PhotoniX Life

Open Access
PhotoniX Life (2026)1:7
https://doi.org/10.3724/PXLIFE.2025-0010

Research

Artifact-suppressed and adaptive self-inspired learning denoising for super-resolution fluorescence microscopy

Jingyang Zhu[1], Yuzhen Liu[1], Yuanyuan Huang[1], Xianxin Ye[2], Shiqun Zhao[2], Xiangyan Ding[1], Liangyi Chen[2,3,4], Haoyu Li[1,5,6], Liying Qu[1*] AND Weisong Zhao[1,5,6*]

[1]Innovation Photonics and Imaging Center, School of Instrumentation Science and Engineering, Harbin Institute of Technology, Harbin 150001, China;
[2]State Key Laboratory of Membrane Biology, Beijing Key Laboratory of Cardiometabolic Molecular Medicine, Institute of Molecular Medicine, National Biomedical Imaging Center, School of Future Technology, Peking University, Beijing 100871, China;
[3]PKU-IDG/McGovern Institute for Brain Research, Beijing 100871, China;
[4]Beijing Academy of Artificial Intelligence, Beijing 100084, China;
[5]Key Laboratory of Ultra-Precision Intelligent Instrumentation of Ministry of Industry and Information Technology, Harbin Institute of Technology, Harbin 150001, China;
[6]State Key Laboratory of Space Environment Interaction with Matters, Harbin Institute of Technology, Harbin 150001, China
* Corresponding authors (emails: quliying0123@163.com (L.Q.); weisongzhao@hit.edu.cn (W.Z.))

Received 17 May 2025; accepted 26 January 2026; published online 31 March 2026

Abstract Self-supervised denoising methods have advanced fluorescence microscopy by improving photon efficiency and eliminating the need for paired training data, enabling fast and long-term live-cell imaging. However, most computational approaches still struggle with background artifacts that compromise biological interpretation. These artifacts—often amplified by normalization strategies in deep learning approaches—manifest as background inconsistencies and fabricated structures that distort relative intensity relationships. Conventional background removal procedures, whether applied as pre- or post-processing steps, prove inadequate when dealing with diverse biological samples that exhibit varying signal distributions and complex background characteristics. Here, we introduce adaptive-Self-inspired Noise2Noise (SN2N), an adaptive-SN2N framework that employs risk-aware adaptive normalization and Gaussian-weighted overlap inference to suppress artifacts. Adaptive-SN2N was validated on both structured illumination and spinning-disk confocal-based structured illumination microscopes, demonstrating improvement in photon efficiency and effective suppression of background artifacts. We further show its potential to improve downstream segmentation accuracy, particularly for challenging biological structures with sparse distributions and dynamic processes. We anticipate this advancement will enable more robust and accurate biological interpretations across a wide range of imaging conditions.

Keywords self-supervised learning, image denoising, biological image analysis, deep learning, live-cell imaging

1. Introduction

Fluorescence microscopy has revolutionized our ability to visualize cellular dynamics and structures, but achieving high-resolution imaging over extended periods presents a fundamental challenge. As researchers push toward higher spatial resolution and longer imaging sessions for capturing biological processes, they must minimize light exposure to reduce photobleaching and phototoxicity, inevitably resulting in images with compromised signal-to-noise ratios (SNRs) [1-3]. This trade-off between photon efficiency and image quality becomes particularly critical in live-cell super-resolution microscopy [4], where both spatial detail and temporal continuity are essential for understanding dynamic biological processes.

Self-supervised denoising methods have emerged as breakthrough solutions for fluorescence microscopy challenges [5-10]. These methods, including pioneering works like Noise2Noise [5], Noise2Self [6], and Noise2Void [7], have fundamentally changed the paradigm of microscopy image restoration. Subsequent advancements have further refined these concepts, for example, by incorporating knowledge of noise structure in StructN2V [11], employing blind-spot network architectures to prevent learning trivial solutions [12], or utilizing innovative training strategies like Neighbor2Neighbor [13] that decouple noise instances. These approaches collectively obviate the need for clean ground truth data, opening new avenues for biological imaging analysis. Self-inspired Noise2Noise (SN2N) [10] represents one such approach that leverages internal data generation strategies and incorporates consistency constraints, allowing training using only single noisy images. This methodology has been shown to improve photon efficiency by approximately one to two orders of magnitude [10], contributing to improved imaging capabilities for super-resolution microscopy of biological specimens over extended periods.

Despite these advantages, computational methods [14-17] face a critical challenge with background artifacts that compromise biological interpretation. During preprocessing or the denoising

© The Author(s) 2026. This is an Open Access article distributed under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

process itself, most approaches introduce artifacts [14,15,18-21] that may appear as background inconsistencies, artificial structures, or distorted intensity relationships. In deep learning-based methods, these artifacts are often exacerbated by normalization strategies applied during training. Recent studies have demonstrated that tile-wise or patch-wise feature normalization—commonly used when processing large images in smaller patches—can be a primary cause of artifacts in biological image segmentation and restoration [22,23]. Specifically, when applying local normalization to patches extracted from images with heterogeneous intensity distributions (e.g., large background regions with sparsely distributed signals), the normalization artificially amplifies background noise patterns and creates systematic distortions at patch boundaries [22]. This occurs because patch-wise normalization computes statistics independently for each local region, causing the network to learn inconsistent intensity mappings that fail to preserve global intensity relationships between foreground structures and background [24,25]. The problem is further compounded in low signal-to-noise scenarios typical of fluorescence microscopy, where background pixels with near-zero intensities become artificially stretched during local normalization [26].

To clarify the context of this work, we define "background" primarily as background regions, referring to areas within the field of view that are devoid of discernible fluorescently-labeled biological structures. The signal within these regions is not truly zero but is a composite of several components. It is dominated by stochastic noise, which our method aims to remove, including photon shot noise inherent to the fluorescence emission process and electronic noise from the camera sensor (e.g., readout noise). Additionally, these regions may contain non-stochastic signal components, such as diffuse, slowly varying stray light or out-of-focus fluorescence from planes above or below the focal point. It is important to note that our self-supervised denoising approach, which builds upon the Noise2Noise principle, is specifically designed to suppress pixel-wise stochastic noise by exploiting its statistical independence. The method does not perform deconvolution and therefore treats structured, non-stochastic signals like out-of-focus light as part of the underlying low-frequency signal to be preserved, rather than as noise to be removed. The "background artifacts" we address herein are primarily the erroneous structures and intensity elevations generated by the denoising algorithm's misinterpretation of amplified stochastic noise, rather than the removal of out-of-focus light.

These background artifacts have profound implications for biological interpretation. They compromise structural boundary definition, reducing the precision of cell or structure segmentation [27]; mask faint signals, affecting the accuracy of particle tracking [28]; and introduce pixel intensity uncertainties that undermine reliable quantification of protein expression levels or molecular localizations [29,30]. Unlike random noise, these artificial structures are particularly deceptive because they often mimic biological patterns. For example, in cancer cell imaging, background artifacts might be misidentified as micro-metastases [31]; in neuronal imaging, artifacts could be mistaken for synaptic junctions [32]; and in drug screening, background inconsistencies lead to measurement fluctuations that affect efficacy assessments [33]. Conventional approaches to mitigate these issues typically involve separate background removal steps either before or after processing. However, these additional procedures create significant challenges when working with diverse biological samples that naturally exhibit varying signal distributions and complex background characteristics. Manual parameter adjustment and sample-specific optimization are often required, limiting throughput and reproducibility.

To address these challenges, we developed adaptive-SN2N, an enhanced self-inspired learning framework that builds upon self-inspired Noise2Noise while specifically targeting background artifact reduction. Our approach includes three key components: First, we introduce a risk-aware adaptive normalization strategy that prevents artifacts by customizing normalization to each image's characteristics. Second, we leverage the core self-inspired learning mechanism, combining self-supervised data generation with self-constrained learning, which increases both photon efficiency and denoising effectiveness. Third, we implement a Gaussian-weighted overlap inference mechanism that eliminates stitching artifacts and boundary discontinuities during full-image reconstruction. Together, adaptive-SN2N not only achieves the one to two orders of magnitude photon efficiency improvement of self-supervised methods but also suppresses background artifacts from the source, making it particularly valuable for quantitative biological imaging applications.

We validate our approach on diverse microscopy datasets, including the public biomedical imaging dataset for super-resolution microscopy (BioSR) [34] benchmark and custom super-resolution spinning disk-based structured illumination microscopy (SD-SIM) [35,36] data. Our approach demonstrates improved photon efficiency, background artifact suppression, and enhanced downstream segmentation accuracy. By substantially improving image quality and addressing background limitations, adaptive-SN2N provides a more reliable computational tool for extracting accurate biological information from noisy microscopy images, supporting precise analysis and quantitative discovery in modern bioimaging research.

2. Results

2.1 Principle of adaptive-Self-inspired Noise2Noise

The proposed adaptive-SN2N framework includes three key components: a risk-aware adaptive normalization, a self-inspired learning denoising framework, and a Gaussian-weighted overlap inference mechanism (Figure 1). Together, these components achieve dual benefits of enhanced image quality with improved photon efficiency while suppressing background artifacts, making it valuable for quantitative biological imaging applications.

2.1.1 Risk-aware adaptive normalization

Patch-wise normalization enhances structural definition by maximizing local contrast within individual regions, demonstrating efficacy for biological structures with homogeneous signal distribution such as dense cytoskeletal networks. However, this approach exhibits limitations when applied to images containing extensive background regions or sparsely distributed fluorescent signals. In such instances, patch-wise normalization amplifies low-magnitude noise patterns in background regions while distorting

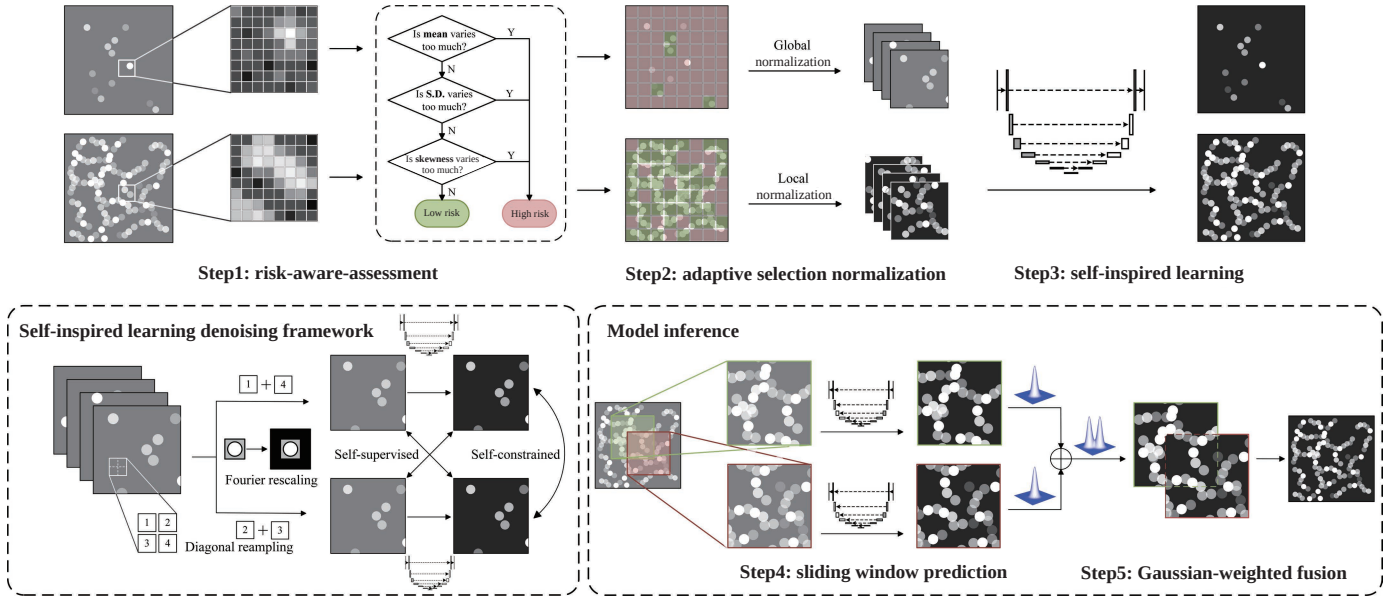


Figure 1 Overview of adaptive-SN2N. Step 1: risk-aware assessment. For each raw image, overlapping patches are extracted. Statistical metrics (mean, standard deviation, and skewness) are calculated for each patch, which is then classified as either "low risk" or "high risk" based on predefined thresholds. Step 2: adaptive selection normalization. The proportion of high-risk patches determines the optimal normalization strategy. If this proportion exceeds a set threshold, image-wise normalization is applied to preserve global intensity relationships and reduce background artifacts. Otherwise, patch-wise normalization is selected to enhance local contrast and structural details. Step 3: self-inspired learning. As detailed in the bottom-left panel ("Self-inspired Learning Denoising Framework"), the SN2N model is trained through a self-inspired process. From a single noisy input image ($H \times W$), diagonal resampling generates two downsampled images ($H/2 \times W/2$). These are subsequently rescaled to the original $H \times W$ dimensions via Fourier interpolation, preserving frequency information. These rescaled twin images, which possess identical content but different noise realizations, serve as both inputs and targets for a U-Net architecture. The network is trained using a self-constrained learning mechanism to enforce consistency between its predictions for these twin inputs. Step 4: sliding window prediction. As depicted in the bottom-right panel ("model inference"), large noisy test images are processed using a sliding window approach, creating overlapping chunks to avoid boundary effects. Each chunk undergoes the same adaptive normalization strategy determined during the training phase before being fed into the trained SN2N model for prediction. Step 5: Gaussian-weighted fusion. Predicted chunks are combined using a Gaussian weighting function, also visualized in the bottom-right panel, which assigns higher confidence to the central regions of each chunk while ensuring smooth transitions in overlapping areas. This weighted fusion mechanism effectively eliminates stitching artifacts and boundary discontinuities, producing a seamless, full-size denoised image with consistent background characteristics.

relative intensity relationships. These transformations manifest as non-uniform background elevation and artifactual structures, potentially inducing biological misinterpretations, including false organelle detection and inaccurate protein expression quantification. On the other hand, image-wise normalization preserves absolute intensity relationships throughout the entire field of view, maintaining background uniformity and preventing selective noise amplification. This methodology provides enhanced quantitative accuracy for sparse signal distributions and preserves intensity gradients essential for rigorous quantitative analyses. However, this approach may diminish local contrast enhancement in regions with densely distributed structures, potentially reducing the details of subtle intracellular features.

