluorescence microscopy setup follows this model [14]:

$$y = \alpha \mathcal{P} \left(\frac{H \otimes x}{\alpha} \right) \downarrow_s + \eta \quad (1)$$

where $x \in \mathbb{R}^{m \times n}$ is the ground truth high-resolution image, $y \in \mathbb{R}^{m' \times n'}$ is the observed low-resolution image, H is the point spread function (PSF) of the imaging system, \otimes denotes convolution operation and \downarrow_s is the downsampling factor s (with $m = s \times m'$, $n = s \times n'$). $\mathcal{P}(\cdot)$ is the signal-dependent Poisson noise component scaled by α and $\eta \sim \mathcal{N}(0, \sigma^2)$ is signal-independent Gaussian noise with variance σ^2 [14]. The parameters α and σ^2 jointly determine the relative contributions of Poisson and Gaussian noise components in the imaging process.

In a supervised neural network parametrized by learnable weights θ , our objective is to estimate a super-resolved image \hat{x} from corrupted, downsampled observations y such that $\hat{x} \simeq x$. This is achieved by training on paired datasets $\{y_i, x_i\}_{i=1}^N$, where x denotes high-resolution (HR) ground truth, y represents low-resolution (LR) input, and \hat{x} is the super-resolved (SR) output. The main contributions of this work are as follows:

- Inspired by the human visual system’s feedback mechanism, our model iteratively refines reconstruction errors through successive upsampling and downsampling layers, enhancing feature refinement.
- Multi-scale external attention integration from different depths selectively amplify salient features while suppressing noise, improving contextual awareness during reconstruction.
- Features from all upsampling blocks are concatenated via skip connections, preserving structural details across network depths.
- We demonstrate that our model super-resolves blurred, noisy widefield images to achieve spatial resolution comparable to structured illumination microscopy (SIM) ground truth.
- The proposed model is trained and tested on a fluorescence microscopy dataset spanning four structurally complex biological targets: clathrin-coated pits (CCP), endoplasmic reticulum (ER), microtubules (MT) and F-actin fibers. This hierarchical evaluation (from the simplest to the most complex structures) rigorously validates generalization capabilities.

4 The Proposed Methodology

Conventional neural network architectures for image super-resolution typically restrict upsampling operations to either the initial or final network stages [9,25].

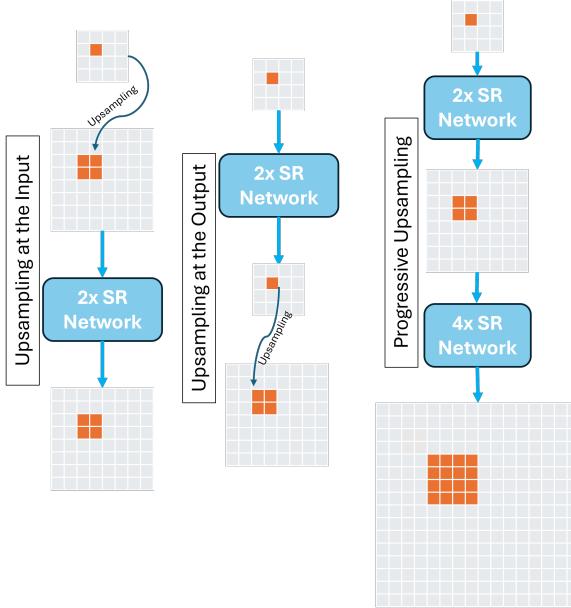


Fig. 1. Different Upsampling strategies in a typical SR network.

Since the model does not participate in the upsampling process, these designs fail to capture the gradual resolution enhancement. While progressive frameworks accommodate multiple scaling factors (e.g., 2 \times , 4 \times , 8 \times) [16], they suffer from conflicting optimization objectives across scales, often degrading performance. Joint multi-scale training further risks scale-specific overfitting without rigorous regularization. Critically, these approaches—illustrated comparatively in Fig. 1—cannot preserve contextual details during the image-to-image translation from low-resolution (LR) inputs to high-resolution (HR) targets. Moreover, artifacts introduced during early upsampling propagate through subsequent stages, particularly in architectures lacking robust residual learning, where sub-optimal feature extraction fundamentally limits final reconstruction quality. To address these limitations, we introduce a feedback mechanism that explicitly mimics the iterative up/down-sampling dynamics observed in fluorescence microscopy, ultra-zoom optical systems, and digital zoom pipelines (e.g., smartphone imaging). This biologically inspired feedback corrects the critical absence of error-correction pathways in vanilla SR networks, enabling the model to progressively refine reconstruction details while preventing error accumulation.

