Research Article

Gabriel Bon, Daniel Sapède, Cédric Matthews* and Fabrice Daian*

μPIX: leveraging generative AI for enhanced, personalized and sustainable microscopy

https://doi.org/10.1515/mim-2024-0024
Received November 19, 2024; accepted March 25, 2025;
published online April 23, 2025

Abstract: Fluorescence microscopy is a critical tool in biocellular research, enabling the visualization of biological tissues and cellular structures. However, the inevitable aging of microscopes can degrade their performance posing challenges for long-term scientific investigations. In this study, we introduce *μPIX*, a personalized deep learning workflow based on a Generative Adversarial Network (GAN) utilizing a Pix2Pix architecture. The network is trained in a supervised manner to denoise images, optimize preprocessing for binary segmentation, and compensate for equipment aging. Our results, evaluated using standard image quality and binary segmentation metrics, demonstrate that *μPIX* outperforms popular deep learning architectures based on convolutional auto-encoder networks for similar tasks. Additionally, our generative model effectively rejuvenates older detectors to perform on par with newer ones, not only by improving image quality but also by preserving resolution in depth and maintaining a near-linear response between original and generated images in terms of pixel intensity (crucial for quantitative imaging). These findings suggest that generative deep learning approaches can significantly contribute to more sustainable, cost-effective microscopy, fostering continued innovation and discovery in biological research.

Keywords: deep learning; generative AI; image processing; image enhancement; quantitative imaging; sustainable microscopy

***Corresponding authors: Cédric Matthews**, Aix Marseille Univ, CNRS, IBDM, Marseille, France, E-mail: cedric.matthews@univ-amu.fr. https://orcid.org/0009-0004-2338-2969; and **Fabrice Daian**, Aix Marseille Univ, CNRS, LIS, Marseille, France, E-mail: fabrice.daian@lis-lab.fr
Gabriel Bon and Daniel Sapède, Aix Marseille Univ, CNRS, IBDM, Marseille, France

1 Introduction

Microscopy is an indispensable tool in biological research, enabling the visualization of structures at the cellular and molecular levels. However, the performance of microscopes degrades over time leading to diminished signal-to-noise ratio (SNR), reduced resolution and other artifacts that impair image quality. These issues necessitate frequent maintenance that is generally not readily available, hardware upgrade and eventually, replacement of expensive microscopy equipment. In an era where sustainability and cost-effectiveness are paramount, extending the functional lifespan of existing microscopy systems is both economically and environmentally beneficial. The aging of microscopes manifests in several ways, including decreased light throughput, increased background noise, and deteriorating optical alignment. These factors collectively reduce the SNR, making it challenging to discern fine details and thus make accurate quantitative measurements on images of biological samples. Historically, image restoration in fluorescence microscopy relied on techniques like deconvolution and filtering.

Classical methods for image denoising share the common goal of reducing noise while preserving key image features, such as edges and fine details. These approaches, whether operating in the spatial or frequency domain, employ techniques like thresholding, local averaging, or statistical assumptions to identify and mitigate noise. Notable methods includes classical filtering (gaussian, median, low-pass, Wiener,…) and more advanced like Non Local mean denoising (NLM) [1], Total Variation denoising [2], Block-Matching and 3D-filtering (BM3D) [3] and wavelet-based denoising [4]–[6]. These approaches effectively balance noise suppression with the preservation of crucial image structures, ensuring enhanced image quality without compromising essential details. Through spatial or multi-scale analysis, these techniques prioritize the retention of essential structures. However, despite their utility, these methods often fall short in achieving the high performance required for modern, high-resolution imaging tasks, where more advanced approaches are necessary to fully restore and enhance complex images.

Point Spread Function (PSF)-based methods, such as Richardson-Lucy deconvolution, have been widely used to reduce noise and correct blur in microscopy images [7]–[10]. These methods work by reversing the effects of the PSF, which describes how a point source of light is spread out by the optical system. However, they often require precise knowledge of the PSF, which can be challenging to obtain and may vary depending on the imaging system and conditions. Additionally, the computational intensity of these methods can limit their applicability in real-time imaging scenarios, and the PSF information is sometimes expensive or difficult to acquire, particularly when provided by equipment manufacturers [11].

Recent advances have begun to merge traditional PSF-based deconvolution with deep learning techniques. Instead of treating AI and deconvolution as separate approaches, researchers are now leveraging neural networks to accelerate the convergence of classical algorithms or refine their initialization. For instance, deep learning models can predict better initial conditions for deconvolution algorithms, improving both stability and computational efficiency [12]–[14]. Additionally, generative models, such as GANs, have been employed to learn blur kernel distributions, providing a more compact and effective prior for blind image deblurring [15], [16]. Beyond these hybrid approaches, some methods explicitly integrate physics-based imaging model into deep learning frameworks to better simulate real-world degradation and improve reconstruction fidelity. Zhang et al. [17] proposed a confocal imaging degradation model based on confocal imaging theory, enabling the generation of synthetic low-resolution images for training, thus eliminating the need for precise image alignment. Xypakis et al. [18] proposed a physics-informed deep neural network architecture that incorporates the Poisson probability distribution of photo detection into the training loss function, achieving significant SNR improvements for low-exposure microscopy data. These hybrid approaches demonstrate a growing trend where AI enhances rather than replaces conventional image restoration methods, offering a more flexible and data-driven alternative to purely PSF-dependent techniques.

Building on these developments, Deep Learning has emerged as a powerful standalone approach for image restoration, offering end-to-end solutions that learn to directly map degraded images to high-quality outputs. Convolutional neural networks (CNNs), Generative Adversarial Networks (GANs), and more recently Diffusion models have shown promising results in denoising, super-resolution, and artifact removal compared to traditional methods [19]. For example, Content-Aware Image Restoration (CARE) [20] networks use a classical U-Net architecture in the context of a convolutional auto-encoder to restore images by learning from pairs of low-quality and high-quality images, real or synthetic. CellPose [21], [22], a generalist deep learning algorithm offers a robust image restoration capabilities alongside its primary function of cell segmentation. Cellpose3 [23], built on the same architecture, extends its capabilities by enabling joint training of a denoiser and segmenter networks. It includes specialized models for denoising, deblurring and upsampling, tailored for both cytoplasmic and nuclear channels. This flexibility allows it to handle a wide range of image degradation issues and offers an accurate segmentation of cells. Moreover, notable results have been achieved with modified U-Net architecture incorporating Residual Channel Attention Blocks, as seen in RCAN [24], further enhancing restoration performance. These approaches have proven effective in reducing noise and enhancing resolution without requiring PSF information. Noise2Void [25], Noise2Noise [26] and more recently Noise2Fast [27], are notable techniques that further simplify the training process by eliminating the need for training paired data in an unsupervised manner. The W2S framework [28], [29] combines wavelet transformations together with CNNs to enhance both the resolution and the contrast of fluorescent microscopy images. This method leverages multi-scale information to improve image quality. Similarly, the Deep-Z framework [30] enables virtual refocusing in 3D fluorescence microscopy, significantly extending the depth of field and correcting for optical aberrations. High-throughput imaging of 3D samples, such as tumor spheroids and organoids has also benefited from Deep Learning. Techniques that combine axial z-sweep image acquisition with CNN-based restoration allow for faster imaging with reduced photo-toxicity, crucial for live imaging. These methods can generate high-quality 2D projections from low-quality z-sweep images enabling real-time analysis with minimal exposure times.