To exploit the complementary advantages of both normalization paradigms while circumventing their respective limitations, we developed a risk-aware adaptive normalization strategy that quantitatively evaluates the statistical properties of each image patch to determine its potential artifact risk. The algorithm analyzes three key patch-level metrics—mean (μ_i), standard deviation (S.D.) (σ_i), and skewness (s_i)—and compares them against thresholds within a unified coordinate space. This approach enables adaptation to images with varying overall brightness and contrast levels.

Through this statistical analysis, we identify two primary categories of "high-risk" patches where patch-wise normalization is likely to generate artifacts. The first category is low-signal, low-contrast background, characterized by patches where the mean, standard deviation, and skewness are all significantly lower than their respective global image statistics. The second category is sparse, high-intensity signals, typically found in patches with a low mean intensity but a disproportionately high standard deviation and a strong positive skewness, indicating a few extremely bright pixels on a dark background.

When the proportion of these high-risk patches exceeds a predetermined threshold (T_{ratio}), it signals that the image is prone to normalization artifacts. In such cases, image-wise normalization is applied to the entire field of view, preserving absolute intensity relationships and ensuring background integrity. Conversely, for images with a low proportion of high-risk patches, which predominantly contain structures amenable to local contrast enhancement, patch-wise normalization is retained. This risk-aware adaptive approach optimizes preprocessing for each specific image's morphological characteristics, preventing artifact formation while maximizing structural visualization. A detailed

description of the specific threshold implementation is provided in Sect. 3.1.2.

2.1.2 Self-inspired learning denoising framework

Our denoising framework builds upon SN2N, which combines self-supervised data generation with self-constrained learning to enable denoising without paired ground truth data.

SN2N leverages the spatial redundancy inherent in super-resolution microscopy images to create its own training pairs from single noisy images. Through a diagonal resampling strategy, every 2 × 2 adjacent pixels are treated as one unit, with diagonal pixels averaged to create two distinct subimages (Figure 1). Because each pixel is independently sampled and the PSF has spatial symmetry, these subimages share identical structural content but contain different noise realizations—a configuration for Noise2Noise training. After diagonal resampling produces two smaller subimages, Fourier interpolation is applied to rescale them back to the original size. Unlike conventional spatial methods that introduce artificially smooth transitions between pixels, Fourier interpolation leverages the fact that the optical transfer function of super-resolution microscopes has only finite support. By padding in the frequency domain and back-transforming, this approach adds no new information but increases spatial sampling in the spatial domain. Through this combination of diagonal resampling and Fourier interpolation (Figure 1), SN2N generates twin images that share identical content but have different noise realizations. This data generation strategy overcomes the traditional requirement for multiple acquisitions of the same field of view.

The self-constrained learning process further enhances denoising performance by enforcing consistency between predictions from twin noisy inputs. This consistency requirement is mathematically formalized through a specialized loss function:

$$L = \|\tilde{x}_1 - x_2\|_1 + \|\tilde{x}_2 - x_1\|_1 + \lambda\|\tilde{x}_1 - \tilde{x}_2\|_1, \quad (1)$$

where x_1 and x_2 represent the two generated noisy subimages sharing an identical underlying signal but containing different noise realizations. These images are independently processed through the neural network, producing outputs $\tilde{x}_1$ and $\tilde{x}_2$, respectively. The first two terms represent the conventional Noise2Noise loss, where each network output ($\tilde{x}_1$ or $\tilde{x}_2$) is compared against the other noisy image (x_2 or x_1) as its target. The third term introduces a self-constraining component with weight λ, which enforces consistency between the two network predictions. Through this constraint, the network effectively learns to extract the common underlying structure while disregarding input-specific noise variations. This mechanism particularly stabilizes low-frequency components and substantially improves the model's denoising capability, especially in scenarios with limited training data. By integrating this self-inspired learning mechanism with a self-supervised data generation strategy, SN2N achieves effective denoising with significant photon efficiency improvement.

2.1.3 Gaussian-weighted overlap inference mechanism

We further implemented a Gaussian-weighted overlap inference mechanism to address boundary discontinuities inherent to patch-based processing in deep neural networks. In standard implementations of SN2N and similar architectures, multiple factors contribute to visible artifacts at patch boundaries during reconstruction: (1) statistical inconsistencies between adjacent patches due to independent normalization, (2) intensity distribution shifts between training and inference stages, and (3) the limited receptive field at patch edges to a lesser extent. When patches are naively stitched together, these inconsistencies manifest as abrupt transitions, artificial edges, or intensity discontinuities that can compromise biological interpretation by creating false structures or breaking continuous cellular components.

Our approach employs a sliding window methodology with substantial inter-block overlap (50% in each dimension), creating redundant predictions for each patch. Rather than simple averaging of these multiple predictions, we implement a spatially-varying confidence weighting scheme using a 2D Gaussian function, which assigns maximum confidence to predictions at block centers while reducing confidence toward block peripheries. The final reconstruction is obtained through weighted accumulation of all overlapping predictions, effectively eliminating stitching artifacts and preserving structural continuity throughout the entire field of view. This capability is particularly critical for accurate morphological assessment of complex biological structures with branching topologies and organelle morphology.

2.2 Ablation study and comparative analysis

To systematically evaluate the contributions of its key components, we conducted a comprehensive ablation study of our proposed adaptive-SN2N framework on the BioSR dataset and the fixed-cell SD-SIM data. This study was designed to quantitatively compare our adaptive normalization strategy against standard methods by assessing the performance of four distinct model configurations:

(1) Baseline (patch-wise): this model employs the core self-inspired learning framework (SN2N) with the standard patch-wise normalization.

(2) I-SN2N (image-wise): this variant replaces the patch-wise normalization in the baseline with an image-wise normalization strategy.

(3) Adaptive-SN2N (w/o overlap): this configuration incorporates our proposed risk-aware adaptive normalization strategy but omits the Gaussian-weighted overlap inference mechanism to isolate the sole impact of the adaptive normalization.

(4) Adaptive-SN2N (w/ overlap): this is our complete proposed framework, integrating both the adaptive normalization and the Gaussian-weighted overlap inference.

By comparing these configurations, we dissected the individual and combined effects of our contributions. The results, summarized in Table 1, were evaluated using peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), mean squared error (MSE), and a dedicated artifact quantification metric, the background variance ratio (BVR), defined in Eq. (8) in the methods.

The results presented in Table 1 reveal a fundamental trade-off between normalization strategies and validate the necessity of an adaptive approach. For all sparsely distributed structures (BioSR microtubules (MTs), clathrin-coated pits (CCPs), endoplasmic reticulum (ER), and all fixed-cell datasets), the baseline (patch-

Table 1 Ablation study and quantitative comparison of normalization strategies

Dataset	Organelle	Metric	Baseline (patch-wise)	I-SN2N (image-wise)	Adaptive-SN2N (w/o overlap)	Adaptive-SN2N (w/ overlap)
BioSR	MTs	MSE	552.041,1	48.478,3	48.478,3	41.906,2
		PSNR	20.711,1	29.427,8	29.427,8	31.908,0
		SSIM	0.583,9	0.907,2	0.907,2	0.921,4
		BVR	1.564,8	1.071,9	1.071,9	0.884,7
	CCPs	MSE	900.817,4	13.835,3	13.835,3	11.098,4
		PSNR	20.006,8	33.792,7	33.792,7	37.678,2
		SSIM	0.380,4	0.883,2	0.883,2	0.915,0
		BVR	4.739,0	0.867,5	0.867,5	0.836,2
	ER	MSE	1,010.890,8	136.278,7	136.278,7	76.281,6
		PSNR	22.597,6	27.722,0	27.722,0	29.306,6
		SSIM	0.610,8	0.757,2	0.757,2	0.881,6
		BVR	1.735,6	1.368,8	1.368,8	1.261,3
	F-actin	MSE	65.512,4	360.361,3	65.512,4	64.189,2
		PSNR	37.678,2	34.967,6	37.678,2	38.957,1
		SSIM	0.915,0	0.885,2	0.915,0	0.921,5
		BVR	1.234,2	1.174,4	1.234,2	1.204,8
Fixed-cell	Lys	MSE	2,417.664,3	393.525,9	393.525,9	332.606,7
		PSNR	19.372,3	35.274,9	35.274,9	37.798,7
		SSIM	0.41	0.901,8	0.901,8	0.923,0
		BVR	3.876,0	1.192,4	1.192,4	1.037,5
	Mito	MSE	4,028.7	98.274,9	98.274,9	62.139,2
		PSNR	21.503,1	35.820,1	35.820,1	36.923,2
		SSIM	0.504,3	0.862,1	0.862,1	0.890,8
		BVR	1.858,9	1.249,5	1.249,5	0.921,6
	MTs	MSE	3,453.805,3	118.798,0	118.798,0	96.301,3
		PSNR	16.719,7	27.309,8	27.309,8	30.145,5
		SSIM	0.576,4	0.865,9	0.865,9	0.878,6
		BVR	1.389,5	1.163,8	1.163,8	0.821,4

wise) method performed poorly, exhibiting high MSE and BVR values (e.g., 4.739,0 for CCPs), which quantitatively confirms the severe background artifacts it introduces. In stark contrast, I-SN2N (image-wise) dramatically improved all metrics for these samples, demonstrating its superiority in preserving background integrity. However, for the dense filamentous actin (F-actin) network, this trend was reversed: the baseline method significantly outperformed I-SN2N on key structural metrics (PSNR 37.68 vs. 34.97; SSIM 0.915,0 vs. 0.885,2), as its patch-wise approach excelled at enhancing local contrast. These quantitative trends and the comparative performance of different strategies are further visualized in the bar charts in Figure 2g–i.

This highlights the core dilemma: no single fixed normalization strategy is optimal for all types of biological images. Our risk-aware adaptive normalization strategy effectively resolves this conflict. The performance of the adaptive-SN2N (w/o overlap) configuration provides direct evidence of its intelligent decision-making. As this model implements a binary choice between the two fixed strategies, its performance metrics are, by design, identical to the superior method for each specific case. For all sparse samples, its metrics match those of I-SN2N; for the dense F-actin, its metrics match those of the Baseline. This differential performance across diverse biological structures provides strong empirical justification for our adaptive normalization strategy, which intelligently selects the optimal approach based on image content characteristics—employing image-wise normalization for sparsely distributed structures (CCPs, MTs, and ER) to maintain background integrity, while selecting patch-wise normalization for densely distributed patterns (F-actin) to maximize structural detail preservation.