The incorporation of a feedback mechanism introduces a self-correcting capability to the super-resolution (SR) task in fluorescence microscopy. Specifically, back-projection enables iterative refinement of high-frequency textural details, thereby preserving the intrinsic relationship between the low-resolution (LR) input and its high-resolution (HR) counterpart. This cyclical process of successive upsampling and downsampling facilitates the computation and correction of re-

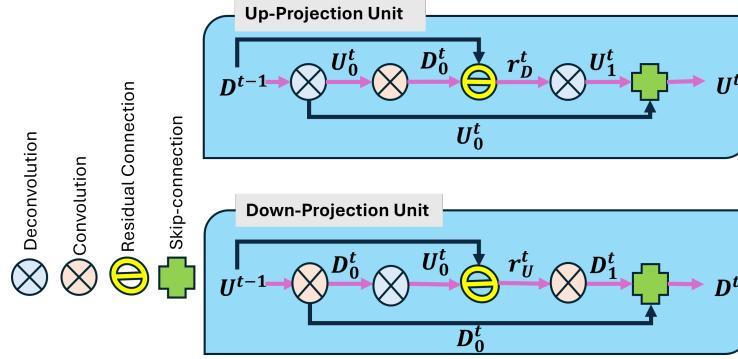


Fig. 2. Up-Projection and Down-Projection Units in the main SR network.

construction errors at multiple stages throughout training, ultimately enhancing image fidelity and structural accuracy.

4.1 Successive Up-Down Sampling

The main building block of the proposed scheme are the projection units (see Fig. 2) which are trained in an end-to-end fashion. More precisely, the mapping from LR \rightarrow HR is called the up-projection while the transformation from HR \rightarrow LR is called the down-projection unit. A typical up-projection unit is calculated as follows:

- **Intermediate Scale Up:** $U_0^t = (D^{t-1} \otimes_{\uparrow} u_0^t) \uparrow_s$
- **Intermediate Scale Down:** $D_0^t = (U_0^t \otimes_{\downarrow} d_0^t) \downarrow_s$
- **Residual:** $r_t^D = D_0^t - D^{t-1}$
- **Residual Scale Up:** $U_1^t = (r_t^D \otimes_{\uparrow} u_1^t) \uparrow_s$
- **Up-Projection Output:** $U^t = U_0^t + U_1^t$

The LR feature map D^{t-1} obtained from the previous down projection unit is convolved with u_0^t to obtain the intermediate HR map U_0^t which is further downsampled by the deconvolutional layer d_0^t to obtain the intermediate LR map D_0^t . The residual r_t^D between the current D_0^t and previous D_0^{t-1} downsampling operation is upsampled with the convolutional layer u_1^t to obtain the residual up-projection map U_1^t . The final up-projection U^t is obtained by adding the intermediate U_0^t and residual U_1^t upscale map.

Similarly, the down-projection unit is designed as follows:

- **Intermediate Scale Down:** $D_0^t = (U^{t-1} \otimes_{\downarrow} d_0^t) \downarrow_s$
- **Intermediate Scale Up:** $U_0^t = (D_0^t \otimes_{\uparrow} u_0^t) \uparrow_s$
- **Residual:** $r_t^U = U_0^t - U^{t-1}$
- **Residual Scale Down:** $D_1^t = (r_t^U \otimes_{\downarrow} d_1^t) \downarrow_s$
- **Down-Projection Output:** $D^t = D_0^t + D_1^t$

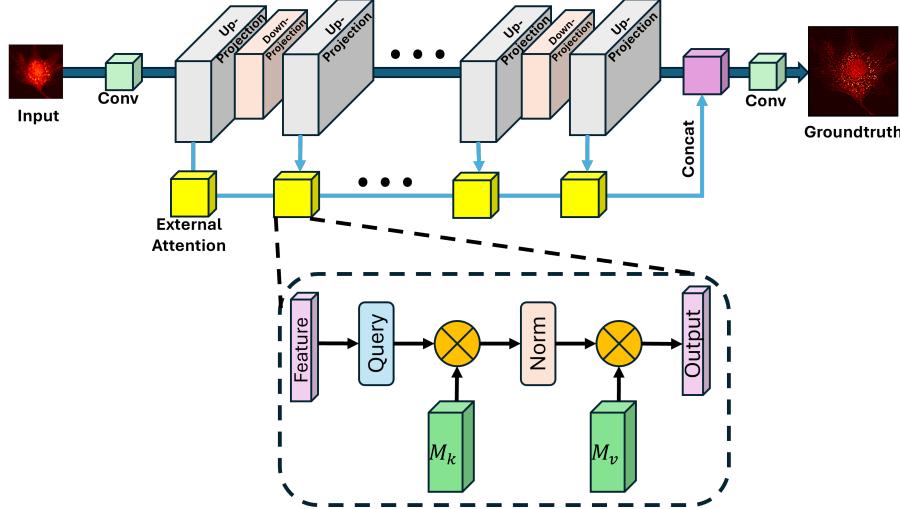


Fig. 3. The complete model architecture with up-projection, down-projection and pluggable external attention blocks.