The advent of diffusion models [31] has revolutionized the field of image generation. They have demonstrated exceptional performance in tasks such as both unconditional image generation [32], image restoration [33], [34], image super-resolution [35], [36] and image denoising [37], [38]. While these models have excelled in terms of image quality metrics, they have largely been trained on large general-purpose datasets, with few specifically tailored to address the unique challenges of microscopy image modalities. However, recent efforts have focused on adapting diffusion models for microscopy tasks, including synthetic dataset generation [39], super-resolution in optical microscopy [40], and EMDiffuse [41], a model

specifically trained on electron microscopy (EM) images, which has shown promising results in denoising and resolution enhancement tasks for electron microscopy imaging.

Despite these advancements, challenges remain in generalizing deep learning models across different imaging conditions and microscopy setups. Our proposed workflow leverages these advances to address both immediate and long-term challenges in microscopy image restoration. In contrast to the current state-of-the-art model based on UNET architectures [42], our *μPIX* workflow leverages the use of a conditional generative adversarial network (cGAN) [43], more specifically the use of Pix2Pix network [44] to efficiently tackle the classical denoising and segmentation problematics. The conditional aspect of such a network ensures that the model is trained on paired image data, enforcing a direct mapping between noisy and clean images to generate accurately denoised outputs. Moreover, our approach allows us to tailor a highly specialized model by implementing a precision-focused strategy ensuring that our enhancements are optimally aligned with the specific deficiency and operational context of the equipment leading to superior image restoration, extended utility of the microscope and ultimately doing quantitative biology.

2 Results

2.1 *μPIX* is built on a generative Pix2Pix architecture

Image denoising is known to be one of the major problems in the field of image analysis and deep Learning based solutions have proven their superior capabilities in this task in comparison to traditional denoising algorithms. As of today, the architecture based on convolutional autoencoder are considered as the state of the art to tackle this problem. UNET and similar convolutional autoencoders operate primarily through pixel-wise predictions, optimizing pixel-level accuracy using reconstruction loss functions. While this approach ensures a good and fast overall image reconstruction, it can struggle with preserving fine details and textures, especially in denoising tasks. In this context, we first based our approach on the use of a classical UNET network working in combination with a classical pixel-wise loss (mean squared error – MSE) and a perceptual loss network (VGG16) to make our network aware of high level perceptual and semantic differences between original and predicted images. Unfortunately, even if we improved slightly in image reconstruction quality, in comparison to classical UNET networks used for such tasks, the results were not satisfying (Supplementary Table 4). To address these limitations, we chose to base our workflow on Pix2Pix generative network (Figure 1). Unlike UNET, *μPIX* leverages the cGAN architecture to produce high-quality and realistic denoised images. This choice allows us to train a model that is not only capable of tackling classical image challenges like denoising but also able to address the specific defects of particular hardware. By leveraging the flexibility of the Pix2Pix architecture, *μPIX* can adapt to the unique characteristics and imperfections of specific imaging equipment, effectively building a specialized prosthesis for each device. This adaptability ensures that our model can provide optimized solutions tailored to the nuances of different hardware, enhancing overall performance and image quality in a way that traditional convolutional autoencoders cannot

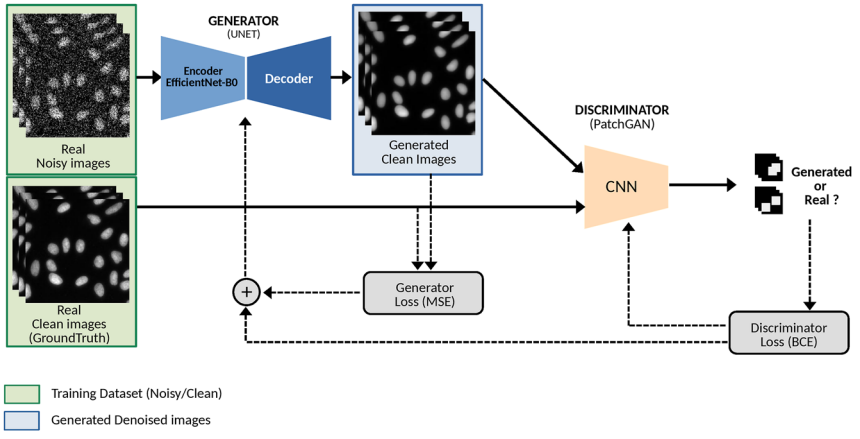


Figure 1: *μPIX* architecture is based on a Pix2Pix generative network. *μPIX* consists of two subnetworks: a generator, based on a UNet architecture with an EfficientNet-b0 backbone, and a discriminator (PatchGAN). During supervised training, a noisy image is input to the generator, which generate an image. This output is compared to the real clean image using a pixel-wise loss function (MSE). Pairs of real and generated images are then passed to the discriminator, which classifies them as real or fake using a binary cross-entropy loss (BCE). Both subnetworks are progressively refined through adversarial loss during training. In the inference phase, only the trained generator is used to generate clean images.

achieve. Our network introduces an adversarial loss that encourages the network to generate images that are not only accurate but also perceptually convincing and realistic. This adversarial training helps preserve fine details and textures that are often lost in pixel-wise approaches. The network is composed of two subnetworks working together: a generator and a discriminator. The generator network is trained to learn how to transform an input image to resemble a reference image provided during the supervised training. Classically this subnetwork architecture is based on a classical UNET network. The encoding part of the generator has been chosen carefully as we wanted to keep a good compromise between training/inference speed and performance. We decided to choose a lightweight and performant EfficientNet-b0 [45] as the backbone for this subnetwork. To ensure that the training phase will be efficient and will not collapse quickly, we decided to not use a pre-trained version of this backbone to avoid at start a too big gap between generator and discriminator capabilities. The discriminator subnetwork is based on a PatchGAN network derived from a classical convolutional neural network which has proven its superior discrimination capabilities in such architectures [46]. The main objective of the discriminator is to assess whether the given images as input are generated or real images. Through adversarial training [47], in a supervised manner, the generator and discriminator iteratively improve their accuracy, leading to the generator becoming increasingly accurate at producing images that closely resemble the reference images. At the end of these steps, the inference phase will only use the generator subnetwork for image generation.

2.2 *μPIX* outperforms state-of-the-art denoisers on CSBDeep denoising benchmark dataset

Image denoising and restoration are central challenges in the analysis of microscopy data. To assess the accuracy of our workflow, we chose to use the CSBDeep Denoising Dataset as a benchmark dataset to evaluate our approach [20]. This dataset, derived from the Broad Bioimage Benchmark Collection [48], comprises pairs of clean reference images containing cell nuclei and their corresponding artificially noised counterparts from the human U20S cell line. The noising involves synthetically adding significant readout and shot noise, along with additional 2×2 pixel binning to mimic acquisitions at very low light levels. To evaluate our workflow comprehensively, we used a set of denoising and image restoration quality metrics. We employed the mean-squared error (MSE) as a measure of signal-to-noise ratio by comparing the reference clean images to the images generated by *μPIX* from the noisy image. The Structural Similarity Index (SSIM) was used to measure the preservation of overall object shapes.