Furthermore, the Gaussian-weighted overlap inference provides a significant and synergistic performance boost. By comparing the adaptive-SN2N (w/o overlap) configuration with our final adaptive-SN2N (w/ overlap) model, a consistent and substantial improvement is observed across nearly all metrics and datasets. For instance, PSNR for BioSR MTs increased from 29.43 to 31.91, and SSIM for fixed-cell lysosomes (Lys) improved from 0.901,8 to 0.923,0. This demonstrates the critical role of the weighted inference mechanism in eliminating stitching artifacts and enhancing overall image fidelity, validating it as an

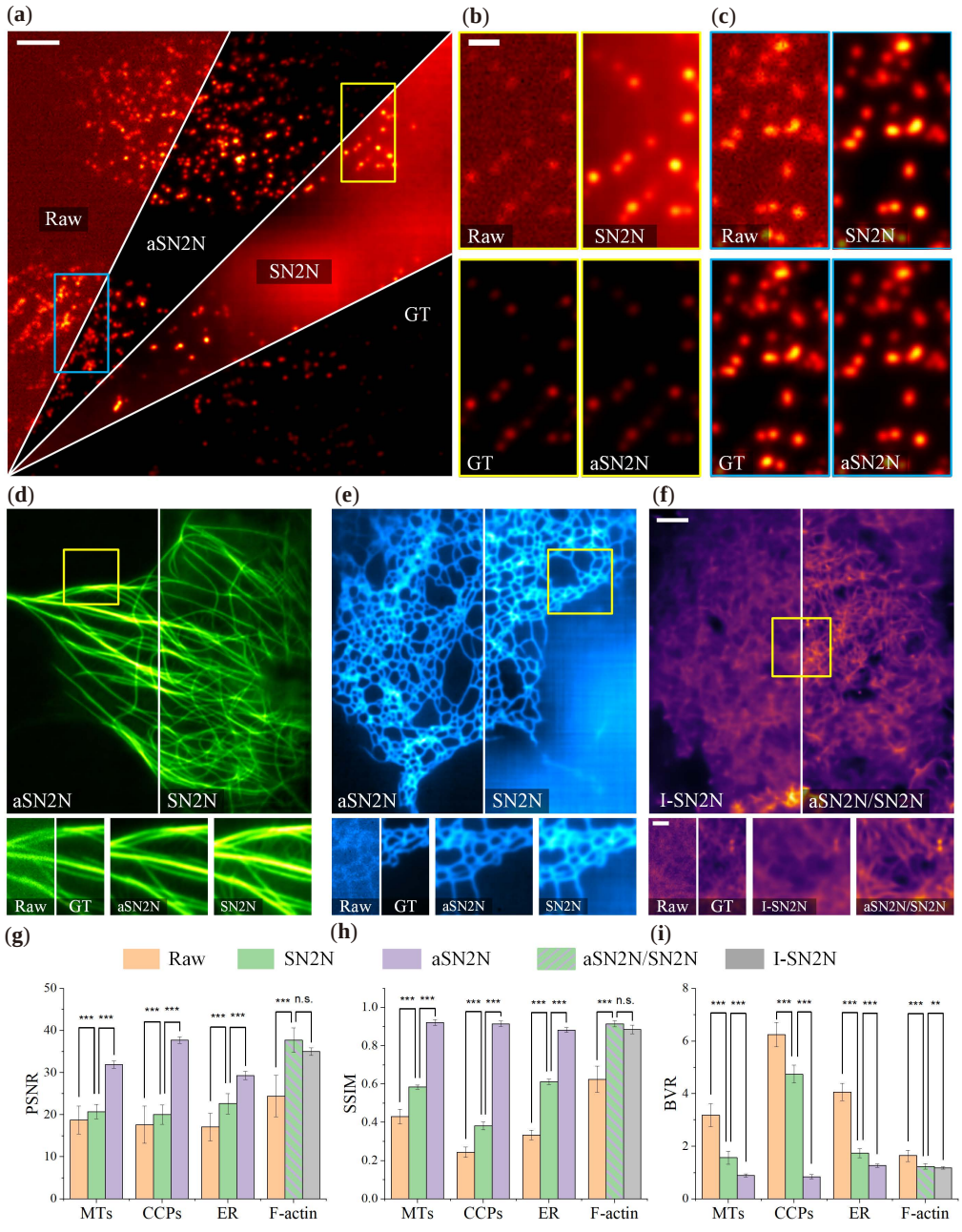


Figure 2 Performance comparison of SN2N and aSN2N on the BioSR dataset. **a.** Full field-of-view comparison for CCPs, showing raw input (raw, left), aSN2N output (top left), SN2N output (top right), and ground truth (GT, right); **b–c.** ROIs from **b** the yellow and **c** blue boxes in **a**, detailing Raw, SN2N output, GT, and aSN2N output for CCPs; **d.** comparison of aSN2N (left) and SN2N (right) for MTs. Inset ROIs show Raw, GT, aSN2N, and SN2N; **e.** comparison of aSN2N (left) and SN2N (right) for ER. Inset ROIs show Raw, GT, aSN2N, and SN2N; **f.** denoising of F-actin. Left: I-SN2N (image-wise normalized SN2N). Right: aSN2N/SN2N, representing the outcome where both aSN2N and SN2N utilize patch-wise normalization for this specific dataset. Inset ROIs compare Raw, GT, I-SN2N, and aSN2N/SN2N. **g–i.** Quantitative evaluation by **g** PSNR, **h** SSIM, and **i** BVR. For MTs, CCPs, and ER, metrics compare aSN2N against SN2N. For F-actin, metrics compare aSN2N/SN2N against I-SN2N. The bar height represents the mean metric, and the error bars indicate the standard error of the mean (SEM) across all test images. Statistical significance between selected pairs was assessed using a two-sided Wilcoxon signed-rank test. Brackets indicate the pairs being compared. Asterisks denote significance levels: $^{**}p<0.01$, $^{***}p<0.001$. "n.s." indicates no significant difference ($p \geqslant 0.05$). Scale bars: 3 μm (**a**; **d–f** main images); 1 μm (**b** and **c**; **d–f** ROIs).

indispensable component of our framework.

Notably, an important nuance is observed in the F-actin results: while our full model achieves the highest PSNR and SSIM, its BVR is slightly elevated compared to that of I-SN2N (1.204,8 vs. 1.174,4). This outcome has a clear mechanistic origin rooted in the inherent properties of the normalization methods themselves. The adaptive framework correctly classifies the predominantly dense F-actin images as low-risk, thus applying patch-wise normalization to maximize local contrast and structural detail. While this is the optimal choice for overall image fidelity, it has a

predictable side effect. For the subset of images within the F-actin dataset that contain minor background regions, the patch-wise process inevitably amplifies the noise variance within those specific areas. In contrast, I-SN2N, by uniformly applying global normalization, circumvents this local amplification, thus achieving a marginally lower average BVR score across the dataset.

2.3 Performance evaluation on biomedical imaging dataset for super-resolution microscopy benchmark dataset

Following the quantitative ablation analysis, we visually evaluated the reconstruction quality on the BioSR dataset (Figure 2).

For CCPs (Figure 2a), the original low-SNR acquisitions (Raw) showed notable noise that limited structural detail visibility. Denoising with SN2N reduced noise but introduced diffuse background elevations, particularly apparent in regions with sparse CCP distribution and intervening cytoplasmic areas (Figure 2a, global view and Figure 2b, magnified region of interest (ROI) indicated by a yellow border). These background alterations created a non-uniform appearance that deviated from the dark cytoplasmic background observed in the ground truth (GT) reference. In contrast, aSN2N suppressed noise while preserving background characteristics, maintaining consistent dark cytoplasmic regions across the field of view, with results more similar to the GT (Figure 2a and b). Notably, in regions with dense CCP structures, the background performance of SN2N exhibited an enhancement, converging towards that of aSN2N (Figure 2c, ROI indicated by a blue border). This observation indicates that normalization-induced artifacts primarily occur in regions with sparse signal distribution or extensive background areas. The mechanism underlying these background alterations stems from the patch-wise normalization's effect on local intensity statistics. By normalizing each patch independently to the [0, 1] range, the algorithm removes global intensity context and relative gradients, while amplifying low-amplitude noise to occupy the same dynamic range as biological signals. This transformation leads the neural network to interpret background noise patterns as structural features. During inference, these interpretations manifest as background elevations and artificial structures that affect quantitative analyses.

For the ROI region of MT images (Figure 2d), which we classify as containing sparse high-intensity signals, SN2N produced notable directional fibrillar artifacts in the inter-microtubule spaces. The origin of these biologically-mimicking artifacts is twofold. First, patch-wise normalization dramatically amplifies the noise in the low-signal regions between bright filaments. Second, the neural network, having been trained on MT data, develops a strong "learned prior" for linear structures. When confronted with the amplified, unstructured noise, the network attempts to interpret it based on this prior, effectively 'hallucinating' fibrillar patterns from random noise. This phenomenon creates structures that could be misinterpreted as genuine cytoskeletal components, posing significant challenges for biological analyses.

For ER [37] images (Figure 2e), differences between normalization strategies were less apparent in regions covered by continuous ER labeling. However, at image peripheries and in areas with sparse ER distribution, SN2N exhibited inaccurate background recovery with elevated intensity and textural artifacts, whereas aSN2N robustly restored the deep background. The appearance of artifacts induced by SN2N was highly dependent on the intrinsic noise characteristics of the source data. In the ER image, SN2N processing resulted in prominent grid-like artifacts. Fourier analysis of the raw data revealed that the ER image contains significant structured noise, evidenced by strong horizontal and vertical components in its frequency spectrum. The patch-wise normalization employed by standard SN2N exacerbates this issue. By independently scaling each background patch, it amplifies the faint, grid-aligned fixed-pattern noise to a contrast level comparable to that of the true ER signal. This creates an ambiguous input, confusing the network, which then struggles to distinguish the structured noise from genuine biological signals, ultimately preserving and enhancing this noise pattern.

As described in Sect. 2.2, aSN2N correctly identified the ER image as high-risk and applied image-wise normalization. This global approach preserves the relative intensity separation between the bright ER signal and the faint background noise. Consequently, the network is presented with an unambiguous scenario: the true signal has high intensity, while the structured noise remains at a very low intensity. This allows the network to effectively learn to apply the self-consistency constraint to the high-intensity signal, while mapping the low-intensity background—including the structured noise—to a uniform, clean state. Thus, aSN2N succeeds not by fundamentally altering the SN2N principle to handle structured noise, but by providing the necessary global context for the network to correctly differentiate signal from noise, thereby preventing the artifacts from forming.

Interestingly, for F-actin networks (Figure 2f), characterized by dense, relatively homogeneous signal distribution, our aSN2N selected the same patch-wise normalization approach as SN2N, resulting in identical processing outcomes for this specific biological structure. This selection contrasts with the previous structures, where aSN2N opted for image-wise normalization. In our comparative analysis, while image-wise normalization (I-SN2N) maintained better background uniformity for F-actin, it compromised the preservation of fine structural details. In contrast, patch-wise normalization produced enhanced local contrast with F-actin filaments showing higher brightness and sharper boundaries. This confirms that for dense, uniform structures, the adaptive algorithm correctly identified the advantage of local normalization, matching the optimal visual quality required for structural analysis.

2.4 Validation on fixed-cell spinning disk-based structured illumination microscopy data

We further compared SN2N versus aSN2N under a commercial SD-SIM system (Olympus SpinSR10 with an sCMOS camera; methods). Our analysis encompassed three subcellular structures: mitochondria (Mito) labeled with mGold-TOM20, Lys labeled with LAMP1-EGFP, and microtubules labeled with Qdot525. For Mito (Figure 3a), the original acquisition (raw, Figure 3a, left) exhibited substantial noise that obscured structural details, as evident in the magnified ROI (Figure 3b, raw). Consistent with our previous observations, SN2N (Figure 3a, bottom) reduced