A set of similar operations as in the up-projection stage is performed in the down-projection stage t with the only exception that transformation takes place from HR \rightarrow LR. The up and downsampling modules are appended to the complete architecture as shown in Fig. 3.

4.2 Pluggable Attention Module

While self-attention revolutionized natural language processing by capturing long-range dependencies in word embeddings, its direct application to vision tasks faces two critical constraints. *Firstly*, for feature space dimensions $N = m \times n$ (total pixels), pixel-level attention requires aggregating features across all N positions. This incurs $\mathcal{O}[dN^2]$ complexity (d = feature dimensions), making it prohibitive for high-resolution images. Consequently, attention must be restricted to patch-level computations rather than individual pixels. *Secondly*, the linear projections generating *key* ($K \in \mathbb{R}^{N \times d'}$), *query* ($Q \in \mathbb{R}^{N \times d'}$), and *value* ($V \in \mathbb{R}^{N \times d}$) matrices operate exclusively within individual samples. This fails to leverage structural correlations across samples—particularly detrimental in low-level vision tasks like super-resolution fluorescence microscopy, where datasets contain semantically related images (e.g., all clathrin-coated pits (CCP) or F-actin samples sharing structural patterns). The standard self-attention output F_{out} is computed as:

$$A = (\alpha)_{i,j} = \text{softmax}(QK^T) \quad (2a)$$

$$F_{\text{out}} = AV \quad (2b)$$

where $A \in \mathbb{R}^{N \times N}$ is the attention map and $F_{\text{out}} \in \mathbb{R}^{R \times d}$ is the output feature space.

To overcome the aforementioned limitations of conventional self-attention modules, we employ a novel mechanism known as External Attention Multiple Layer Perceptron (EAMLP) [10]. In EAMLP, attention maps are generated by first computing correlations between self-query and self-key vectors, followed by weighting these maps with the corresponding self-value vectors. Distinctively, the externally learned key (M_k) and value (M_v) memory units are shared and jointly optimized across all samples in the training set, functioning analogously to learnable parameters. Both memory components are realized using linear layers and serve as multi-sample feature aggregation regularizers, each with a memory size s . Importantly, since s is independent of the input feature map size and significantly smaller than the feature dimension d , the computational complexity of external self-attention reduces to $\mathcal{O}(d \cdot s \cdot m \cdot n) = \mathcal{O}(d \cdot m \cdot n)$. A typical EAMLP implementation is structured as follows:

$$A = \text{Norm}(FM_k^T) \quad (3a)$$

$$F_{\text{out}} = AM_v \quad (3b)$$

where $\{M_k, M_v\} \in \mathbb{R}^{s \times d}$ and $\text{Norm}(\cdot)$ is instance normalization. Softmax in self-attention normalizes the attention map such that $\sum_j \alpha_{i,j} = 1$. However, since attention map is sensitive to the scale of the input feature space, double normalization needs to be applied; as follows:

$$\{\tilde{\alpha}_{i,j}\} = FM_k^T \quad (4a)$$

$$\hat{\alpha} = \exp(\tilde{\alpha}_{i,j}) / \sum_k \exp(\tilde{\alpha}_{k,j}) \quad (4b)$$

$$\alpha_{i,j} = \hat{\alpha}_{i,j} / \sum_k \hat{\alpha}_{i,k} \quad (4c)$$

The block diagram for the external attention module is depicted in Fig. 3.

4.3 Network Architecture

The complete model architecture is illustrated in Fig. 3. To facilitate deep feature integration, upsampled features from the high-resolution (HR) up-projection units are directly fed into the successive down-sampling units. However, these upsampled features are concatenated only after passing through the external attention MLP units, which helps recover high-frequency details that may have been lost during previous down-projection stages. The convolutional and deconvolutional operations within the up- and down-projection units are implemented using standard 2D and transposed 2D convolutional layers, respectively, each employing a kernel size of 3×3 and a stride equal to the up/down sampling factor, denoted as s . The number of filters used throughout these layers is 64. Prior to feeding the input images into the up- and down-projection units, a convolutional layer maps the input image channels to a 256-dimensional feature space. Similarly, after the concatenation layer, a final convolutional layer reduces the feature