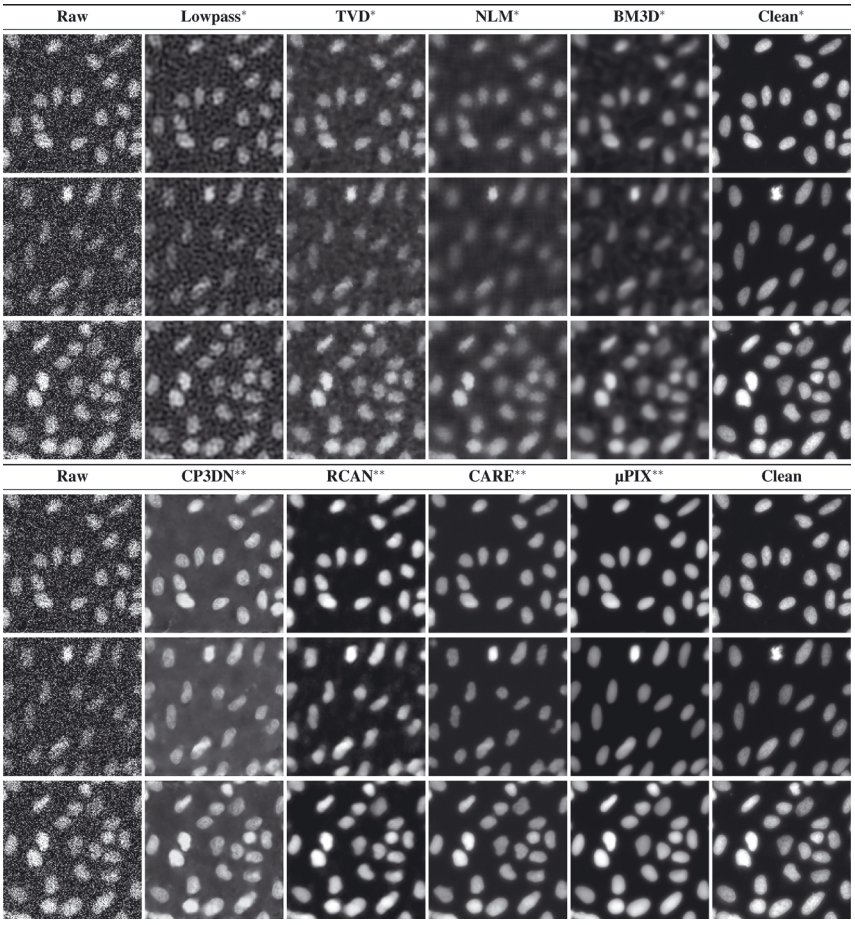
We compared our approach, μPIX, against a range of image denoising algorithms, including both traditional non-deep learning-based methods and deep learning-based models. The non-deep learning-based methods include Low-Pass filtering, TVD [2], NLM [1], and BM3D [3]. For deep learning-based approaches, we evaluated our μPIX alongside CARE [20], the Residual Channel Attention Network (RCAN) [24], and the "Denoise Nuclei" model from CellPose3 [23], which are all based on UNET-like architectures. To ensure a fair comparison and minimize bias in generalization, we trained both CARE, RCAN and *μPIX* from scratch using the same dataset. However, for the Denoise Nuclei model from CellPose3, we used the provided pre-trained version, as there is currently no publicly available method to train this denoiser from scratch on a custom dataset.

Finally, we evaluated these architectures and our approach concurrently on the same test dataset consisting of various nuclei images extracted from the original benchmark dataset and we averaged the metrics. As shown in (Table 1) and (Table 2), our approach clearly outperforms all compared methods, including both non-deep learning-based and deep learning-based denoising algorithms, in terms of signal-to-noise ratio and structural preservation. For the best algorithms, CARE, we improved MSE by 51 % and SSIM by almost 4 % (MSE: 117.68, SSIM: 0.94). This

Table 1: MSE and SSIM metrics for evaluating the quality of denoised images. Mean Squared Error (MSE) and Structural Similarity Index (SSIM) are used to assess the quality of denoised images. The performance of *μPIX* is compared against both traditional non-deep learning-based methods (Low-Pass Filter, Total Variation Denoising (TVD), Non-Local Means (NLM), and Block Matching 3D (BM3D)) and deep learning-based methods (CellPose3 Denoise Nuclei, Residual Channel Attention Network (RCAN), and CARE). MSE and SSIM are computed between the noisy and denoised images from the CSBDeep Denoising Benchmark Dataset test set. For reference, the "Raw" row represents the metrics between the original noisy image and the clean (ground truth) image. Bold values means best scores in terms of MSE and SSIM.

	MSE	SSIM
Raw	2,434.03	0.491
Lowpass filter	887.64	0.688
TVD	749.18	0.733
NLM	741.06	0.742
BM3D	509.48	0.834
Denoise Nuclei (Cellpose3)	1,457.07	0.656
RCAN	505.86	0.825
CARE	241.26	0.90
μPIX	**117.68**	**0.948**

Table 2: Denoising results on the CSBDeep Benchmark Dataset, including *μPIX*. *Traditional non-deep learning-based denoising algorithms. **Deep learning-based denoising algorithms, including our proposed method *μPIX*.



superiority can be attributed to several factors: the adversarial process helps the generator produce more realistic and high-quality images, the conditioning provides more context-aware denoising capabilities, and the adversarial loss, compared to pixel-wise loss, encourages the generation of sharper and more structurally accurate images.

2.3 *μPIX* improves binary segmentation quality when used as an image denoiser in the preprocessing steps

Object segmentation is another fundamental task in image analysis and a good preprocessing of images can greatly enhance further binary segmentation performance. We then wanted to assess how *μPIX* performs as the main image denoiser in a segmentation workflow (Figure 2). We chose to evaluate *μPIX* against both CARE and Cellpose3 "Denoise Nuclei" as preprocessing steps to enhance image segmentation. As we did not have segmentation ground truth included in the CSBDeep Denoising Dataset and given its robustness and superior performance in segmenting nuclei, we relied on Stardist [49] as image segmenter, using its pre-trained model "2D_versatile_fluo" to infer binary segmentation from clean images which served as the reference for perfect segmentation. Using the same test dataset previously described, we assessed the impact of different preprocessing methods on binary segmentation performance using classical segmentation metrics: Intersection Over Union (IoU), Precision, Recall, and F1-score. IoU measures the overlap between predicted and true segmentations, offering a direct measurement of segmentation accuracy. In the context of binary segmentation, Precision and Recall are crucial for understanding the balance between over-segmentation and under-segmentation. Precision indicates the proportion of true positive results among all positive predictions, thereby reflecting the extent of over-segmentation due to false positives. Recall indicates the proportion of true positives among all actual positives,