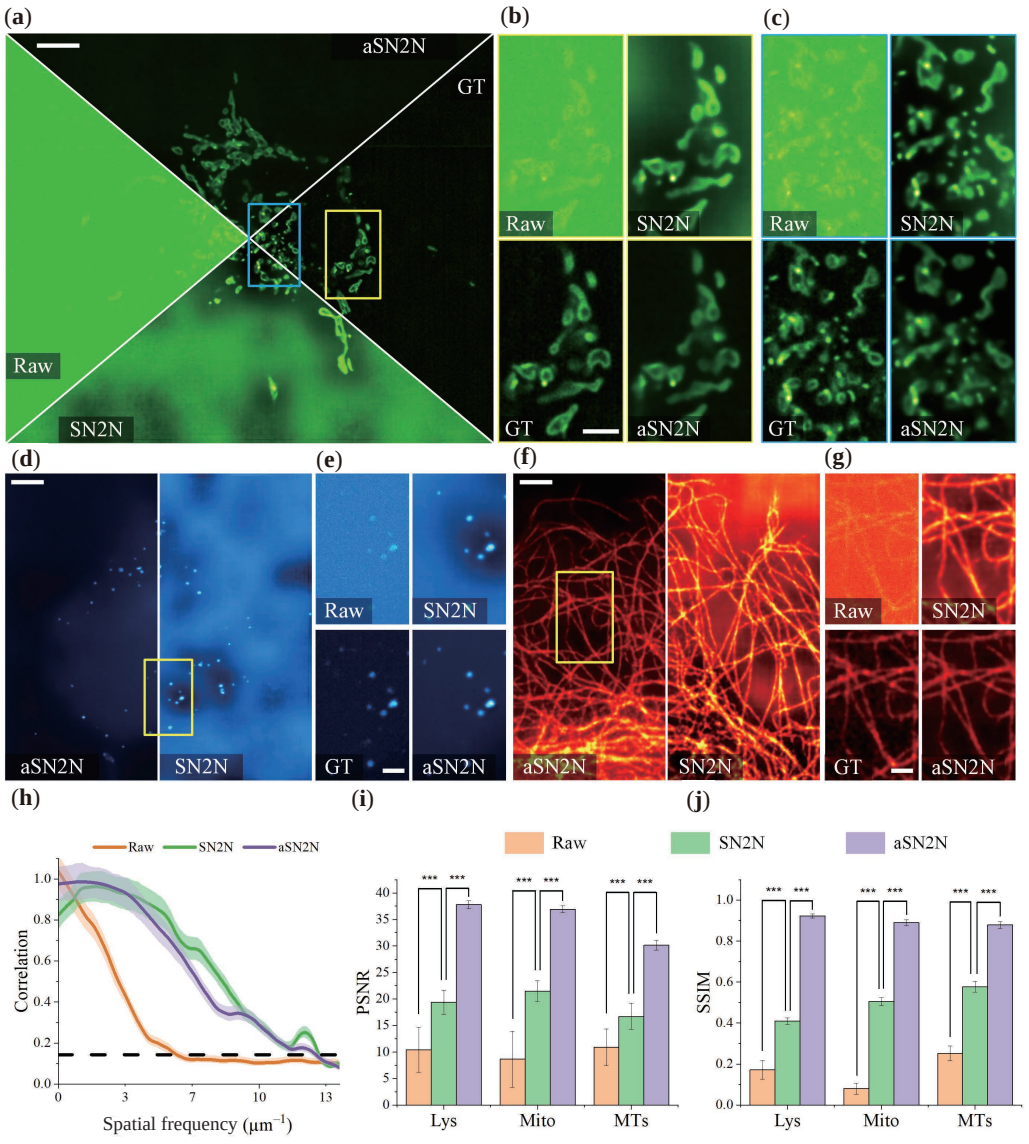


Figure 3 Adaptive-SN2N improves SD-SIM image denoising in fixed COS-7 cells by suppressing background artifacts. **a–g**. Imaging validation of different cellular structures in COS-7 cells: **a** representative wide-field (WF) images of Mito (mGold-TOM20 labeled), comparing the raw input (raw, left), SN2N output (bottom), adaptive-SN2N (aSN2N) output (top), and ground truth (GT, right); **b–c** enlarged regions from the **b** yellow and **c** blue boxes indicated in **a**, showing Raw, SN2N output, GT, and aSN2N output for detailed comparison; **d** denoising comparison for lysosomes (LAMP1-EGFP labeled), displaying aSN2N output (left) and SN2N output (right); **e** enlarged ROI from the yellow box in **d**, comparing Raw, SN2N output, GT, and aSN2N output; **f** denoising comparison for microtubules (Qdot525 labeled), displaying aSN2N output (left) and SN2N output (right); **g** enlarged ROI from the yellow box in **f**, comparing Raw, SN2N output, GT, and aSN2N output. **h**. FRC curves for raw data (orange), SN2N output (green), and aSN2N output (purple). The solid lines represent the mean correlation calculated across multiple test images (e.g., n=10), with the shaded areas denoting the SEM. The dashed black line indicates the 1/7 correlation threshold. **i–j**. Quantitative comparison of **i** PSNR and **j** SSIM for Lys, Mito, and MTs. Comparisons are presented for raw data (orange), SN2N output (green), and aSN2N output (purple). The bar height represents the mean metric, and the error bars indicate the SEM across all test images. Statistical significance between selected pairs was assessed using a two-sided Wilcoxon signed-rank test. Brackets indicate the pairs being compared. Asterisks denote significance levels: ***p<0.001. Scale bars: 3 μm (**a**, **d**, **f**); 1 μm (**b**, **c**, **e**, **g**).

random noise but introduced background elevations, particularly in regions with sparse mitochondrial distribution. In contrast, aSN2N (Figure 3a, top) suppressed noise while maintaining a uniform background across the entire field, producing results more consistent with the GT (Figure 3a, right) reference. Even in denser mitochondrial regions (Figure 3c) where SN2N performed better, adaptive-SN2N still provided improved background fidelity.

Similar patterns emerged for lysosomal structures (Figure 3d), where SN2N generated diffuse background artifacts, while aSN2N preserved background clarity and rendered well-defined lysosomal signals. In microtubule network imaging (Figure 3f), SN2N created uneven background intensity, particularly between filaments, whereas aSN2N effectively suppressed these artifacts while preserving microtubule continuity and structural definition

with maintained background integrity. These results consistently demonstrated aSN2N's superior performance across diverse subcellular architectures.

Quantitative assessments further validated the aSN2N approach. Resolution was evaluated using Fourier ring correlation (FRC) [38,39] analysis (Figure 3h) of the lysosomal (LAMP1-EGFP) dataset (Figure 3d) with a standardized 1/7 correlation threshold. As anticipated, the FRC curve for raw acquisitions declined most rapidly, indicating inferior resolution. While both SN2N and aSN2N improved resolution over the raw input and reached comparable theoretical limits, their FRC profiles revealed a critical difference in stability and fidelity.

Notably, the FRC curve for standard SN2N (green line) exhibits a physically anomalous trend: it initially rises at low spatial frequencies before decaying. This is not a random fluctuation but a direct consequence of the large-scale background elevations introduced by patch-wise normalization, a core issue addressed in this paper. These low-frequency artifacts disrupt the correlation between the two independent image halves used for FRC calculation, artificially depressing the FRC value near the origin and causing the abnormal rise. This phenomenon serves as a quantitative indicator of the background corruption caused by standard SN2N.

In contrast, the FRC curve for aSN2N (purple line) displays a stable, monotonically decreasing profile, which is theoretically expected. This demonstrates that by effectively suppressing background artifacts, aSN2N not only enhances visual quality but also ensures a more faithful structural representation across all spatial frequencies, leading to more reliable and accurate quantitative measurements such as FRC-based resolution assessment.

2.5 Adaptive-Self-inspired Noise2Noise enables high-fidelity multicolor live-cell super-resolution imaging

Live-cell multicolor super-resolution imaging is critical for understanding dynamic cellular processes, but poses unique challenges for image denoising. Background artifacts are particularly problematic in such contexts, as they can create false structures that mimic biological features, leading to erroneous interpretation of organelle dynamics and compromising downstream quantitative analyses such as segmentation. To evaluate aSN2N's effectiveness in addressing these specific challenges, we conducted dual-color SD-SIM time-series imaging on live COS-7 cells co-expressing Tom20-mCherry (mitochondria, shown in green) and Sec61β-EGFP (ER, shown in magenta) (Figure 4).

The original SD-SIM image (Figure 4a, top) at t = 0 s exhibits significant noise that obscures fine organelle structures. After processing with SN2N (Figure 4a, bottom), some noise was suppressed, but diffuse artifacts appeared in both channels, particularly in background regions. These artifacts reduced image contrast and blurred organelle boundaries. In the mitochondrial channel, background elevation led to severe segmentation errors. As shown in Figure 4b (middle row, SN2N segmentation panel), Otsu's thresholding misidentified background regions as foreground, compromising quantification accuracy.

In contrast, aSN2N (Figure 4a, middle) effectively eliminated noise while maintaining background uniformity and low signal levels, rendering well-defined organelle structures. In dual-color merged images, aSN2N enabled a more accurate representation of mitochondria-ER spatial relationships. The clean background provided by aSN2N allowed accurate mitochondrial segmentation, avoiding the background mis-segmentation issues present in standard SN2N results (Figure 4b, bottom row). For ER structures, which are characterized by their complex morphology and low contrast, aSN2N's artifact suppression is critical for accurate downstream analysis. To investigate this, we performed skeletonization on the denoised ER images (Figure 4b). While standard SN2N processing allowed for the extraction of the main ER network, the resulting skeleton suffered from significant fragmentation, especially in dense regions (Figure 4b, middle row). In stark contrast, the skeleton derived from the aSN2N output is substantially more continuous and complete, preserving the fine, web-like architecture of the organelle (Figure 4b, bottom row). This demonstrates that by providing a cleaner and more faithful representation of the underlying biology, aSN2N enables more reliable and detailed morphological and interaction analyses, which are particularly crucial for studying dynamic and complex organelles.

For dynamic process visualization (Figure 4c), aSN2N time-series (Figure 4c, bottom row) allowed clear observation of mitochondrial events, including fission (yellow arrowheads) and fusion (red arrowheads). These processes exhibited excellent spatiotemporal continuity and structural clarity, facilitating tracking of morphological changes before and after events (white arrowheads). Conversely, in SN2N sequences (Figure 4c, middle row), background artifacts and reduced structural definition hampered event observation, complicating accurate identification of mitochondrial dynamics. Raw images (Figure 4c, top row) rendered these subtle changes nearly indiscernible due to noise.

For dual-color live-cell SD-SIM time-series imaging, aSN2N outperformed SN2N in static image quality, dynamic process fidelity, and segmentation accuracy, which is essential for extracting reliable quantitative information from super-resolution images. These findings further validate the effectiveness of our approach for processing complex biological imaging data, particularly in live-cell dynamic contexts.

3. Theory/calculation

3.1 Adaptive normalization strategy

3.1.1 Design rationale

To theoretically elucidate the potential mechanism by which patch-wise normalization introduces artifacts in specific image patches, we examine its impact on noise statistical properties. Assume the pixel intensity x within an image patch P_i can be modeled as the sum of a true signal s and additive noise n, i.e., $x = x + n$, where the noise n is typically assumed to have a zero mean ($E(n) \approx 0$) and a certain variance σ_n^2.

Image-wise normalization (for which "global" will be used in formula notations for brevity) first normalizes the entire large image I_{large} to the [0, 1] interval. The pixel x'_{global} in a subsequently extracted patch $P'_{i,\text{global}}$ is the result of scaling the original pixel x by the global range $\Delta I_{\text{large}} = \max\left(I_{\text{large}}\right) - \min\left(I_{\text{large}}\right)$:

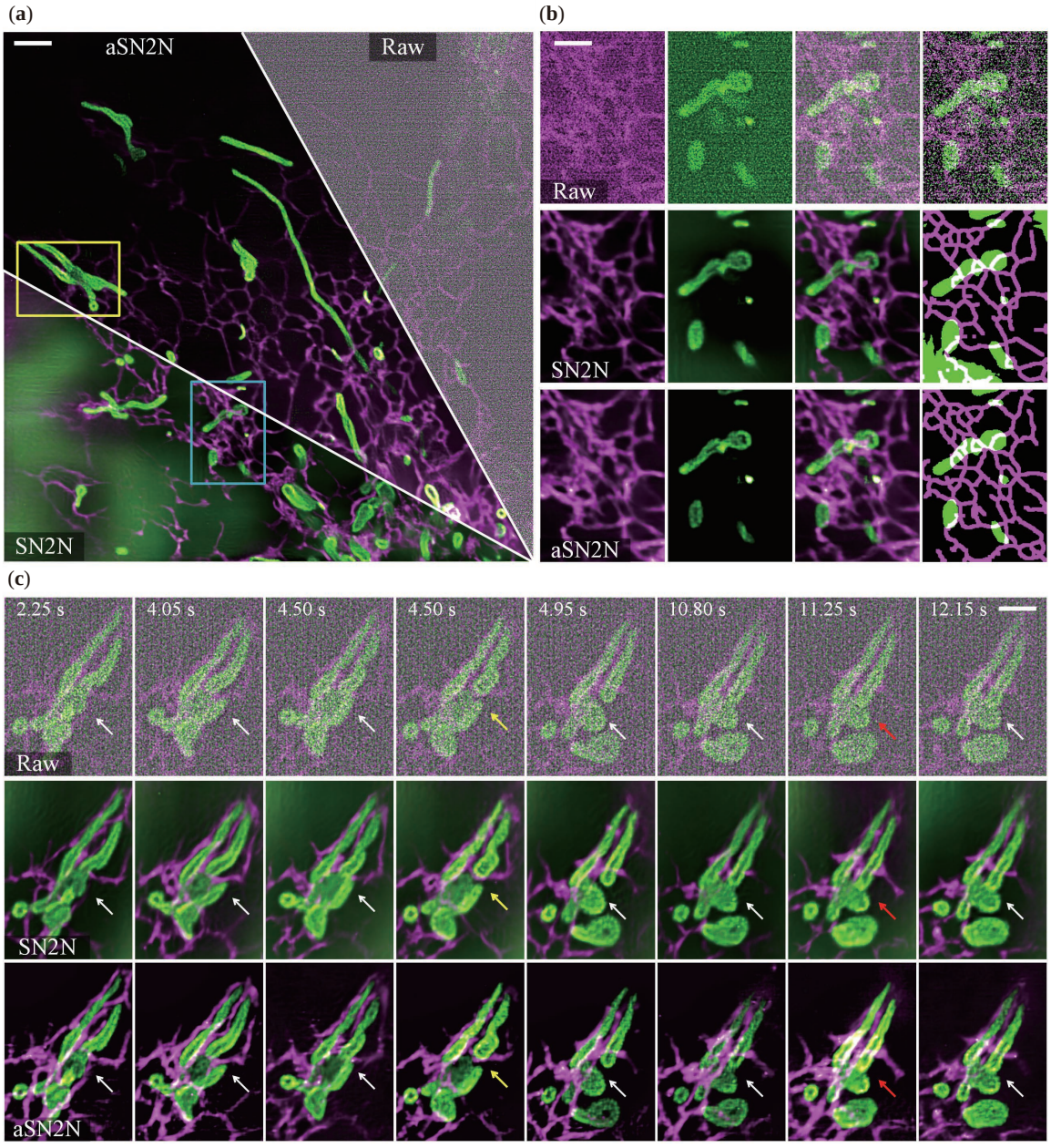


Figure 4 Adaptive-SN2N enhances live-cell dual-color SD-SIM imaging of organelle dynamics by reducing background artifacts and improving segmentation fidelity. **a**. Representative WF SD-SIM image of a live COS-7 cell at t=0 s, co-expressing Tom20-mCherry (mitochondria, shown in green) and Sec61β-EGFP (ER, shown in magenta). **b.** Magnified views of the region indicated by the blue box in **a**, comparing raw input (raw), SN2N, and adaptive-SN2N (aSN2N) processed results. For each method, representative images of individual organelles (mitochondria and ER) and their dual-color merged overlay are presented. Corresponding segmentation masks, obtained using Otsu's hard thresholding (with subsequent skeletonization of ER for SN2N and adaptive-SN2N outputs), are also shown. **c.** Time-lapse image sequences from a representative dynamic region, comparing Raw input, SN2N, and adaptive-SN2N. White arrowheads indicate pre-event structures, yellow arrowheads point to mitochondrial fission events, and red arrowheads highlight mitochondrial fusion events. Time stamps as indicated in the figure. Scale bar: 2 μm (**a**–**c**).

$$x'_{\text{global}} = \frac{x - \min(I_{\text{large}})}{\Delta I_{\text{large}}}. \tag{2}$$

Under this transformation, the original noise variance σ_n^2 is uniformly scaled to:

$$\sigma^2_{n,\text{global}} = \frac{\sigma_n^2}{(\Delta I_{\text{large}})^2}. \tag{3}$$

Crucially, all image patches share the same scaling factor, maintaining their original relative intensity relationships and noise levels.