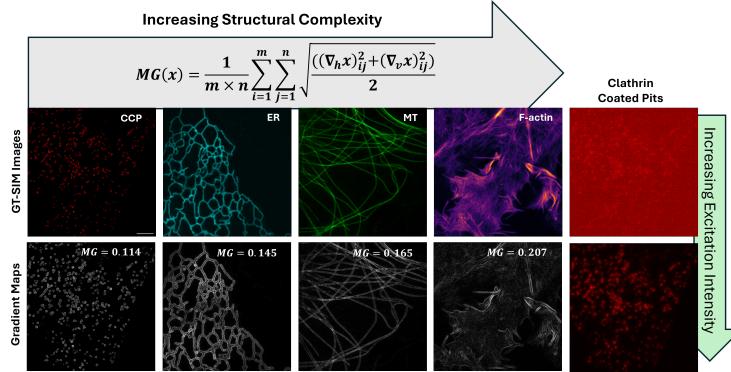


Fig. 4. **Row 1:** Different samples obtained from structured illumination microscopy (SIM) (acting as groundtruth data) in increasing levels of structural complexity: clathrin-coated pits (CCP), endoplasmic reticulum (ER), microtubules (MT) and F-actin fibers. **Row 2:** The corresponding gradient maps where structural complexity is measured in terms of grayscale mean gradient (MG) where $\nabla_h x$ and $\nabla_v x$ are the gradients across horizontal and vertical axes respectively. The representative widefield CCP images for two different excitation intensities serving as input samples during model training (last column).

space back to the original image dimensions. The linear layers M_k and M_v within the attention blocks are initialized with random values. This architecture ensures selective feature fusion guided by attention mechanisms, preventing information loss from iterative down-projection, while maintaining architectural coherence through consistent kernel sizes and strides. The 256-dimensional embedding balances representational capacity with computational efficiency, making the model both effective and practical for super-resolution tasks in fluorescence microscopy.

The model was trained for 100 epochs using the ADAM optimizer with an initial learning rate of 1×10^{-4} , which was reduced by a factor of 10 after 50 epochs. A batch size of 1 was maintained throughout all training sessions. To address residual misalignment between estimated and target images, we employed frequency distribution loss (FDL) [19] as the primary optimization criterion. This loss function operates in the Fourier domain, separately quantifying amplitude discrepancies and phase differences to ensure precise structural alignment.

5 Results and Discussions

5.1 Experimental Setup

All experiments utilized the publicly available BioSR dataset ¹, comprising well-registered LR-HR image pairs acquired through multi-modality structured illumination microscopy (SIM) in two configurations: grazing incidence (GI) and total internal reflectance fluorescence (TIRF). The dataset features nine distinct