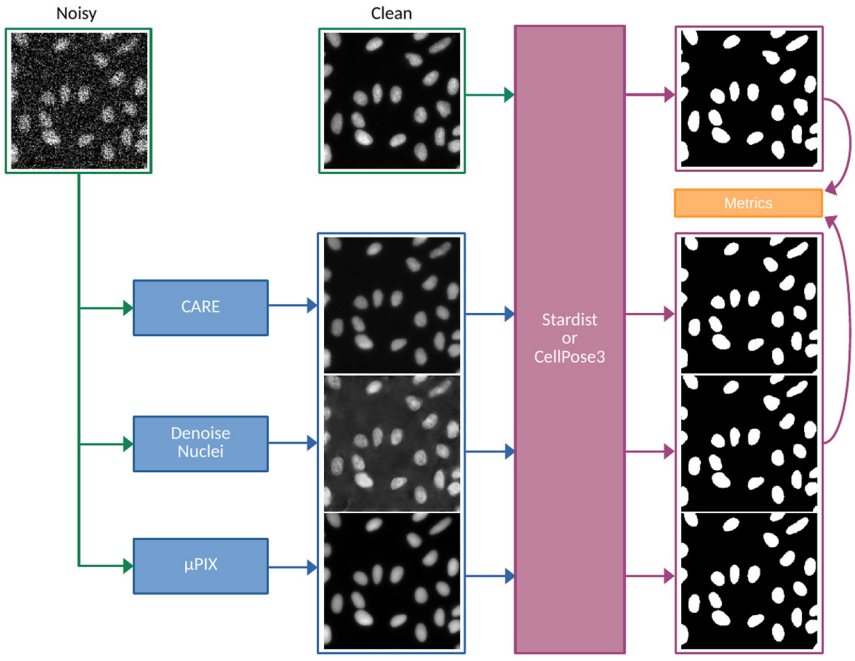


Figure 2: Workflow describing the evaluation process for denoising performance in binary segmentation. To assess the quality of binary segmentation after image denoising by various softwares, pairs of noisy and clean images are used. The noisy images are processed through each of the denoising algorithms: CARE, Denoise Nuclei, and *μPIX* (in blue). The denoised images are then fed into either the Stardist or CellPose3 segmentation models (purple). In parallel, the clean image is also processed by these segmentation models (purple). The resulting binary mask from the clean image serves as the ground truth and is compared with the binary masks generated from the denoised images (orange). The IoU, Precision, Recall, and F1-score metrics are then calculated for performance evaluation.

highlighting the degree of under-segmentation caused by false negatives. The F1-score, as the harmonic mean of Precision and Recall, provides a single metric that balances both aspects, ensuring a robust evaluation of binary segmentation performance. As shown in (Table 3) and (Table 4), using *μPIX* as a preprocessing step for segmentation outperforms the other methods across all metrics, yielding the best results in terms of IoU (0.861), Recall (0.903), and F1-score (0.9253) on the test dataset. Although CARE achieved the highest Precision score (0.9537), *μPIX* maintained a strong balance with its high Recall. We then assessed whether our approach could be generalized to other state-of-the-art segmentation tools. We used the Cellpose3 segmenter as a reference and employed CARE, Cellpose3 "Denoise Nuclei", and *μPIX* as preprocessing steps to benchmark segmentation quality using the same metrics and test dataset. As shown in (Table 3) and (Table 4), *μPIX* yielded again the best overall results across the defined segmentation metrics (IoU: 0.861, Recall: 0.903, and F1-score: 0.9253). Additionally, switching from Cellpose3 "Denoise Nuclei" to *μPIX* within the Cellpose3 workflow improved the F1-score by nearly 3 % meaning that the actual CellPose3 workflow could be improved by using a trained *μPIX* as denoiser. Consequently, *μPIX* is not only a strong candidate for pure image denoising but also enhances performance when used in conjunction with state-of-the-art models in the context of binary image segmentation.

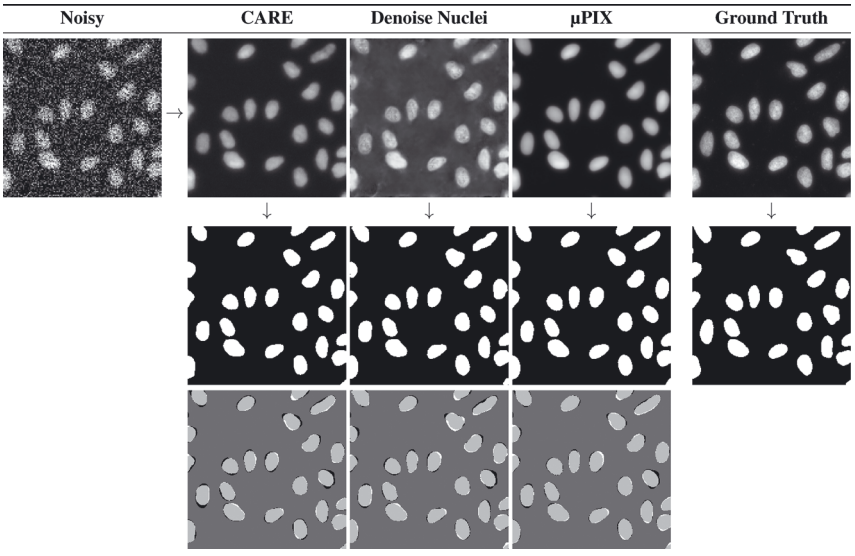
2.4 *μPIX* enables effective rejuvenation of microscope detectors

A major challenge faced by all microscopy platforms is the aging of equipment, which inevitably introduces acquisition artifacts, making it increasingly complex to analyze acquired data. To demonstrate *μPIX* capabilities toward the challenge of hardware senescence, we chose to simulate one specific case: *detector rejuvenation*. To this end, since we based our architecture on a supervised approach, we chose to simulate detector degradation by building a dedicated dataset consisting of pairs of images acquired simultaneously using an older Multi-Alkali detector [50] and its newer and more performant counterpart using a GaAsP detector [51]. We acquired this tailored dataset using a biphoton system able to generate two images at the same position within the sample as illustrated in (Figure 3). This microscopy setup freed us from the tedious steps of image re-alignment during the acquisition and the dataset preprocessing steps. Furthermore, it allowed us to maintain

Table 3: Segmentation metrics using Stardist and CellPose3 as Segmenter. Noisy images from the CSBDeep Benchmark Dataset were denoised using CARE, Denoise Nuclei, and μPIX. The denoised images were then segmented using either StarDist or CellPose3. Segmentation metrics (IoU, Precision, Recall, F1-score) were calculated by comparing the binary segmentations of clean and denoised images for both StarDist and CellPose3. Bold values means best scores in terms of MSE and SSIM.

		IoU	Precision	Recall	F1-score
Stardist	Raw	0.7441	0.9226	0.7936	0.8532
	CARE (CSBDeep)	0.7895	**0.9647**	0.813	0.8824
	Denoise Nuclei (Cellpose3)	0.801	0.9622	0.8273	0.8897
	μPIX	**0.8471**	0.9597	**0.8783**	**0.9172**
CellPose3	Raw	0.7342	0.8967	0.802	0.8467
	CARE (CSBDeep)	0.8032	**0.9537**	0.8358	0.8908
	Denoise Nuclei (Cellpose3)	0.8167	0.9524	0.8514	0.8991
	μPIX	**0.861**	0.9488	**0.903**	**0.9253**

Table 4: Binary segmentation results on the CSBDeep Benchmark Dataset. (Top) Three noisy input examples from the CSBDeep Dataset, along with their denoised counterparts generated by CARE, Denoise Nuclei, and μPIX, followed by the ground truth. (Middle) Binary segmentation results alongside the ground truth mask. (Bottom) Visual representation of segmentation differences: in white false positive, in black false negative and in light gray true positive and dark gray true negative.



a supervised learning context mandatory for our Pix2Pix architecture. As we wanted to be as close as real use cases, we decided to use biological samples consisting of Gastruloids [52]. We acquired two complete stacks, one serving as a training/validation set and the other one as a test set (see Methods). We trained μPIX from scratch on this training dataset taking as input the image acquired with the Multi-Alkali detector and considering the corresponding image acquired with the GaAsP detector as ground truth. To conclude whether the detector rejuvenation was effective, we decided to check for three different features on image preservation: image quality, intensity preservation along the Z-axis, linear preservation of pixel intensity level between the original and rejuvenated detector. We used MSE and SSIM metrics between images acquired with GaAsP detector and predicted images to assess the quality of the detector rejuvenation. We chose to compare our approach only to CARE because, as of today, there is no published way to train or fine tune a Cellpose3 "denoise nuclei" model on its own dataset. As shown in (Table 5) and (Supplementary Table 1), our approach greatly improves SNR and the quality of structure restoration. Moreover, we can see that μPIX successfully handled challenging restoration tasks, such as delineating detected objects, managing complex structures, addressing intense contrast variations (both low and high) and reconstructing objects even in cases where limited information was available. These results demonstrate that our approach is effective and reliable for