In contrast, patch-wise normalization (for which "local" will be used in formula notations for brevity) performs min-max normalization individually on each image patch P_i extracted from the original large image I_{large}. In this case, a pixel x within the patch is transformed to:

$$x'_{\text{local}} = \frac{x - \min(P_i)}{\Delta P_i}, \tag{4}$$

where $\Delta P_i = \max(P_i) - \min(P_i)$ is the local dynamic range of that patch.

Correspondingly, the noise variance is scaled to:

$$\sigma^2_{n,\text{local}} = \frac{\sigma_n^2}{(\Delta P_i)^2}, \tag{5}$$

where the scaling factor ΔP_i is local and varies significantly across different image patches, which is the root cause of the problem.

We now analyze two "high-risk" scenarios. The first involves low-signal, low-contrast background patches. In such patches, the signal's intensity is low and varies smoothly, resulting in a very small local dynamic range ΔP_i. According to the formula (Eq. (5)) for the effect of patch-wise normalization on noise variance, $\sigma^2_{n,\text{local}} = \sigma_n^2 / (\Delta P_i)^2$, a small denominator ΔP_i leads to a drastic amplification of the normalized noise variance $\sigma^2_{n,\text{local}}$. This means that insignificant noise fluctuations in the original image are forcibly stretched to the entire [0, 1] range. When the neural network encounters these training samples with disproportionately amplified noise, it may erroneously learn and amplify these noise patterns, ultimately manifesting as fine artifacts during the inference stage. Image-wise normalization, due to its larger, constant denominator I_{large}, avoids this noise amplification effect, helping to maintain the smoothness of background regions.

The second scenario involves sparse strong signal patches, characterized by large areas of low-signal background and a few pixels or structures of extremely high brightness. In such patches, $\min(P_i)$ is determined by the background, close to the background baseline, while $\max(P_i)$ is determined by the brightest signal, far exceeding the background, causing the local range $\Delta P_i = \max(P_i) - \min(P_i)$ to become very large. Consider a background pixel $x_{\text{bg}} = s_{\text{low}} + n_{\text{bg}}$. Its value after patch-wise normalization is $x'_{\text{bg,local}} = \left[x_{\text{bg}} - \min(P_i)\right] / \Delta P_i$. Since x_{bg} is very close to $\min(P_i)$ and the denominator ΔP is large, $x'_{\text{bg,local}}$ is compressed into an extremely narrow range very close to 0. This treatment distorts the original data's intensity distribution by mapping most background pixels to near-zero values. More severely, this creates a critical distribution shift between the training and inference stages. The neural network, during training, learns to associate "background" with these near-zero compressed values. However, during whole-image inference, patches from regions distant from any bright signals do not undergo such extreme compression. The network then misinterprets these uncompressed, "normal" background pixel values as non-zero signals, leading to the generation of unnatural background elevation artifacts. Image-wise normalization, by preserving the relative intensity relationships between pixels, avoids extreme compression of the background and reduces the risk of distribution shift.

In summary, theoretical analysis indicates that indiscriminate use of patch-wise normalization risks amplifying noise when processing low-contrast background patches and may produce artifacts due to distorted intensity distributions and induced distribution shifts when processing sparse strong signal patches.

Based on the above analysis, to enhance the adaptability of SN2N to different image contents and improve denoising stability, we propose an adaptive normalization strategy. Its core idea is to first assess the overall proportion of potentially problematic regions for normalization within a single large original image during the training data preprocessing stage. Then, based on this assessment, a single normalization scheme—either image-wise normalization or patch-wise normalization—is uniformly selected for that entire image. This aims to intelligently balance the trade-offs: when the overall image content is judged to be susceptible to the negative effects of patch-wise normalization, image-wise normalization, which preserves global information, is prioritized; conversely, when the risk is deemed low, the potential local contrast enhancement benefits of patch-wise normalization are leveraged.

3.1.2 Workflow

The specific implementation workflow of this adaptive normalization strategy is as follows: for an input raw image I_{large}, we first normalize the pixel intensities to the range [0, 1] by clipping the top 0.01% outliers to mitigate the impact of hot pixels. This normalization maps diverse biological images into a unified statistical space, allowing for the definition of robust, absolute decision boundaries across different datasets. Subsequently, a series of image patches $\{P_i\}$ are densely extracted from I_{large} according to a predefined patch size (e.g., 128 pixel × 128 pixel) and stride (e.g., 64 pixel, allowing overlap). Next, for each extracted image patch P_i, key statistical metrics are calculated internally to quantify its content features. These metrics primarily include: mean $\mu_i = (1/N) \sum_{p \in P_i} p$ (reflecting overall brightness), standard deviation $\sigma_i = \sqrt{(1/N) \sum_{p \in P_i} (p - \mu_i)^2}$ (measuring internal contrast), and skewness $s_i = \left[(1/N) \sum_{p \in P_i} (p - \mu_i)^3 \right] / \sigma_i^3$ (measuring the asymmetry of the intensity distribution, where high positive skewness often indicates sparse bright features).

Guided by the theoretical analysis in Sect. 3.1.1 and confirmed through statistical scatter plots of patches from diverse biological samples, we identified two distinct clusters corresponding to artifact-prone regions:

(1) Low-signal, low-contrast background: these patches are prone to noise amplification. In the unified statistical space, they cluster in the region of low mean intensity and low standard deviation. Additionally, while signal-bearing regions typically show positive skewness, background noise tends to exhibit a symmetric or slightly negative skewness distribution.

(2) Sparse strong signal: these patches can cause intensity distribution shifts. They exhibit a distinct signature: a low mean intensity (dominated by background) but a disproportionately high standard deviation and high positive skewness, driven by sparse, high-intensity outliers.

Based on these data-driven observations, we define specific boundaries (detailed in Sect. 6.2.3) to identify these high-risk patches.

After completing the risk assessment for all patches within I_{large}, the total number of patches identified as high-risk, $N_{high\ risk}$, is counted, and its proportion relative to the total number of extracted patches, N_{total}, is calculated as $R_{high\ risk} = N_{high\ risk} / N_{total}$.

Finally, this high-risk patch proportion $R_{high\ risk}$ guides the image-wise normalization decision for the entire large image I_{large}. We set a proportion threshold $T_{ratio} = 0.1$. Our sensitivity analysis (Table 2) reveals a clear separation between artifact-prone datasets (e.g., CCPs consistently >25%) and safe datasets (e.g., F-actin <8.5%). We selected 10% as a conservative boundary situated just above the maximum observed ratio for dense structures. This design prioritizes artifact prevention, ensuring that any image deviating even slightly from the safe structural baseline defaults to the safer image-wise normalization. If $R_{high\ risk} > T_{ratio}$, it indicates that the image contains a significant number of regions likely to cause problems with patch-wise normalization. In this case, image-wise normalization is applied to this image. Specifically, the entire large image I_{large} first undergoes image-wise normalization, and then all training patches are extracted from the normalized I'_{large}. Conversely, if $R_{high\ risk} \leqslant T_{ratio}$, the image is considered to have an overall low risk and may benefit from local contrast enhancement; thus, patch-wise normalization is applied. This involves extracting all training patches P_i from the original large image I_{large}, and then performing patch-wise normalization independently on each patch P_i.

By implementing this adaptive procedure, we determine a single and unified normalization scheme for each original large image I_{large} in the training dataset. Consequently, the final generated training dataset consists of a series of image patches characterized by a key feature: all training patches originating from the same original large image either entirely carry intensity information adjusted by global context (if that image was deemed high-risk) or have all undergone independent local intensity stretching (if that image was deemed low-risk). This unified processing approach, based on global risk assessment, aims to make the data distribution encountered by the model during training more aligned with practical application scenarios. It allows for the preservation of global information to suppress artifacts when necessary, while also leveraging local contrast information when appropriate, thereby optimizing the model's overall denoising performance and stability across different types of biological images.

3.2 Weighted overlap inference

Deep learning models based on the U-Net [40] architecture (including the SN2N network used as a basis in this study) face a challenge due to their inherent limited receptive field (RF) when applied to large-sized images. This means that when the network predicts pixel values in the edge regions of an image, it cannot access complete contextual information, potentially leading to reduced prediction accuracy. A straightforward strategy to address this is to segment the large image into smaller patches for independent inference and then stitch the results together. However, this simple patch-stitching method is highly prone to introducing noticeable discontinuities or artifacts (stitching artifacts) [41,42] at the boundaries between patches, severely compromising the visual quality of the final output image and the reliability of subsequent quantitative analysis. To overcome this limitation, we implement a Weighted Overlap Inference mechanism. The core idea of this mechanism is to employ a sliding window strategy to process the input image and use weighted averaging to smoothly fuse the prediction results from overlapping regions, aiming to generate a seamless and high-quality full-sized denoised image.

Specifically, this inference workflow first defines the basic processing unit of the network—the block size, e.g., B pixel × B pixel (in this study, B = 128), which is typically consistent with or related to the image patch size used during training. Concurrently, a stride S is set, which is smaller than the block size B (e.g., $S = B / 2 = 64$ pixel), to ensure significant overlap between

Table 2 Threshold sensitivity analysis under two typical cellular architectures

Organelle	Parameter	Base value	50%	75%	100%	125%	150%
CCPs	$T_{\mu 1}$	0.200	25.584,9	30.786,5	36.660,0	37.941,8	37.954,2
	$T_{\sigma 1}$	0.028	25.472,9	30.574,9	36.660,0	42.844,7	47.299,7
	T_{s1}	0.000	31.458,4	34.743,7	36.660,0	37.618,2	38.190,6
	$T_{\sigma aux}$	0.050	28.982,1	33.909,9	36.660,0	36.660,0	36.660,0
	$T_{\mu 2}$	0.300	28.683,4	33.773,0	36.660,0	37.269,8	37.269,8
	$T_{\sigma 2}$	0.075	54.766,1	44.288,2	36.660,0	31.931,3	29.280,7
	T_{s2}	3.000	43.093,6	41.737,2	36.660,0	31.446,0	29.367,8
F-actin	$T_{\mu 1}$	0.200	1.891,7	2.547,3	4.213,2	4.331,5	4.331,5
	$T_{\sigma 1}$	0.028	1.848,7	2.375,3	4.213,2	5.889,9	7.813,8
	T_{s1}	0.000	4.041,3	4.095,0	4.213,2	4.643,2	5.761,0
	$T_{\sigma aux}$	0.050	4.062,8	4.159,5	4.213,2	4.492,7	4.944,1
	$T_{\mu 2}$	0.300	3.847,8	3.944,5	4.213,2	4.213,2	4.213,2
	$T_{\sigma 2}$	0.075	4.804,4	4.439,0	4.213,2	4.041,3	3.923,0
	T_{s2}	3.000	8.469,5	5.417,0	4.213,2	3.901,5	3.858,6

adjacent processing blocks (50% in this example). Using this block size and stride, a series of mutually overlapping image blocks $\{P_{\text{noisy},i}\}$ are extracted from the entire input noisy image I_{noisy} via a sliding window approach. For image boundary regions, if the remaining size is less than $B \times B$, mirror padding is used to virtually extend the image, ensuring all regions are fully covered, and each extracted block has the standard $B \times B$ dimensions.

Next, for each extracted noisy block $\{P_{\text{noisy},i}\}$, our trained enhanced SN2N model f_θ is applied for denoising inference. Before feeding the image block into the network, normalization preprocessing is required. Considering that the adaptive normalization strategy was employed during training, and we aim for stability and consistency in the inference phase, we adopt the same data processing approach as the adaptive normalization strategy. That is, images ultimately selected for patch-wise normalization still undergo patch-wise normalization, and similarly for image-wise normalization. Subsequently, the normalized block $P'_{\text{noisy},i}$ is input to the network, yielding the corresponding denoised prediction block $P_{\text{denoised},i} = f_\theta\left(P'_{\text{noisy},i}\right)$.