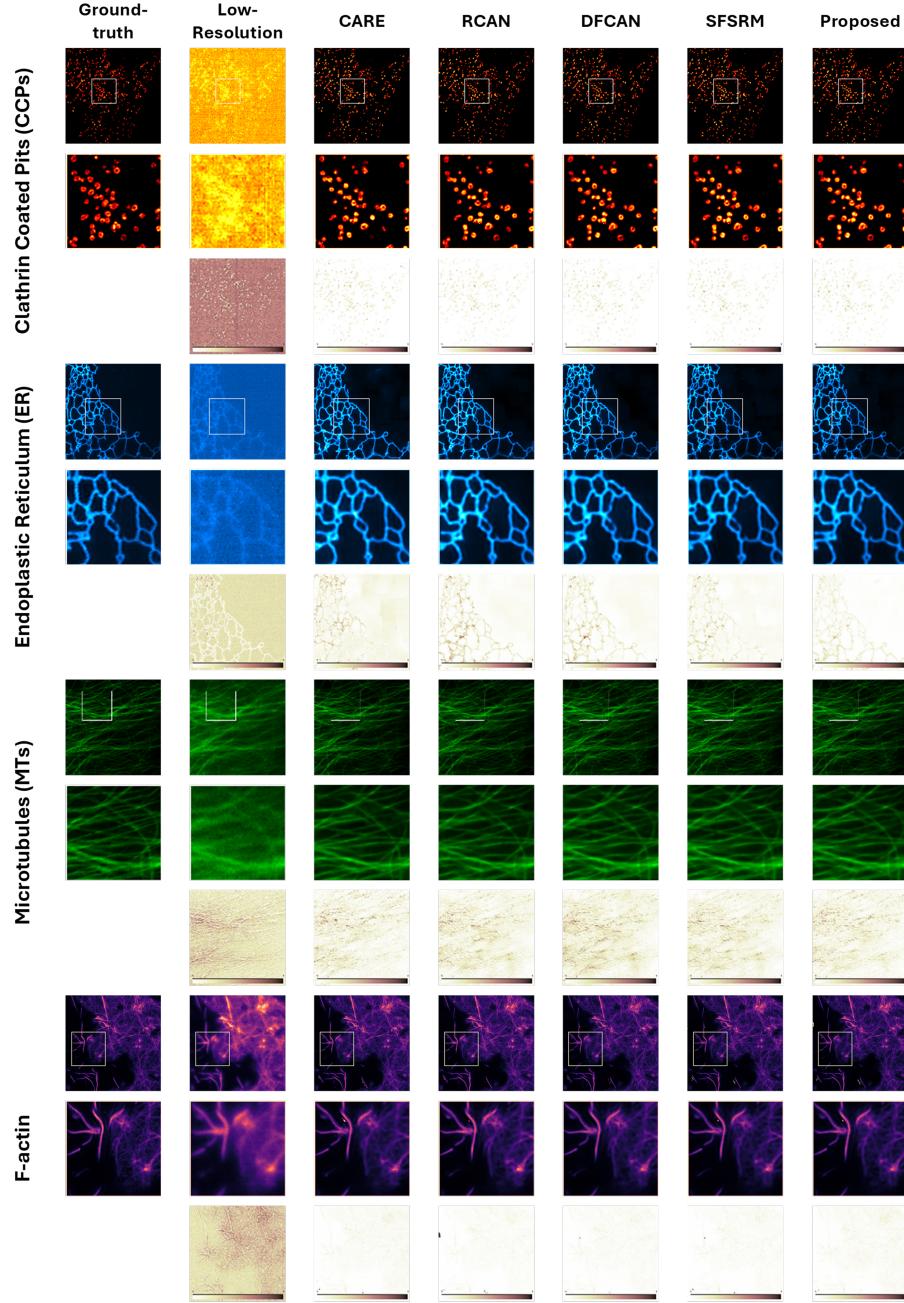


Fig. 5. Rows 1, 4, 7, 10: Super-resolved images inferred from the trained models for different specimen types under low-fluorescence excitation levels using different comparing methods: CARE, RCAN, DFCAN and SFSRM. The corresponding zoomed regions are illustrated in rows (2, 5, 8 and 11) while rows (3, 6, 9 and 12) illustrate the mean absolute error (MAE) between the groundtruth images and the reconstructed SR images.

signal-to-noise ratio (SNR) levels across four biological specimens of increasing structural complexity (see Fig. 4): clathrin-coated pits (CCPs), endoplasmic reticulum (ER), microtubules (MTs), and linear/non-linear F-actin fibers. For each specimen type, 50 fields of view (FOVs) were captured at escalating excitation intensities to simulate SNR variation. Widefield inputs were generated by averaging SIM FOVs, while ground truth targets were acquired at high excitation intensity (120 W cm^{-2} , 10 ms exposure). Of the 50 FOVs per specimen, 35 (30 for training and 5 for validation) were used for model development, while the remaining 15 FOVs were reserved for testing. This resulted in approximately 20,000 paired samples (about 2,200 per SNR level) for each specimen, with input images at a resolution of 128×128 pixels and corresponding target images at 256×256 pixels. For $2\times$ super-resolution evaluation, paired images of size 256×256 and 512×512 were generated from the test FOVs. To ensure consistency and comparability across all samples, percentile-based normalization (PBN) was applied to the ground truth SIM images (x), the widefield input images (y), and the model predictions (\hat{x}) prior to both training and testing.

The results are presented for all sample types at low fluorophore excitation levels². Quantitative evaluation employs peak signal-to-noise ratio (PSNR), normalized root-mean-square error (NRMSE), and multi-scale structural similarity index (MS-SSIM) as key performance indicators. To contextualize our findings, we compare them against several recent state-of-the-art methods specifically designed for fluorescence microscopy, including content-aware image restoration [28], deep neural network-based super-resolution evaluation and development [21], three-dimensional residual channel attention networks for denoising and sharpening [2], and single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging [3].

5.2 Qualitative Results

Visual results across diverse samples are presented in Fig. 5, featuring side-by-side comparisons with state-of-the-art fluorescence microscopy super-resolution methods. While test samples are inherently grayscale, we apply perceptually optimized color maps to enhance structural visibility. Each reconstruction includes corresponding zoomed regions for detailed assessment, demonstrating our model's superior preservation of fine structures through its iterative up/down-sampling architecture. Pixel-level accuracy is quantified via mean absolute error (MAE) maps, which reveal minimal artifacts and exceptional reconstruction fidelity across most samples. However, analysis shows a gradual performance decline with increasing structural complexity—from clathrin-coated pits (CCPs) to F-actin fibers—highlighting the challenge of reconstructing highly intricate biological features.