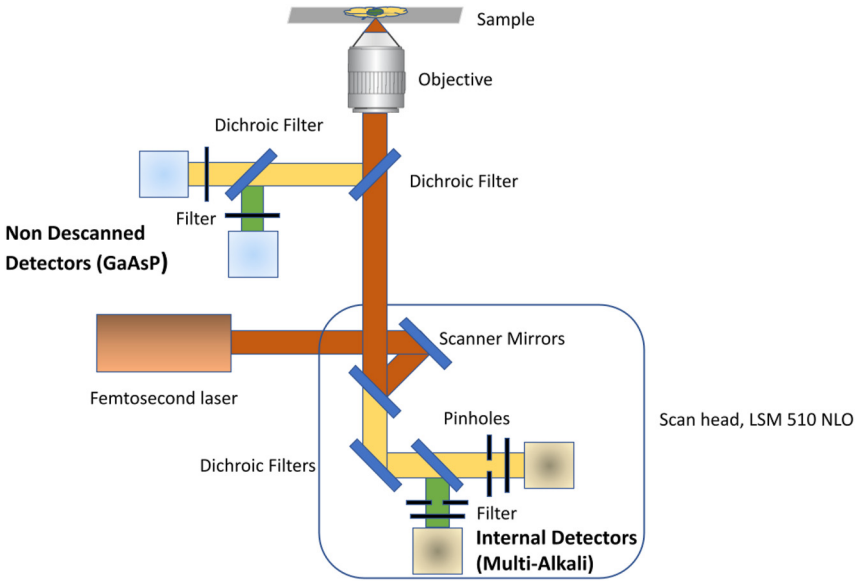
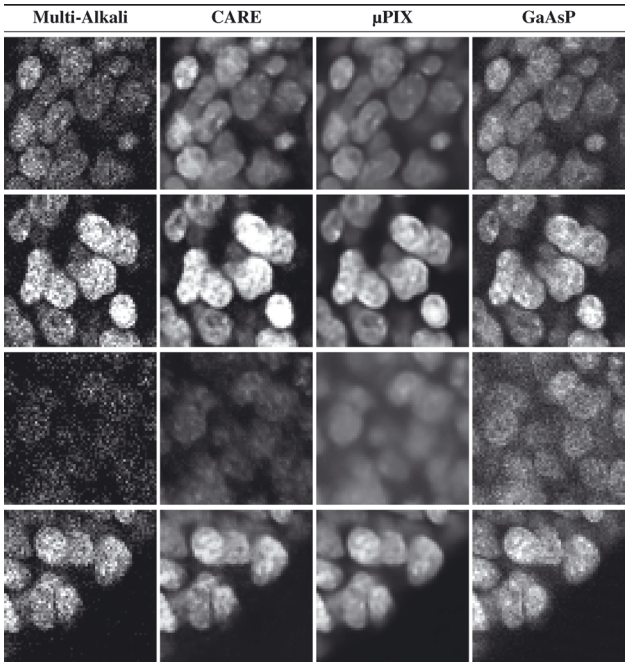


Figure 3: Schematic of a two-photon microscope optical path. Shown here are the two detection channels for the fluorescence signal generated by two-photon absorption. The Non Descanned channel, positioned closest to the objective, is the most sensitive. The GaAsP detectors are new and provide a reference signal optimized for learning. The detection path through the scan head is generally less efficient, even if the pinholes are completely open to collect the scattered emission photons. Multi-Alkali detectors are functional but obsolete, over 15 years old, and will be used to detect ground truth. The signals detected on both types of detectors are almost spatially aligned.

Table 5: Denoising results on Microscope Rejuvenation Test Dataset. Images acquired using the Multi-Alkali detector from different regions with varying pixel intensities were processed by CARE and μPIX, and then compared to the original images obtained with the GaAsP detector.



enhancing image quality by compensating the Multi-Alkali detector aging in generating images resembling to thoses acquired with a GaAsP detector.

Moreover, as it is well known that the signal quality diminishes along the Z-axis during a confocal acquisition due to light scattering, light absorption, refractive

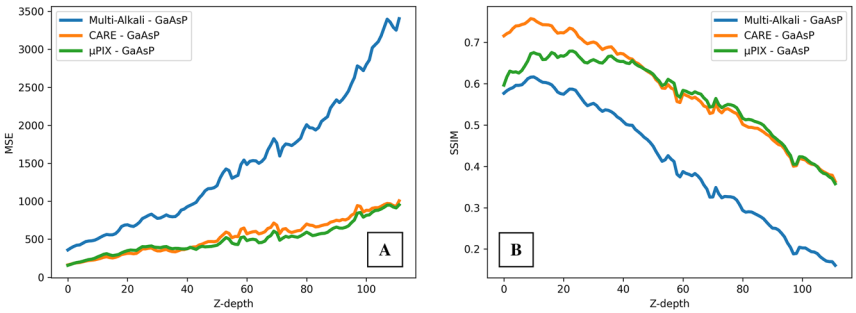


Figure 4: Evaluation of intensity and structural preservation along the *Z*-axis using the Microscope Rejuvenation Test stack. (A) MSE between Multi-Alkali and GaAsP (blue), CARE and GaAsP (orange), *μPIX* and GaAsP (green) (B) SSIM between Multi-Alkali and GaAsP (blue), CARE and GaAsP (orange), *μPIX* and GaAsP (green).

index mismatch, photo-bleaching and detector sensitivity, we wanted to assess how *μPIX* compensates for these artifacts. To do so, we started by measuring these effects by comparing the MSE between an older Multi-Alkali and a newer GaAsP detector for every slice along the *Z*-axis of the test stack. As shown in (Figure 4A), there is a quasi-exponential difference in terms of MSE as we go deeper into the stack along the *Z*-axis. This is explainable by the fact that Multi-Alkali detectors are more prone to the signal intensity attenuation effect compared to the newer GaAsP detectors. We then compared our predicted images to GaAsP and we see that the difference is far more prone than before and quasi-linear in terms of MSE and SSIM (Figure 4B) as we go down into the stack along the *Z*-axis. This means that even if our approach is not able to abolish completely these artifacts, our results in terms of signal intensity are very close and resemble those we would have obtained using GaAsP detectors (Supplementary Table 5).

We next aimed to determine whether the images generated by *μPIX* could be reliably used for quantitative imaging analysis. To do this, we evaluated whether *μPIX* maintains a consistent response at different levels of pixel intensity when compared to images acquired using GaAsP detectors. As shown in (Figure 5A), pixels ranging from 0 to nearly 60 % of maximum intensity (representing approximately 95 % of the total pixels in the test images) are restored with near-perfect linearity by *μPIX*. In contrast (Figure 5B), shows that this consistency is not maintained by CARE. For pixel intensities greater than 60 % and the maximum (representing around 5 % of the image pixels), we observe moderate deviations from perfect restoration. However, this concerns only a small fraction of the pixels, most of which are saturated and therefore contain limited or non-informative data, making these deviations negligible for most practical purposes.

These results suggest that using *μPIX* for detector rejuvenation opens the possibility for users to conduct quantitative imaging in the same way as if they were using a GaAsP detector directly.