To seamlessly fuse these overlapping prediction blocks $\{P_{\text{denoised},i}\}$ into a complete denoised image, we introduce a weighted averaging mechanism. For this purpose, we design a two-dimensional Gaussian weighting function $W(x,y)$ of the form:

$$W(x,y) = \exp\left[-\frac{(x-x_c)^2+(y-y_c)^2}{2\sigma_w^2}\right], \tag{6}$$

where (x,y) are the pixel coordinates within the block, (x_c,y_c) are the center coordinates of the block (i.e., $(B/2, B/2)$), and σ_w is the standard deviation controlling the spatial decay rate of the weights. This function assigns the highest weights (close to 1) to pixels in the central region of the block, with weights smoothly decaying towards the block edges. The choice of σ_w needs to be coordinated with the block size B and stride S to ensure a smooth transition of weights in overlapping regions (in this study, $\sigma_w = B/4$). For each prediction block $P_{\text{denoised},i}$, we generate a corresponding $B \times B$ weight map W_i.

In the final fusion stage, we initialize two zero matrices of the same dimensions as the original input image I_{noisy}: $I_{\text{accumulated result}}$ for accumulating the weighted denoised pixel values, and $I_{\text{accumulated weight}}$ for accumulating the corresponding weight values. Then, we iterate through all predicted denoised blocks $P_{\text{denoised},i}$ and their corresponding weight maps W_i. The weighted result of each block, $P_{\text{denoised},i} \odot W_i$ ($\odot$ denotes element-wise multiplication), is added to the $I_{\text{accumulated result}}$ matrix at the corresponding spatial location. Simultaneously, the weight map W_i itself is also added to the $I_{\text{accumulated weight}}$ matrix at the same location. After iterating through and accumulating all blocks, the final, seamlessly stitched full-sized denoised image I_{denoised} is calculated by element-wise division:

$$I_{\text{denoised}}(x,y) = \frac{I_{\text{accumulated result}}(x,y)}{I_{\text{accumulated weight}}(x,y)+\varepsilon}, \tag{7}$$

where ε is a very small positive number (e.g., 10^{-8}) used to prevent division-by-zero errors in regions where the accumulated weight might be close to zero (especially at the outermost edges of the image). This weighted overlap inference method effectively eliminates boundary artifacts introduced by simple patch stitching by smoothly fusing the contributions from overlapping regions, thereby generating visually natural and spatially continuous high-quality denoised results.

It is important to note that this weighted overlap mechanism, while crucial for eliminating stitching artifacts, inherently increases the computational cost of the inference process. By employing a 50% overlap in each dimension, each pixel in the interior of the image is processed approximately four times, and our empirical measurements confirm an increase in processing time compared to a naive, non-overlapping patch-stitching approach. For a typical 502 pixel × 502 pixel image on an NVIDIA RTX 4090 GPU, inference time with this method is approximately 0.13 s, whereas a naive approach takes about 0.02 s. However, this trade-off is well-justified by the significant improvement in image quality, as quantitatively demonstrated in our ablation study (Sect. 6.3, Table 1). Furthermore, since most fluorescence microscopy denoising tasks are performed offline and are not subject to real-time constraints, we contend that this increased processing time is a practical and acceptable compromise for achieving superior, artifact-free results.

4. Discussion

4.1 Biological fidelity and analytical reliability

The background artifacts generated by deep learning methods during denoising constitute a distortion of biological information, particularly concerning because they can mimic biological structures and lead to misinterpretations of cellular dynamics and organization. Our proposed adaptive-SN2N addresses this challenge through risk-aware optional normalization and Gaussian-weighted prediction mechanisms that adaptively suppress these artifacts, resolving the trade-off between artifact minimization and signal preservation in denoising algorithms.

This failure of indiscriminate patch-wise normalization is not a matter of insufficient training data but a fundamental problem of input data distortion that misleads the self-supervised learning process. Our analysis, both theoretical (Sect. 3.1.1) and experimental (Sect. 2.2), reveals that patch-wise normalization imposes a content-dependent transformation on local data statistics, creating two problematic scenarios. First, for low-contrast background regions, it drastically amplifies the inherent noise by stretching a small native dynamic range to the full [0, 1] interval. This violates a core assumption of self-supervised methods like Noise2Noise—that noise is statistically independent of the signal—causing the learning objective to incorrectly interpret this amplified noise as a structural feature to be preserved. Second, for sparse, high-signal regions, it compresses the vast majority of background pixels into a near-zero intensity range. This creates a critical distribution shift between the data seen during training and the data encountered during full-image inference, where background patches are not as compressed. This discrepancy causes the trained model to misinterpret these normal background fluctuations as meaningful signals, resulting in

artifactual elevations. Therefore, the problem is not a weakness in the self-supervised paradigm itself but an ill-posed learning task created by the preprocessing step, which our adaptive strategy directly resolves by preserving global intensity context when the risk of such distortion is high.

Through systematic analysis of experimental results, we found that adaptive-SN2N performs optimally when selecting image-wise normalization for most biological specimens. Our findings reveal a fundamental insight into biological image processing and challenge conventional computer vision assumptions: in biological contexts, preserving global intensity relationships often takes precedence over enhancing local contrast. Unlike natural images, biological microscopy data exhibit extremely high background-to-signal ratios, structural sparsity, and heterogeneous spatial distributions that render many general-purpose deep learning strategies unsuitable. The relative intensity distribution across an entire field of view carries critical information about organelle relationships and protein expression gradients that patch-wise normalization inadvertently disrupts. This suggests biological image processing requires specialized deep learning paradigms designed explicitly for the intrinsic properties of biological data—from preprocessing to network architecture—with broad implications for future biological image analysis tools.

The practical impact of our approach extends to critical research areas such as neuroscience and cancer biology, where background artifacts pose significant analytical challenges. In neural circuit studies, artifacts from conventional denoising methods are frequently misinterpreted as synaptic structures, with false-positive detections accounting for up to 60% of errors in automated vesicle quantification15 [43]. These artifacts, combined with noise-induced systematic errors at low SNRs, compromise the accuracy of synaptic connectivity analysis [44,45]. In cancer research, the situation is equally problematic: cellular protrusions—critical indicators of metastatic potential—are particularly vulnerable to artifact contamination [46-48]. Cancer cells extend long protrusions (30–100 μm) through basement membranes during invasion, and distinguishing these genuine structures from background artifacts is essential for predicting tumor metastasis [47,49]. However, conventional patch-wise normalization exacerbates these difficulties by amplifying background variations, making it challenging to identify authentic protrusion-localized signals [49]. By providing cleaner backgrounds and preserving structural integrity of fine cellular processes—including synaptic puncta and migratory protrusions—adaptive-SN2N improves analytical reliability across these domains [43,50].

4.2 Limitations and future directions

Despite the demonstrated advantages of adaptive-SN2N in suppressing artifacts and improving biological fidelity, it is important to acknowledge its limitations, which also point toward promising avenues for future research.

First, the core of our adaptive strategy relies on a set of heuristic, rule-based parameters for risk assessment (e.g., the threshold T_{ratio} detailed in Sect. 6.2.3). Although we designed these parameters by unifying images into a standard coordinate space to enhance robustness, their optimal values were still determined empirically through experiments on our validation datasets. This dependency on pre-tuned parameters implies that while the method is adaptive to various cell types, it may require parameter adjustments to generalize optimally to new microscopy modalities or biological samples with fundamentally different statistical properties (e.g., non-fluorescence imaging like brightfield or differential interference contrast (DIC), or images with structured, non-Gaussian noise). Future work could explore data-driven approaches, such as using a meta-learning framework, to learn these decision parameters automatically, thereby creating a more universally applicable and fully automated system.

Second, our adaptive normalization makes a single, binary decision for an entire image—applying either image-wise or patch-wise normalization uniformly. This global choice is a deliberate design to ensure a consistent data distribution is presented to the neural network for a given image, which is crucial for preventing severe artifacts during model training and inference. However, this approach may represent a compromise for highly heterogeneous images that contain both dense, well-structured regions (which would benefit from patch-wise normalization) and large, sparse background areas (which necessitate image-wise normalization).

A seemingly straightforward solution would be a region-adaptive strategy, where normalization is chosen locally on a patch-by-patch basis. We explored this direction during our initial investigations. However, this local approach introduces a more challenging problem: severe inconsistencies in data distribution and scale between adjacent patches. Feeding such heterogeneous data into a single network confuses the learning process. Alternatively, using separate networks for different normalization schemes would create significant discontinuities and stitching artifacts at the boundaries between regions processed differently, undermining the primary goal of artifact suppression. Therefore, a truly effective region-adaptive implementation would require a more sophisticated architecture, potentially involving a gating network to dynamically blend the outputs from different processing pathways or a multi-scale framework that can handle local and global statistics simultaneously. Developing such a model remains a non-trivial challenge and constitutes a promising avenue for future research.

Finally, the performance of adaptive-SN2N is intrinsically linked to the underlying self-supervised denoising model (SN2N). While effective, self-supervised methods assume that noise is statistically independent between pixels. They may struggle with images containing significant structured or correlated noise, which could be misinterpreted as a biological signal. Additionally, like most deep learning models, there remains a risk of over-smoothing or failing to restore extremely fine, low-SNR structures that closely resemble noise, potentially leading to a loss of subtle details.

Addressing these limitations will be key to developing the next generation of robust computational microscopy tools. We believe that our current work provides a strong foundation and a conceptual framework for these future advancements.

4.3 Outlook and conclusion

Looking forward, we envision extending this adaptive framework

beyond denoising to other computational methods in biological imaging. This risk-aware approach could fundamentally transform techniques that currently rely on background removal preprocessing or post-processing steps. Implementing similar adaptive strategies in segmentation, colocalization analysis, and feature extraction algorithms could enable these methods to intrinsically suppress background artifacts without additional processing steps. Furthermore, fully data-driven adaptive schemes —such as meta-networks that predict optimal processing parameters or dynamically adjust strategies for different image regions—could advance toward a universal biological image enhancement paradigm. Combined with meta-learning, such systems could rapidly adapt to diverse imaging modalities and biological structures, addressing challenges from content-aware data compression to multimodal image fusion.

In conclusion, adaptive-SN2N represents a fundamental rethinking of computational microscopy for biological image processing. By prioritizing biological fidelity through risk-aware processing, our framework addresses the unique requirements of quantitative biological imaging, demonstrating that the most effective computational methods in biology are those designed with explicit consideration of biological structures' unique properties and life science research's specific analytical needs.

5. Conclusions

We developed adaptive-SN2N to address the critical challenge of background artifacts in self-supervised denoising algorithms for fluorescence microscopy. Our approach integrates three key innovations: a risk-aware adaptive normalization strategy that selects optimal preprocessing based on image characteristics, the core self-inspired learning mechanism that enables training without paired data, and a weighted overlap inference mechanism that eliminates stitching artifacts during reconstruction.

We evaluated adaptive-SN2N across multiple datasets—including the BioSR benchmark, fixed-cell SD-SIM data, and dual-color live-cell imaging. Our evaluations demonstrated adaptive-SN2N's superior performance in both image quality enhancement and background artifact suppression. Our method inherits the substantial photon efficiency improvements—often by one to two orders of magnitude—characteristic of the underlying self-supervised framework. By synergistically combining this capability with our novel adaptive artifact suppression, our method enhances the visualization of biological structures and improves the reliability of downstream analysis, particularly valuable for analyzing sparsely distributed biological structures and dynamic cellular processes.

6. Materials and methods

6.1 Datasets acquisition

To evaluate the performance of the proposed adaptive-SN2N framework, we employed multiple microscopy imaging datasets exhibiting diverse biological structural features, imaging conditions, and noise characteristics. These datasets comprised both public benchmarks and custom-acquired super-resolution data.

6.1.1 Biomedical imaging dataset for super-resolution microscopy dataset

We utilized the publicly available BioSR dataset, which consists of images acquired using SIM. The BioSR dataset provides images at nine distinct SNR levels (ranging from level 1 to level 9), along with their corresponding raw GT images that serve as a reference. In all experiments conducted for this study, WF images were exclusively employed for both training and inference processes. These WF images were generated by averaging nine raw SIM frames. Specifically, the WF images utilized were derived from the raw SIM frames associated with SNR level 1, representing the most challenging noise condition within the dataset.

6.1.2 Spinning disk-based structured illumination microscopy setup

Two commercial SD-SIM systems were employed for super-resolution confocal image acquisition. The first system utilized was a SpinSR10 (Olympus) equipped with a high-numerical-aperture WF objective (×100/1.49 NA oil immersion, APON, Olympus) and a scientific CMOS camera (ORCA Fusion, Hamamatsu). The system incorporated two laser lines (488 and 561 nm) for multi-channel fluorescence excitation. The optical detection pathway featured an additional ×3.2 magnification, yielding a total system magnification of ×320. Images were acquired in SoRa super-resolution mode, with the capability to switch to conventional SD-confocal mode for comparative imaging. The second system, a Live-SR SD-SIM, was configured around an inverted fluorescence microscope (IX81, Olympus) with a WF objective (×100/1.3 NA oil immersion, Olympus), integrated with a scanning confocal unit (CSU-X1, Yokogawa) and a Live-SR module (GATACA Systems). This system also utilized two laser lines (488 and 561 nm) for excitation. Image detection was performed using either a scientific CMOS camera (C14440-20UP, Hamamatsu).