¹ **Link to the dataset:** <https://figshare.com/articles/dataset/BioSR/>

² **More results at:** <https://aetesamhazique.github.io/SRFM-UpDownSamp/>

Table 1. Metric Values over Different Sample Types

	Noisy	CARE	RCAN	DFCAN	SFSRM	Proposed
CCPs _s	MAE	0.533	0.014	0.012	0.012	0.012
	PSNR	5.35	25.52	26.38	26.35	26.40
	NRMSE	16.644	1.861	1.695	1.694	1.674
	MS-SSIM	0.013	0.894	0.908	0.915	0.918
ER	MAE	0.1687	0.0322	0.0305	0.0277	0.0209
	PSNR	15.04	24.03	25.26	26.58	30.15
	NRMSE	1.354	0.548	0.489	0.420	0.278
	MS-SSIM	0.259	0.868	0.827	0.822	0.889
MTs	MAE	0.131	0.032	0.035	0.044	0.037
	PSNR	16.75	26.81	27.51	25.01	26.22
	NRMSE	0.970	0.305	0.303	0.374	0.325
	MS-SSIM	0.403	0.768	0.808	0.797	0.817
F-actin	MAE	0.129	0.025	0.027	0.021	0.019
	PSNR	16.26	28.12	28.10	27.94	29.17
	NRMSE	1.108	0.274	0.278	0.227	0.223
	MS-SSIM	0.523	0.798	0.806	0.821	0.840

5.3 Quantitative Analysis and Uncertainty Quantification

Table 1 presents a quantitative evaluation of the proposed methodology against other state-of-the-art methods for SR microscopy image reconstruction.

Table 1 provides a quantitative comparison of our proposed approach with leading state-of-the-art methods for super-resolution microscopy image reconstruction, utilizing the evaluation metrics discussed previously. To further validate our results, we employ the Superresolution Quantitative Image Rating and Reporting of Error Locations (SQUIRREL) metric [7]. SQUIRREL assesses the fidelity of super-resolved images by comparing a diffraction-limited reference image to its super-resolution counterpart from the same sample volume. This process involves applying a resolution scaling function (RSF) to the SR image to produce a resolution-scaled version, followed by the computation of pixel-wise absolute differences—termed the resolution scaled error map (RSM)—between the reference and resolution-scaled images. The RSM serves as an artifact map, highlighting local discrepancies and defects. Two summary metrics are then derived: the resolution-scaled error (RSE), which is sensitive to contrast and brightness variations, and the resolution-scaled Pearson’s correlation coefficient (RSP), ranging from -1 to 1, enabling robust cross-modality quality assessment. Figure 6 displays the RSMs for various specimen types, along with their corresponding RSE and RSP values. The scale bar included in the RSM images aids in distinguishing authentic biological features from reconstruction artifacts.

The rolling Fourier-ring correlation (rFRC) [31] overcomes a key limitation of conventional computational imaging techniques, which often fail to account for spatially heterogeneous reconstruction quality. By analyzing local frequency-domain correlations, rFRC quantifies reconstruction uncertainty at the super-resolution scale. This is achieved through a scanning window approach that generates pixel-level resolution maps across the image. Figure 6 presents rFRC maps for various specimen types, alongside corresponding rFRC values and mean resolution measurements at different spatial locations. The accompanying col-