3 Discussion

A common concern among microscope users is the validity of AI generated images, as these are synthetic and may not be suitable for further analysis. The rise of generative AI has

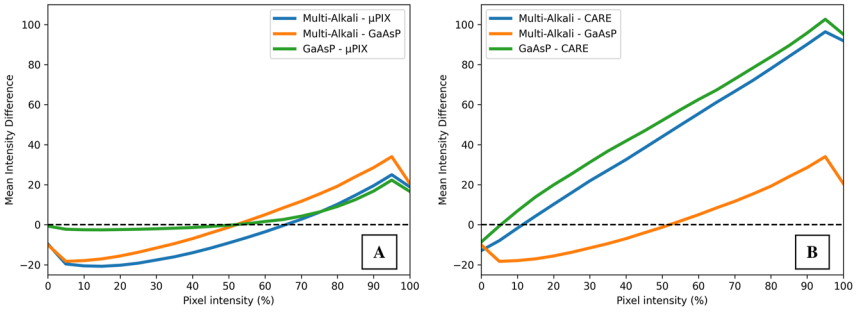


Figure 5: Evaluation of signal linearity preservation using the Microscope Rejuvenation Test Dataset. (A) Mean intensity differences for pixel intensity ranging from 0 (no intensity) to 100 (maximum intensity) between Multi-Alkali and *μPIX* (blue), Multi-Alkali and GaAsP (orange), GaAsP and *μPIX* (green). The dashed line corespond to a perfect linearity. (B) Mean intensity differences for pixel intensity ranging from 0 (no intensity) to 100 (maximum intensity) between Multi-Alkali and CARE (blue), Multi-Alkali and GaAsP (orange), GaAsP and CARE (green). The dashed line corresponds to a perfect linearity.

underscored the need to expand the family of perceptual metrics, focusing on human perception to validate such images. The two most widely used perceptual metrics are the Frechet Inception Distance (FID) and the Inception Score (IS) [53], [54]. These metrics utilize a pre-trained Inception network on the ImageNet 10 k dataset [55] and measure the Wasserstein distance and KL-divergence, respectively, between the embeddings of real and generated images. While these metrics are effective for well-structured images, they may be problematic in the context of image denoising. They tend to prioritize structural conservation over the preservation of overall distribution. Moreover, since these metrics rely on features extracted from a network pre-trained on a generalist images database, they may not be suitable for microscopic images. Microscopic images often contain subtle artifacts that may not align well with the features learned by the Inception model. To our knowledge, there is no perceptual metric specifically adapted to the nature of microscopic images that can effectively capture their unique feature space and accurately measure their perceptual quality. In line with previous attempts to assess image quality through human evaluation [56], we decided to adopt a similar approach. We designed an experiment involving 27 participants. These individuals were familiar with biological imaging but had no specific knowledge of the image categories or the methods used to generate them. Each participant completed seven rounds of evaluation, with each round consisting of four images: two generated by algorithms (μPIX and CARE) and two acquired through real detectors (Multi-Alkali and GaAsP), representing four distinct image categories. For each round, participants were asked to rate each image on a scale from 1 to 4, where 1 indicated the image was highly suitable for analysis, and 4 indicated it was completely unsuitable. The ratings for each image category were collected and their distributions are shown in (Figure 6). We conducted a statistical analysis between categories using the Wilcoxon non-parametric test. The evaluation results revealed that, as expected, a newer detector yields higher perceived image quality, as evidenced by a statistically significant distinction in human perception between images acquired with the older Multi-Alkali detector and those from the newer GaAsP detector ($p-\text{value} = 7.87 \times 10^{-7}$). We then compared the older Multi-Alkali images to both the CARE and μPIX generated images. Interestingly, in terms of human perception, the Multi-Alkali images and those enhanced by CARE were not judged to be significantly different, indicating that the enhancements applied by CARE do not improve the perceived usability of the images for analysis ($p-\text{value} = 0.6151$). In contrast, μPIX images were perceived as more usable for analysis when compared to Multi-Alkali ($p-\text{value} = 6.63 \times 10^{-10}$) and remarkably as more usable than those obtained from GaAsP detectors ($p-\text{value} = 7.23 \times 10^{-3}$). From a broader perspective, it appears that, despite its effectiveness in enhancing images, CARE (built on a convolutional autoencoder architecture) fails to convince users of its full

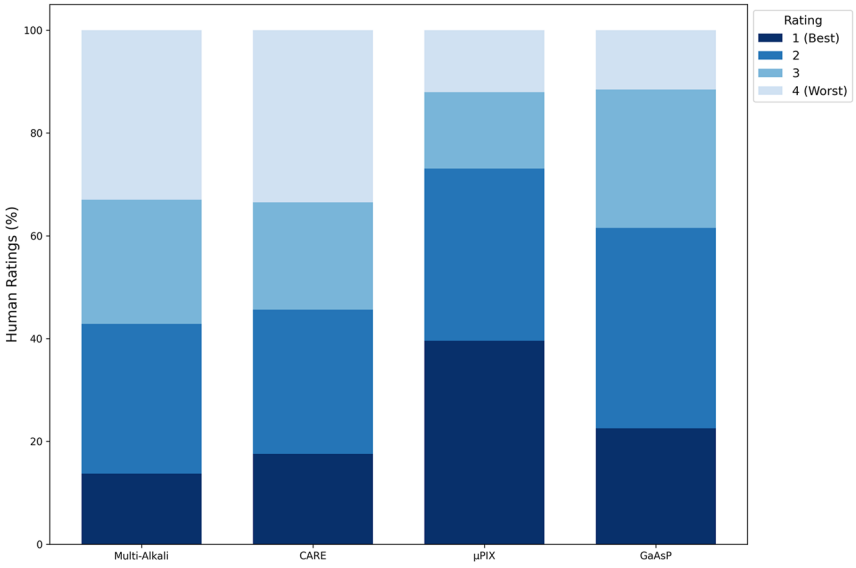


Figure 6: Evaluation of image quality by human observers. Stacked bar plot illustrating the distribution of human ratings for four image acquisition methods: Multi-Alkali, CARE, μPIX, and GaAsP. Each stacked bar represents the percentage of total ratings for each method, with the sum of ratings normalized to 100 %. The ratings range from 1 (best: dark blue) to 4 (worst: light blue). The distribution reflects the overall quality perception across the different image types based on human evaluations.

applicability for image analysis. This is likely due to the known tendency of such networks to introduce blurring and smooth out structures [57]–[59]. In contrast, generative approaches based on Pix2Pix networks aim to maintain the visual coherence of textures and structures while reducing noise. This leads to more realistic images that are better evaluated by human observers, and perceived similarly to real images captured by more advanced detectors.