6.1.3 Sample preparation and preprocessing

To label late endosomes or Lys, we incubated COS-7 cells in 50 nmol/L LysoTracker Deep Red (Thermo Fisher Scientific, L12492) for 45 min and washed them three times in PBS before imaging. To label mitochondria, COS-7 cells were incubated with 250 nmol/L MitoTracker Green FM (Thermo Fisher Scientific, M7514) and 250 nM (1 nM = 1 nmol/L) MitoTracker Deep Red FM (Thermo Fisher Scientific, M22426) in HBSS containing Ca^{2+} and Mg^{2+} or no phenol red medium (Thermo Fisher Scientific, 14025076) at 37 °C for 15 min before being washed three times before imaging. To perform nuclear staining on COS-7 cells, SPY650-DNA (Cytoskeleton, CY-SC501) was diluted to 1:1,000 in PBS for ~1 h and washed three times in PBS. To label cells with genetic indicators, COS-7 cells were transfected with LAMP1-EGFP and mGold-TOM20. The transfections were executed using Lipofectamine 2000 (Thermo Fisher Scientific, 11668019) according to the manufacturer's instructions. After transfection, cells were plated on precoated coverslips. Live cells were imaged in a complete cell culture medium containing no phenol red in a 37 °C live-cell imaging system. For the calcium lantern imaging in SD-SIM, the calcium signal was stimulated with a micropipette containing 10 μmol l−15'-ATP-Na2 solutions (Sigma-Aldrich, A1852).

For the fixed-sample experiments, we manually created GT by enlarging the exposure time to acquire the high-SNR data, which served as labels. To remove potential low baseline background and noise, the GT images were then created from these high-SNR images by subtracting a constant background value and subsequently filtering with a small Gaussian kernel.

6.2 Implementation details

6.2.1 Network architecture

Adaptive-SN2N utilized a 2D U-Net architecture with symmetric encoder and decoder paths, each comprising four levels. The encoder implements downsampling through blocks of two 3×3 convolutions, batch normalization, and Leaky ReLU activation (negative slope 0.02), followed by 2×2 max pooling. Feature channels begin at 64 and double after each downsampling step, reaching 512 at the bottleneck. The decoder employs Bilinear Upsampling, concatenating upsampled feature maps with corresponding encoder maps via skip connections. These concatenated features are processed through similar convolution blocks, progressively halving channel count back to 64. The final layer consists of a 1×1 convolution mapping features to the desired output channels with a Sigmoid activation for generating the denoised image.

6.2.2 Training process

The model was implemented using the PyTorch framework and trained on a workstation with an NVIDIA RTX 4090 GPU (24GB VRAM). For model optimization, we employed the Adam optimizer with hyperparameters set to $\beta_1 = 0.5$ and $\beta_2 = 0.999$, and an initial learning rate of 2×10^{-4}. A learning rate scheduling strategy was utilized, reducing the learning rate by a factor of 0.5 if the validation loss did not improve for 10 consecutive epochs.

The model was trained independently for 100 epochs on each biological structure dataset (e.g., CCPs, ER, MTs from the BioSR benchmark), where the number of raw images for each structure ranges from 51 to 68 (specifically, 54 for CCPs, 68 for ER, 51 for F-actin, and 55 for MTs). To prepare the training data, we systematically extracted patches from each full-sized image using a sliding window approach rather than random cropping. Specifically, for a typical 502 pixel × 502 pixel image from the BioSR dataset, we extracted 128 pixel × 128 pixel patches with a stride of 32 pixels in both dimensions. This procedure yielded 144 overlapping patches per image. To further augment the training set and enhance model robustness, we applied a 90-degree rotation to each extracted patch, effectively doubling the number of training samples to 288 per image. The batch size during training was set to 32.

The self-constraining loss $L = L_1 + L_2 + \lambda L_3$, with the weight for the consistency term λ set to 0.5. Crucially, the loss function was evaluated at the level of individual patches. The entire training process, from patch extraction with adaptive normalization to loss calculation, was performed on these 128 pixel × 128 pixel patches, not on the full-sized images after reconstruction.

6.2.3 Adaptive normalization and inference parameters

The thresholds for identifying high-risk patches were determined by analyzing the statistical distribution of over 40,000 patches extracted from the validation dataset. By projecting patch statistics onto 2D planes (mean-std. and mean-skewness), we established decision boundaries that robustly separate artifact-prone background clusters from structural signal clusters. The specific criteria for the [0, 1] normalized images are:

(1) For low-signal, low-contrast background detection

Mean ($T_{\mu 1} = 0.2$) and standard deviation ($T_{\sigma 1} = 0.028$): Identifies patches significantly darker and smoother than signal regions. The mean threshold of 0.2 captures the majority of background areas, while the standard deviation threshold acts as a strict limit to exclude textured structures.

Skewness ($T_{s1} = 0.0$) and auxiliary std. ($T_{\sigma\,\mathrm{aux}} = 0.05$): background noise typically exhibits symmetric or negative skewness. We enforce a skewness check ($s_i \leqslant T_{s1}$) combined with a secondary standard deviation constraint ($\sigma_i \leqslant T_{\sigma\,\mathrm{aux}}$). This combined rule prevents the misclassification of structural regions that may exhibit negative skewness (e.g., holes within dense structures) but possess a higher standard deviation than noise.

(2) For sparse strong signal detection

Mean ($T_{\mu 2} = 0.3$), standard deviation factor ($T_{\sigma 2} = 0.075$), and skewness ($T_{s2} = 3.0$): targets patches that are predominantly dark (low mean) but contain extreme outliers (very high contrast and skewness).

These specific values were selected to be conservative, prioritizing the correct identification of artifact-prone regions. The statistical separability of these categories and the validity of the chosen decision boundaries are visually demonstrated in Figure 5, where distinct common features are observed in the Mean-Std.-Skewness projection planes. Crucially, we conducted a sensitivity analysis to verify that our method does not rely on precise parameter tuning. We varied each threshold parameter individually from 50% to 150% of its reference value and monitored the stability of the final normalization decision (patch-wise vs. image-wise). As shown in Table 2, the high-risk patch ratio ($R_{\mathrm{high_ratio}}$) for CCPs (a typical sparse sample) consistently remained well above the decision threshold ($T_{\mathrm{ratio}} = 0.1$), while for F-actin (a typical dense sample), it stayed well below 0.1 across the variation range. This demonstrates that the final decision relies on the robust separation of statistical clusters rather than the precise marginal classification of individual patches.

During the weighted overlap inference stage, the block size used was 128 pixel ×128 pixel, with a stride of 16 pixel. The standard deviation σ_w of the Gaussian weighting function was set to one-fourth of the block size, i.e., $\sigma_w = 32$.

6.2.4 Software and hardware

All code was developed in Python with NumPy [51] and scikit-image [52] dependencies. Training and inference were performed on a workstation with NVIDIA RTX 4090 GPU (24 GB VRAM). Data processing utilized Fiji/ImageJ. Figures were prepared using Fiji/ImageJ, Microsoft Visio, and Origin Pro.

6.3 Performance metrics

To quantitatively evaluate the denoising efficacy of our proposed method, we compared the denoised images with their corresponding GT counterparts using several established image quality metrics. This evaluation was performed on the BioSR and fixed-

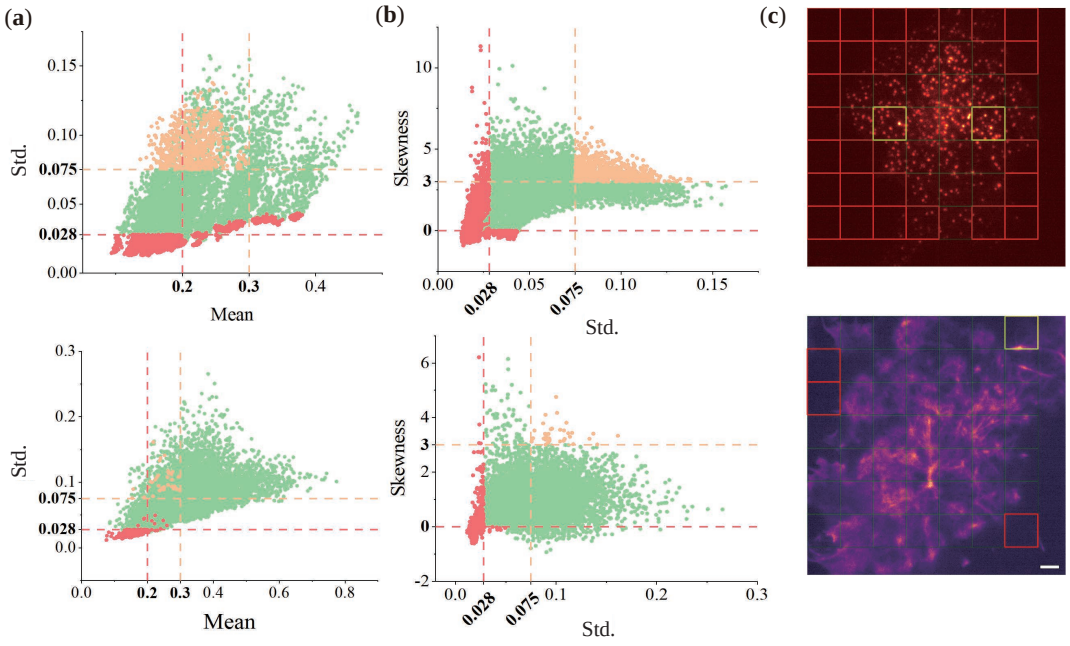


Figure 5 Data-driven statistical rationale for artifact risk assessment. **a**–**b**. Statistical scatter plots of image patches extracted from a representative sparse sample (CCPs, top row) and a dense sample (F-actin, bottom row). Each point represents a single image patch (normalized to [0, 1]). Color coding: red points indicate identified "high-risk background" patches (characterized by low mean, low std, and low/negative skewness); Orange points indicate "high-risk sparse" patches (characterized by low mean but high std. and high positive skewness); green points represent "low-risk" structural patches. Decision boundaries: indicated by dashed lines in corresponding colors, representing the absolute parameter thresholds used in the algorithm. **c**. Spatial visualization of the risk assessment on the corresponding raw images. Colored boxes correspond to the classifications in the scatter plots: red boxes indicate high-risk background; yellow boxes indicate high-risk sparse; green grid lines indicate low-risk (normal) patches. Scale bar: 3 μm **(c)**.

cell SD-SIM datasets. Specifically, we computed MSE to quantify pixel-wise differences, PSNR to measure fidelity, and SSIM to assess perceptual similarity by comparing luminance, contrast, and structure. For these metrics, lower MSE values and higher PSNR and SSIM values (with 1 being a perfect match for SSIM) signify a more accurate denoising result.

To specifically quantify the suppression of background artifacts, we introduced the background variance ratio (BVR). This metric evaluates the smoothness and consistency of the background in a denoised image relative to the ground truth. First, a background mask M_{bg} is generated by thresholding the GT image to identify low-intensity background pixels. The variance of these background pixels is then calculated for both the GT image ($\sigma^2_{\mathrm{bg,GT}}$) and the corresponding pixels in the denoised image ($\sigma^2_{\mathrm{bg,denoised}}$). The BVR is defined as their ratio:

$$\mathrm{BVR} = \frac{\sigma^2_{\mathrm{bg,denoised}}}{\sigma^2_{\mathrm{bg,GT}}}. \tag{8}$$

Given that GT images are also derived from real-world captures (achieved through extremely high light intensity or multi-frame averaging), they inevitably contain a small amount of noise. Therefore, ideally, a lower BVR value is preferable. A BVR less than 1 indicates significant denoising, with background smoothness even surpassing the ground truth. A BVR around 1 suggests good denoising performance, with background noise levels nearly identical to GT. Conversely, a BVR significantly greater than 1 implies that the result contains a considerable degree of background artifacts.

Finally, for super-resolution datasets, specifically the fixed-cell SD-SIM data, FRC was employed to assess the effective spatial resolution of the denoised images. The FRC methodology quantifies image resolution by calculating a correlation value for discrete frequency bins (or rings) in Fourier space. This is achieved by comparing the Fourier transforms, $F_1(r)$ and $F_2(r)$, of two independent images (often generated from splitting a single dataset into two halves, commonly referred to as half-datasets) that represent noisy instances of the same underlying structure. The FRC value between these two images, denoted as FRC, for the i-th frequency bin, r_i, is computed according to the formula:

$$\mathrm{FRC}(r_i) = \frac{\sum_{r\in r_i} F_1(r)\cdot F_2(r)^*}{\sqrt{\left[\sum_{r\in r_i} F_1^2(r)\right]\cdot\left[\sum_{r\in r_i} F_2^2(r)\right]}}, \tag{9}$$

where $F_1(r)$ and $F_2(r)$ represent the Fourier transform coefficients of the two respective images at a specific spatial frequency r within the i-th frequency bin r_i. The term $F_2(r)^*$ denotes the complex conjugate of $F_2(r)$, and the summation $\sum_{r\in r_i}$ is performed over all spatial frequency components r belonging to the bin r_i. An FRC curve (or FRC histogram) is generated by plotting these correlation values as a function of spatial frequency. The effective image resolution is subsequently determined from this curve, typically defined as the cut-off spatial frequency at which the FRC value descends below a pre-established threshold (e.g., common

thresholds include 1/7, approximately 0.143, or 0.5). An FRC curve that sustains higher correlation values at higher spatial frequencies signifies a more faithful preservation of fine structural details and thus indicates a superior effective resolution.