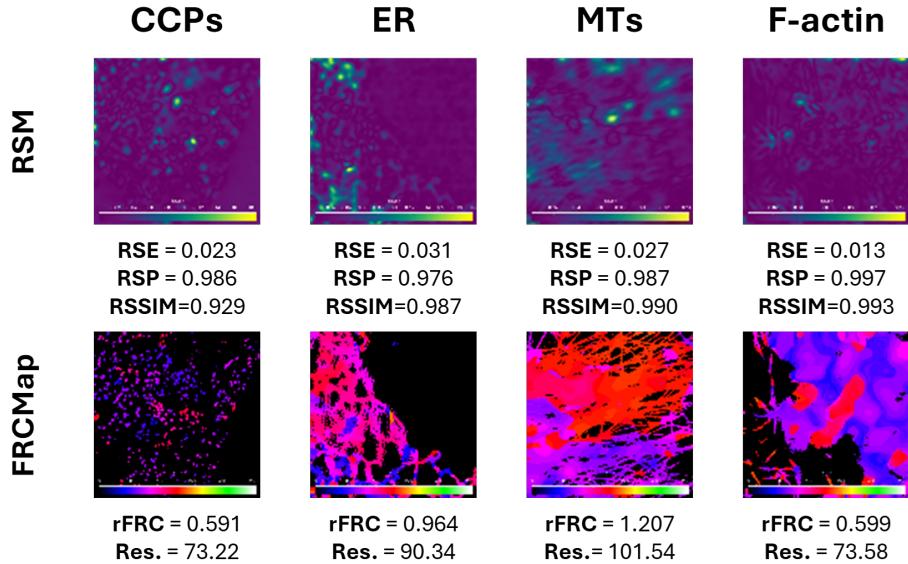


Fig. 6. Resolution Scaled Error Maps (**RSM**) and rolling Fourier Ring Correlation Maps (**rFRC Maps**) for different specimen types under the proposed methodology. Also provided are the corresponding metric values: resolution scaled error (**RSE**), Pearson correlation coefficient (**RSP**) and structural similarity index (**RSSIM**). Uncertainty quantification metrics: **rFRC value** (higher value indicates higher confidence in the reconstructed structures) and mean resolution (**Res.**) obtained for the given input LR image. **Scalebar:** [0, 1] for RSM and [0, 256] for rFRCMap from left to right.

orbar denotes the achievable resolution range in nanometers, spanning [0, 256], thereby facilitating precise spatial assessment of image quality.

6 Conclusion

In this study, we introduced a deep learning framework for super-resolution in fluorescence microscopy, leveraging real widefield images as inputs and supervising the training with ground truth data acquired from SIM microscopy. The architecture's successive up- and downsampling strategy effectively preserves contextual relationships between low-resolution inputs and high-resolution targets. To further enhance reconstruction fidelity, external attention modules are integrated after each up-projection layer, with features from all depths concatenated to capture multi-scale dependencies. Both qualitative and quantitative evaluations across diverse sample types and varying fluorescence excitation levels demonstrate the robustness and effectiveness of our approach. Building on insights from uncertainty quantification, we propose training dedicated models for different fluorophore excitation levels in future work. We anticipate that this direction will clarify the relationship between attainable mean resolution and sample type under varying signal-to-noise ratio conditions.

Acknowledgments. The work carried out in this paper was funded by the BIT Mesra under Seed Grant Scheme with grant number DRIE/SMS/DRIE-10886/2024-25.

References

1. von Chamier, L., Laine, R.F., Jukkala, J., Spahn, C., Krentzel, D., Nehme, E., Lerche, M., Hernández-Pérez, S., Mattila, P.K., Karinou, E., et al.: Democratising deep learning for microscopy with zerocostdl4mic. *Nature communications* **12**(1), 2276 (2021)
2. Chen, J., Sasaki, H., Lai, H., Su, Y., Liu, J., Wu, Y., Zhovmer, A., Combs, C.A., Rey-Suarez, I., Chang, H.Y., et al.: Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nature methods* **18**(6), 678–687 (2021)
3. Chen, R., Tang, X., Zhao, Y., Shen, Z., Zhang, M., Shen, Y., Li, T., Chung, C.H.Y., Zhang, L., Wang, J., et al.: Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging. *Nature Communications* **14**(1), 2854 (2023)
4. Chen, R., Zhao, Y., Li, M., Wang, Y., Zhang, L., Fei, P.: Efficient super-resolution volumetric imaging by radial fluctuation bayesian analysis light-sheet microscopy. *Journal of biophotonics* **13**(8), e201960242 (2020)
5. Christensen, C.N., Ward, E.N., Lu, M., Lio, P., Kaminski, C.F.: Ml-sim: universal reconstruction of structured illumination microscopy images using transfer learning. *Biomedical optics express* **12**(5), 2720–2733 (2021)
6. Combs, C.A., Shroff, H.: Fluorescence microscopy: a concise guide to current imaging methods. *Current protocols in neuroscience* **79**(1), 2–1 (2017)
7. Culley, S., Albrecht, D., Jacobs, C., Pereira, P.M., Leterrier, C., Mercer, J., Henriques, R.: Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nature methods* **15**(4), 263–266 (2018)
8. Descloux, A., Grußmayer, K.S., Radenovic, A.: Parameter-free image resolution estimation based on decorrelation analysis. *Nature methods* **16**(9), 918–924 (2019)
9. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence* **38**(2), 295–307 (2015)
10. Guo, M.H., Liu, Z.N., Mu, T.J., Hu, S.M.: Beyond self-attention: External attention using two linear layers for visual tasks. *IEEE transactions on pattern analysis and machine intelligence* **45**(5), 5436–5447 (2022)
11. Heintzmann, R., Huser, T.: Super-resolution structured illumination microscopy. *Chemical reviews* **117**(23), 13890–13908 (2017)
12. Hoebe, R., Van Oven, C., Gadella Jr, T., Dhonukshe, P., Van Noorden, C., Manders, E.: Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging. *Nature biotechnology* **25**(2), 249–253 (2007)
13. Holden, S.J., Uphoff, S., Kapanidis, A.N.: Daostorm: an algorithm for high-density super-resolution microscopy. *Nature methods* **8**(4), 279–280 (2011)
14. Li, J., Luisier, F., Blu, T.: Pure-let image deconvolution. *IEEE Transactions on Image Processing* **27**(1), 92–105 (2017)
15. Li, Y., Xu, F., Zhang, F., Xu, P., Zhang, M., Fan, M., Li, L., Gao, X., Han, R.: Dlbi: deep learning guided bayesian inference for structure reconstruction of super-resolution fluorescence microscopy. *Bioinformatics* **34**(13), i284–i294 (2018)