Currently, the predominant paradigm for using Deep Learning in image treatment and analysis revolves around the use of U-Net networks, as seen in tools like CARE, CellPose3 or RCAN. The main advantages of these networks include their ease of training and deployment by non-specialist users, thanks to their training stability and the relatively short training time required to achieve good-quality results. This user-centered approach allows anyone to efficiently train their own model from scratch using their own dataset. However, we consider this approach unsustainable as it operates at the end of the imaging pipeline without addressing upstream factors such as hardware defects and specificities. Our *μPIX* workflow proposes an alternative paradigm, shifting from a user-centered to a hardware-centered perspective for model development and training. We envision that such a workflow should be developed over the long term by microscopy platforms to directly address hardware constraints and gradually build robust and flexible models for personalized microscopy. While it is well known that training GANs has disadvantages compared to U-Nets such as longer training times (ranging from hours to days), training instability that can lead to mode collapse, and the challenge of determining an optimal stopping criterion; the results we present in terms of image denoising, segmentation, and hardware rejuvenation suggest that investing effort in developing models tailored to specific hardware by a deep learning specialist could be more beneficial. These models could then be shared by many users, as opposed to the current practice of many users developing rather identical models for the same hardware. In the era of *frugal AI*, we believe that our approach has the potential to save significant resources, both in terms of user time and global computational costs. Additionally, it is important to recognize that the ground truth datasets provided to *μPIX* are not of infinite quality and do not encompass every possible microscope artifact requiring correction. As the pix2pix architecture is based on a supervised learning approach, the system is inherently limited by the quality and variety of the images included in the training set. Consequently, *μPIX* cannot generate images of higher quality than those contained in the dataset, implying that it does not effectively correct hardware defects but rather learns to replicate them, producing images that closely resemble the ground truth provided.

Our approach opens up a wide range of possibilities for enhancing microscopy acquisition workflows and setups. Live imaging applications that require high prediction rates could particularly benefit from the use of *μPIX*. A major limitation of live imaging is phototoxicity, caused by the prolonged use of lasers. By utilizing a *μPIX* model pre-trained on a custom dataset with varying laser intensities, laser power can be reduced, potentially extending the lifespan of the samples being imaged and improving the image SNR. While the adversarial training process of Pix2Pix networks is time consuming, the inference stage is relatively fast, as it only involves the network generator part, a lightweight UNET network. Benchmarking performed on an Nvidia A6000 GPU with 48 GB of VRAM revealed that *μPIX* can process 473 images of size 256×256 per second, 119 images of size 512×512, and 31 images of size 1024×1024. These results demonstrate that real-time image correction during live imaging sessions is highly feasible, suggesting that such models can be seamlessly integrated into existing microscope setups. We have also begun exploring the application of *μPIX* for post-acquisition correction of temperature and oil refractive index mismatches.

We also explored whether an unsupervised approach could yield better results. The primary limitation of *μPIX* stems from the fact that Pix2Pix relies on a supervised learning architecture, which requires paired ground truth images registered in the same position for training. To address this, we investigated the use of *CycleGAN* architectures, which only require unpaired images under different conditions, which makers it easier to construct a training dataset. While the results were acceptable, the image quality was noticeably lower compared to what we achieved with Pix2Pix. We believe that this limitation could be mitigated by developing a more robust and diverse dataset (Supplementary Table 2).

While emerging models such as diffusion networks and transformer architectures, particularly Visual Transformers (ViT), hold great promise for image generation and vision-related tasks, we remain cautious about applying diffusion models in our context. The main challenge is that state-of-the-art diffusion models are typically trained on large, general-purpose datasets with millions of images, which does not align with the specialized datasets used in microscopy. Furthermore, these models are rarely trained in a supervised manner. Another concern is that for real-time denoising applications, such as live imaging, the image generation process in diffusion models is too slow due to the sampling cycles required during inference. Despite these

limitations, we investigated EM Diffuse [41], a supervised diffusion model developed for electron microscopy, which we retrained on our own datasets. We provide preliminary results as a point of comparison (Supplementary Table 3).

In conclusion, our work introduces *μPIX* as an innovative solution to the pressing challenges of microscopy image denoising and restoration, effectively tackling issues arising from aging hardware and acquisition artifacts. Through rigorous comparisons with established denoising methods either non or deep learning-based, we have demonstrated that *μPIX* significantly enhances image quality, preserves structural integrity, and improves object segmentation accuracy. Our findings indicate that *μPIX* excels in rejuvenating images captured by older detectors, effectively bridging the gap to newer technologies while compensating for artifacts associated with light absorption along the *Z*-axis and maintaining a quasi-linear relationship in pixel intensity between the original and rejuvenated detector. By adopting a hardware-centered paradigm for model development, we emphasize the importance of creating specialized solutions tailored to specific microscopy setups, promoting a sustainable approach to image analysis that reduces computational burdens on users. Furthermore, our research underscores the transformative potential of *μPIX* in microscopy workflows, enabling researchers to confidently analyze synthetic images that closely mimic those obtained from advanced detectors. This work paves the way for future advancements and developments in generative AI for microscopy platforms, ultimately supporting the pursuit of more accurate and reliable imaging outcomes.

4 Materials and methods

4.1 Microscope setups

Fluorescence imaging was performed on a Zeiss LSM510 confocal scanning microscope (CLSM) equipped for two-photon imaging. The excitation laser is a Spectra Physics Mai Tai infrared tunable laser, used at 900 nm for the scope of these experiments. The scan head is mounted on a Zeiss Axiovert200M stand. We used a Plan-Apochromat 20×/0.75 Numerical Aperture objective. The room is well stabilized and controlled at a temperature of 21 °C (±1 °C). The stand and sample environment was isolated in a black-painted incubation chamber, providing light isolation to prevent external signal pollution. The heating unit is not on, but the presence of the chamber still further helps with the temperature stability of the sample environment. Two beam paths were exploited for fluorescence detection. First, the more efficient non-descanned (NDD) beam path using two specially integrated Hamamatsu GaAsP detectors (new integration work in 2021 by ALPhANOV company). The detection range is set by filters from [500/550] nm for the *green* channel (the *red* channel was acquired but not used for the training). Second is the descanned configuration, using the internal Multi-Alkali PMT old detector of the scan head after the pinhole was fully opened. Filter sets of the internal beam path are chosen to fit a detection range as close as possible to the range of the NDD configuration of [500/550] nm. The system driving the acquisition software is Zeiss Efficient Navigation (ZEN) version 2009. To generate data without acquiring the same area, we use the tiling option with the overlap set to 0 %.

4.2 Two photons imaging

One-color two-photon imaging of immunostained samples was performed on an inverted Zeiss LSM510 confocal as described above. Multiposition imaging was used to automatically acquire image *Z* stacks on multiple gastruloids mounted on the same sample slide. The sampling parameters remain the same for all samples (pixel width 0.62 μm, voxel depth 1.2 μm), with 114 identical sections acquired sequentially on the GaAsP and Multi-Alkali detectors, reaching a depth in *Z* of 137 μm. The images were acquired with the full field-of-view (318 μm). The power of laser excitation and gains of detectors are optimized to exploit detector dynamics (8 bits) while avoiding any saturated pixels.

4.3 Microscope rejuvenation dataset

4.3.1 Sample preparation

Gastruloids were generated using the protocol described previously in [52], from a H2B-GFP mouse embryonic stem cells line (a generous gift from Kat Hadjantonakis). Briefly, 200 cells were seeded and aggregated for 48 h in low-adherence 96-well plates (Costar ref: 7007) and subsequently pulsed with the Wnt agonist Chiron, which was washed out after 24 h, i.e. at 72 h of aggregate culture. We imaged 96 h old gastruloids, which exhibited polarized morphologies.

4.3.2 Dataset construction

As a result, we acquired 10 stacks of size 512 × 512 pixels with depth varying from 22 to 130 slices. Among those, we used 8 stacks to train/validate our model and two stacks to test the model. For the preparation of training and validation data, we tiled images into smaller regions of size 256 × 256 pixels with an overlap of 64 pixels to ensure

comprehensive coverage. Employing a *reflect mode* for tile border padding helped mitigate any potential blank spaces resulting from the overlapping procedure. Furthermore, to standardize image intensity distributions, we applied percentile normalization between 1 and 99 %, resulting in pixel values ranging from −1 to 1. During the training phase, we implemented data augmentation using the Albumentations [60] Python library. We decided to include transformations consistent with the biological objects used. We then used *shift scale rotate*, *elastic transformation*, *optical distorsion*, *randomrotate90* and *horizontal/vertical flip*.