To determine whether the observed differences in performance metrics were statistically significant, we performed a Wilcoxon signed-rank test, which is a non-parametric test suitable for paired non-normally distributed data. For each organelle dataset, we treated the metric values from each processed image as a paired sample. We conducted pairwise comparisons between key model configurations to evaluate the impact of specific architectural choices (e.g., comparing patch-wise vs. image-wise normalization, and our full model vs. its ablated version). A p-value of less than 0.05 was considered statistically significant. In our results, significance levels are denoted as $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Complementing the quantitative metrics, which are contingent upon GT availability, a rigorous qualitative visual assessment was conducted for all datasets. This is particularly crucial for datasets such as live-cell imaging, where GT is inherently absent. This visual evaluation critically examined the denoised images to ensure the faithful preservation of fine structural details and edge sharpness, while meticulously avoiding over-smoothing that could obscure subtle features. Concurrently, the assessment focused on the restoration of background regions to a uniform and clean state, ensuring they were devoid of unnatural intensity elevations, residual noise patterns, or other inconsistencies. Of paramount importance was the algorithm's ability to suppress artifacts and, critically, to prevent the introduction of spurious or algorithm-generated structures (e.g., network-induced hallucinations), blotches, or undesirable blurring.

To visually substantiate the performance differences in these qualitative aspects, results will be presented using side-by-side comparative figures. These figures will display original images, the outputs from adaptive-SN2N and other comparative methods, alongside GT images (where available). Magnified views of selected ROIs will further elucidate subtle differences. Such detailed visual analysis is indispensable for a comprehensive understanding of the algorithm's practical utility and its fidelity in the context of imaging complex biological phenomena.

Data availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request. Our implementation of the aSN2N code can be found at https://github.com/SR-Wiki/aSN2N.

Authors' contributions Liying Qu and Weisong Zhao conceived the research and supervised the project. Jingyang Zhu designed and implemented the core algorithms. Shiqun Zhao and Xianxin Ye performed the experiments and collected the data across all datasets. Yuanyuan Huang and Yuzhen Liu analyzed the data and prepared the figures. Liying Qu and Jingyang Zhu wrote the manuscript with input from all authors. All authors discussed the results, contributed to the manuscript revision, and approved the final manuscript.

Compliance and ethics The author(s) declare that they have no conflict of interest.

Funding This work was supported by the National Key Research and Development Program of China (2024YFC3406600, 2022YFC3400600, 2025YFF0518103, 2023YFC3402600, 2024YFA1307200), the National Natural Science Foundation of China (NSFC) (32422052, 62305083, 32227802, 624B2054, T2222009), the Scientific Research Innovation Capability Support Project for Young Faculty (ZYGXQNJSKYCXNLZCXM-H7), the Young Elite Scientists Sponsorship Program by China Association for Science and Technology (2023QNRC001), the China National Postdoctoral Program for Innovative Talents (BX20250481), the China Postdoctoral Science Foundation (2025M784357).

References

1. Balasubramanian H, Hobson CM, Chew TL, Aaron JS. Imagining the future of optical microscopy: Everything, everywhere, all at once. Commun Biol. 2023; 6: 1096.
2. Eling N, Morgan MD, Marioni JC. Challenges in measuring and understanding biological noise. Nat Rev Genet. 2019; 20: 536–48.
3. Mishin AS, Lukyanov KA. Live-cell super-resolution fluorescence microscopy. Biochemistry Mosc. 2019; 84: 19–31.
4. Schermelleh L, Ferrand A, Huser T, Eggeling C, Sauer M, Biehlmaier O, et al. Super-resolution microscopy demystified. Nat Cell Biol. 2019; 21: 72–84.
5. Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M, et al. Noise2Noise: Learning image restoration without clean data. In: Proceedings of the 35th International Conference on Machine Learning. Stockholm: International Machine Learning Society; 2018. p. 2965–74.
6. Batson J, Royer L. Noise2Self: Blind denoising by self-supervision. In: Proceedings of the 36th International Conference on Machine Learning. California: International Conference on Machine Learning; 2019. p. 524–33.
7. Krull A, Buchholz TO, Jug F. Noise2Void-learning denoising from single noisy images. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). California: IEEE; 2019. p. 2124–32.
8. Eom M, Han S, Park P, Kim G, Cho ES, Sim J, et al. Statistically unbiased prediction enables accurate denoising of voltage imaging data. Nat Methods. 2023; 20: 1581–92.
9. Zhang G, Li X, Zhang Y, Han X, Li X, Yu J, et al. Bio-friendly long-term subcellular dynamic recording by self-supervised image enhancement microscopy. Nat Methods. 2023; 20: 1957–70.
10. Qu L, Zhao S, Huang Y, Ye X, Wang K, Liu Y, et al. Self-inspired learning for denoising live-cell super-resolution microscopy. Nat Methods. 2024; 21: 1895–908.
11. Broaddus C, Krull A, Weigert M, Schmidt U, Myers G. Removing structured noise with self-supervised blind-spot networks. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). Iowa City: IEEE; 2020. p. 159–63.
12. Wu XH, Liu M, Cao Y, Ren DW, Zuo WM. Unpaired learning of deep image denoising. In: European Conference on Computer Vision 2020. Cham: Springer International Publishing; 2020. p. 352–68.
13. Huang T, Li SJ, Jia X, Lu HC, Liu JZ. Neighbor2Neighbor: Self-supervised denoising from single noisy images. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE; 2021. p. 14776–85.
14. Huang X, Fan J, Li L, Liu H, Wu R, Wu Y, et al. Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy. Nat Biotechnol. 2018; 36: 451–9.
15. Zhao W, Zhao S, Li L, Huang X, Xing S, Zhang Y, et al. Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy. Nat Biotechnol. 2022; 40: 606–17.
16. Ingaramo M, York AG, Hoogendoorn E, Postma M, Shroff H, Patterson GH. Richardson–Lucy deconvolution as a general tool for combining images with complementary strengths. Chemphyschem. 2014; 15: 794–800.
17. Hou Y, Wang W, Fu Y, Ge X, Li M, Xi P. Multi-resolution analysis enables fidelity-ensured deconvolution for fluorescence microscopy. eLight. 2024; 4: 14.
18. Godin AG, Lounis B, Cognet L. Super-resolution microscopy approaches for live cell imaging. Biophys J. 2014; 107: 1777–84.
19. Valli J, Garcia-Burgos A, Rooney LM, Vale de Melo E Oliveira B, Duncan RR, Rickman C. Seeing beyond the limit: A guide to choosing the right super-resolution microscopy technique. J Biol Chem. 2021; 297: 100791.
20. Nixon-Abell J, Obara CJ, Weigel AV, Li D, Legant WR, Xu CS, et al. Increased spatiotemporal resolution reveals highly dynamic dense tubular matrices in the peripheral ER. Science. 2016; 354: aaf3928.
21. Valm AM, Cohen S, Legant WR, Melunis J, Hershberg U, Wait E, et al. Applying systems-level spectral imaging and analysis to reveal the organelle interactome. Nature. 2017; 546: 162–7.
22. Buglakova E, Archit A, D'Imprima E, Mahamid J, Pape C, Kreshuk A. Tiling artifacts and trade-offs of feature normalization in the segmentation of large biological images. arXiv. 2025, doi: 10.48550/arXiv.2503.19545.
23. Lin A, Lu AX. Incorporating knowledge of plates in batch normalization improves generalization of deep learning for microscopy images. bioRxiv. 2022, doi: 10.1101/2022.10.14.512286.
24. Nam H, Kim HE. Batch-instance normalization for adaptively style-invariant neural networks. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc; 2018. p. 2563–72.

25. Shinohara RT, Sweeney EM, Goldsmith J, Shiee N, Mateen FJ, Calabresi PA, et al. Statistical normalization techniques for magnetic resonance imaging. NeuroImage Clin. 2014; 6: 9–19.
26. Marston DJ, Slattery SD, Hahn KM, Tsygankov D. Correcting artifacts in ratiometric biosensor imaging; an improved approach for dividing noisy signals. Front Cell Dev Biol. 2021; 9: 685825.
27. Lu M, Christensen CN, Weber JM, Konno T, Läubli NF, Scherer KM, et al. ERnet: A tool for the semantic segmentation and quantitative analysis of endoplasmic reticulum topology. Nat Methods. 2023; 20: 569–79.
28. Sekh AA, Opstad IS, Godtliebsen G, Birgisdottir ÅB, Ahluwalia BS, Agarwal K, et al. Physics-based machine learning for subcellular segmentation in living cells. Nat Mach Intell. 2021; 3: 1071–80.
29. Hugelier S, Colosi PL, Lakadamyali M. Quantitative single-molecule localization microscopy. Annu Rev Biophys. 2023; 52: 139–60.
30. Mazidi H, Ding T, Nehorai A, Lew MD. Quantifying accuracy and heterogeneity in single-molecule super-resolution microscopy. Nat Commun. 2020; 11: 6353.
31. Prabhakaran S, Yapp C, Baker GJ, Beyer J, Chang YH, Creason AL, et al. Addressing persistent challenges in digital image analysis of cancerous tissues. bioRxiv. 2023, doi: 10.1101/2023.07.21.548450.
32. Su F, Wei MP, Sun M, Jiang LX, Dong ZQ, Wang J, et al. Deep learning-based synapse counting and synaptic ultrastructure analysis of electron microscopy images. J Neurosci Meth. 2023; 384: 109750.
33. Askr H, Elgeldawi E, Aboul Ella H, Elshaier YAMM, Gomaa MM, Hassanien AE. Deep learning in drug discovery: An integrative review and future challenges. Artif Intell Rev. 2023; 56: 5975–6037.
34. Qiao C, Li D, Guo Y, Liu C, Jiang T, Dai Q, et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. Nat Methods. 2021; 18: 194–202.
35. Müller CB, Enderlein J. Image scanning microscopy. Phys Rev Lett. 2010; 104: 198101.
36. Hayashi S, Okada Y. Ultrafast superresolution fluorescence imaging with spinning disk confocal microscope optics. Mol Biol Cell. 2015; 26: 1743–51.
37. Lu M, van Tartwijk FW, Lin JQ, Nijenhuis W, Parutto P, Fantham M, et al. The structure and global distribution of the endoplasmic reticulum network are actively regulated by lysosomes. Sci Adv. 2020; 6: eabc7209.
38. Nieuwenhuizen RPJ, Lidke KA, Bates M, Puig DL, Grünwald D, Stallinga S, et al. Measuring image resolution in optical nanoscopy. Nat Methods. 2013; 10: 557–62.
39. Koho S, Tortarolo G, Castello M, Deguchi T, Diaspro A, Vicidomini G. Fourier ring correlation simplifies image restoration in fluorescence microscopy. Nat Commun. 2019; 10: 3103.
40. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer; 2015. p. 234–241.
41. Böhland M, Bruch R, Bäuerle S, Rettenberger L, Reischl M. Improving generative adversarial networks for patch-based unpaired image-to-image translation. IEEE Access. 2023; 11: 127895–906.
42. Reina GA, Panchumarthy R, Thakur SP, Bastidas A, Bakas S. Systematic evaluation of image tiling adverse effects on deep learning semantic segmentation. Front Neurosci. 2020; 14: 1.
43. Fan J, Xia X, Li Y, Dy JG, Wong STC. A quantitative analytic pipeline for evaluating neuronal activities by high throughput synaptic vesicle imaging. NeuroImage. 2012; 62: 2040–54.
44. Verstraelen P, Van Dyck M, Verschuuren M, Kashikar ND, Nuydens R, Timmermans JP, et al. Image-based profiling of synaptic connectivity in primary neuronal cell culture. Front Neurosci. 2018; 12: 389.
45. Wang YL. Noise-induced systematic errors in ratio imaging: Serious artefacts and correction with multi-resolution denoising. J Microsc. 2007; 228: 123–31.
46. Nazari SS, Doyle AD, Bleck CKE, Yamada KM. Long prehensile protrusions can facilitate cancer cell invasion through the basement membrane. Cells. 2023; 12: 2474.
47. Muscarella AM, Dai W, Mitchell PG, Zhang W, Wang H, Jia L, et al. Unique cellular protrusions mediate breast cancer cell migration by tethering to osteogenic cells. npj Breast Cancer. 2020; 6: 42.
48. Meyer AS, Hughes-Alford SK, Kay JE, Castillo A, Wells A, Gertler FB, et al. 2D protrusion but not motility predicts growth factor-induced cancer cell migration in 3D collagen. J Cell Biol. 2012; 197: 721–9.
49. Zhang Y, Zeng R, Xia Y, Han W, Luan Y, Zhang Y, et al. Visualization of protrusion-localized STAT3 mRNA using a self-powered lipidic nanoflare for predicting hepatocellular carcinoma metastasis. Microchim Acta. 2025; 192: 241.
50. Chaudhary S, Moon S, Lu H. Fast, efficient, and accurate neuro-imaging denoising *via* supervised deep learning. Nat Commun. 2022; 13: 5165.
51. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020; 585: 357–62.
52. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011; 12: 2825–30.