16. Mei, Y., Fan, Y., Zhang, Y., Yu, J., Zhou, Y., Liu, D., Fu, Y., Huang, T.S., Shi, H.: Pyramid attention network for image restoration. *International Journal of Computer Vision* **131**(12), 3207–3225 (2023)
17. Möckl, L., Roy, A.R., Moerner, W.: Deep learning in single-molecule microscopy: fundamentals, caveats, and recent developments. *Biomedical optics express* **11**(3), 1633–1661 (2020)
18. Nehme, E., Weiss, L.E., Michaeli, T., Shechtman, Y.: Deep-storm: super-resolution single-molecule microscopy by deep learning. *Optica* **5**(4), 458–464 (2018)
19. Ni, Z., Wu, J., Wang, Z., Yang, W., Wang, H., Ma, L.: Misalignment-robust frequency distribution loss for image transformation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2910–2919 (2024)
20. Ouyang, W., Aristov, A., Lelek, M., Hao, X., Zimmer, C.: Deep learning massively accelerates super-resolution localization microscopy. *Nature biotechnology* **36**(5), 460–468 (2018)
21. Qiao, C., Li, D., Guo, Y., Liu, C., Jiang, T., Dai, Q., Li, D.: Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature methods* **18**(2), 194–202 (2021)
22. Sage, D., Kirshner, H., Pengo, T., Stuurman, N., Min, J., Manley, S., Unser, M.: Quantitative evaluation of software packages for single-molecule localization microscopy. *Nature methods* **12**(8), 717–724 (2015)
23. Schermelleh, L., Ferrand, A., Huser, T., Eggeling, C., Sauer, M., Biehlmaier, O., Drummen, G.P.: Super-resolution microscopy demystified. *Nature cell biology* **21**(1), 72–84 (2019)
24. Shah, Z.H., Müller, M., Wang, T.C., Scheidig, P.M., Schneider, A., Schüttelpelz, M., Huser, T., Schenck, W.: Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images. *Photonics Research* **9**(5), B168–B181 (2021)
25. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1874–1883 (2016)
26. Speiser, A., Müller, L.R., Hoess, P., Matti, U., Obara, C.J., Legant, W.R., Kreshuk, A., Macke, J.H., Ries, J., Turaga, S.C.: Deep learning enables fast and dense single-molecule localization with high accuracy. *Nature methods* **18**(9), 1082–1090 (2021)
27. Wang, Z., Zhu, L., Zhang, H., Li, G., Yi, C., Li, Y., Yang, Y., Ding, Y., Zhen, M., Gao, S., et al.: Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. *Nature methods* **18**(5), 551–556 (2021)
28. Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., et al.: Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods* **15**(12), 1090–1097 (2018)
29. Xu, J., Ma, H., Liu, Y.: Stochastic optical reconstruction microscopy (storm). *Current protocols in cytometry* **81**(1), 12–46 (2017)
30. Xu, Y., Xu, R., Wang, Z., Zhou, Y., Shen, Q., Ji, W., Dang, D., Meng, L., Tang, B.Z.: Recent advances in luminescent materials for super-resolution imaging via stimulated emission depletion nanoscopy. *Chemical Society Reviews* **50**(1), 667–690 (2021)
31. Zhao, W., Huang, X., Yang, J., Qu, L., Qiu, G., Zhao, Y., Wang, X., Su, D., Ding, X., Mao, H., et al.: Quantitatively mapping local quality of super-resolution

- microscopy by rolling fourier ring correlation. *Light: Science & Applications* **12**(1), 298 (2023)
32. Zhu, L., Zhang, W., Elnatan, D., Huang, B.: Faster storm using compressed sensing. *Nature methods* **9**(7), 721–723 (2012)