4.4 Objectives

4.4.1 GAN objectives

GANs are generative models that learn a mapping from a random noise vector z to an output image y using a generator network G, which can be either an encoder-decoder or a U-Net:

$$G(z) \to y$$

The generator G is trained to produce outputs that are indistinguishable from real images by an adversarially trained discriminator D, which aims to correctly distinguish real images from generated ones. The standard GAN objective is:

$$\mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}_y[\log\, D(y)] + \mathbb{E}_z[\log(1 - D(G(z)))].$$

In contrast, conditional GANs (cGANs) learn to map an observed condition x and a random noise vector z to a corresponding output y:

$$G(x, z) \to y$$

The cGAN objective is given by:

$$\begin{aligned}\mathcal{L}_{\text{cGAN}}(G, D) = {} & \mathbb{E}_{x,y}[\log\, D(x, y)] \\ & + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]\end{aligned}$$

Pix2Pix is a special case of cGANs where the goal is to condition the output image to resemble the input image as closely as possible, eliminating the need for the noise vector z. In this case, the generator learns a direct mapping from the input image to the output image:

$$G(x) \to y$$

To enhance the similarity between the generated image and the ground truth, a L_1 reconstruction loss is incorporated. This loss encourages the generator to produce outputs that closely resemble the target image by minimizing the absolute differences between the corresponding pixel values, thereby preserving structural details and reducing large deviations in pixel intensities. To control the trade-off between adversarial loss and reconstruction loss, a weighting factor λ is introduced. The objective function of Pix2Pix can be written as:

$$\begin{aligned}\mathcal{L}_{\text{pix2pix}}(G, D) = {} & \mathbb{E}_{x,y}[\log\, D(x, y)] + \mathbb{E}_x[\log(1 - D(x, G(x)))] \\ & + \lambda\big(\mathbb{E}_{x,y}[\|y - G(x)\|_1]\big)\end{aligned}$$

Finally, the full Pix2Pix objective can be formulated as a Min-Max optimization problem, which defines the adversarial game between the generator G and the discriminator D, where G^* represents the optimal generator that solves this Min-Max problem:

$$G^* = \arg\min_G \max_D \mathcal{L}_{\text{pix2pix}}(G, D)$$

4.4.2 *μPIX* objectives

Our architecture is a variant of the classical Pix2Pix that introduces several key modifications to improve performance. First, rather than using a single balancing hyperparameter λ, we introduce two distinct weight parameters, w_1 and w_2, to independently regulate the contributions of adversarial loss and reconstruction loss, respectively. Second, we replace the L_1 loss with an L_2 loss, which has been shown to provide better results in terms of image generation quality in our setup. Third, as the discriminator in *μPIX* is a PatchGAN, which operates by classifying local patches of the image, rather than the entire image, we defined the *μPIX* objective function as:

$$\begin{aligned}\mathcal{L}_{\mu\text{pix}}(G, D) = {} & w_1\Bigg(\mathbb{E}_{x,y}\left[\frac{1}{N}\sum_{i=1}^{N}\log\, D(x_i, y_i)\right] \\ & + \mathbb{E}_x\left[\frac{1}{N}\sum_{i=1}^{N}\log\big(1 - D(x_i, G(x_i))\big)\right]\Bigg) \\ & + w_2\big(\mathbb{E}_{x,y}\big[\|y - G(x)\|_2^2\big]\big)\end{aligned}$$

where $D(x_i, y_i)$ and $D(x_i, G(x_i))$ are the discriminator PatchGAN's outputs for the i-th patch of the real image y given the input image x and the generated image $G(x)$ given the input image x, respectively, and N is the number of patches in the image. Finally, the full *μPIX* objective can be formulated as:

$$G^* = \arg\min_G \max_D \mathcal{L}_{\mu\text{pix}}(G, D)$$

4.5 Losses

4.5.1 Reconstruction loss for image generation (generator loss)

The Mean Squared Error (MSE) quantifies the average squared difference between the pixel values of the real and generated images. Given two images, Y (real) and $\hat{Y}$ (generated), each of dimensions $H \times W$, the MSE is computed as:

$$\mathrm{MSE}(Y, \hat{Y}) = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(y_{i,j} - \hat{y}_{i,j}\right)^2$$

where $y_{i,j}$ represents the pixel intensity of Y at spatial location (i, j), and $\hat{y}_{i,j}$ denotes the corresponding pixel intensity in $\hat{Y}$. A lower MSE value indicates that the generated image $\hat{Y}$ is closer to the real image Y in terms of pixel-wise similarity.

4.5.2 Discriminator loss for patch classification

The Binary Cross-Entropy (BCE) measures the difference between the predicted probability that a given image patch is real and the true label (real or fake) of that patch. Given a patch with true label $p \in \{0, 1\}$, where $p = 1$ indicates a real patch and $p = 0$ indicates a fake patch, and the predicted probability $\hat{p}$, the BCE loss is computed as:

$$\mathrm{BCE}(p, \hat{p}) = -\left[p \log(\hat{p}) + (1 - p) \log(1 - \hat{p})\right]$$

4.5.3 Adversarial loss for *μPIX network*

The adversarial loss is derived from the *μPIX* objective function, which combines both discriminator and reconstruction losses, and the user chosen parameters w_1 and w_2:

$$\mu\mathrm{PIX\_loss}(Y, \hat{Y}) = w_1 \cdot \mathrm{MSE}(Y, \hat{Y}) + w_2 \cdot \mathrm{BCE}(p, \hat{p})$$

4.6 Metrics

4.6.1 Evaluation metrics for image generation

We evaluated the performance of our image generation using two common metrics: MSE (described above) and Structural Similarity Index (SSIM).

SSIM measures the perceptual similarity between the real and generated images by considering luminance, contrast, structure and is computed as follows:

$$\mathrm{SSIM}(Y, \hat{Y}) = \frac{(2\mu_Y \mu_{\hat{Y}} + C_1)(2\sigma_{Y\hat{Y}} + C_2)}{\left(\mu_Y^2 + \mu_{\hat{Y}}^2 + C_1\right)(\sigma_Y^2 + \sigma_{\hat{Y}}^2 + C_2)}$$

where Y and $\hat{Y}$ represent the real and generated images, respectively. The terms μ_Y and $\mu_{\hat{Y}}$ are the mean pixel intensities of Y and $\hat{Y}$, while σ_Y^2 and $\sigma_{\hat{Y}}^2$ represent their variances. The covariance between Y and $\hat{Y}$ is given by $\sigma_{Y\hat{Y}}$, and C_1 and C_2 are small constants to stabilize the division.

SSIM provides a more perceptually relevant measure of image quality, as it accounts for structural information, which is often more aligned with human visual perception compared to pixel-wise differences captured by MSE.

4.6.2 Evaluation metrics for image segmentation

Intersection Over Union (IoU) is a common metric used to evaluate the performance of segmentation models.

Given two binary segmentation masks, S_{pred} (predicted) from a generated image and S_{gt} (ground truth) from a real image, where each pixel is either 0 (background) or 1 (foreground), the IoU is defined as:

$$\mathrm{IoU}(S_{\mathrm{pred}}, S_{\mathrm{gt}}) = \frac{\mathrm{card}\left(\{(i,j) \mid S_{\mathrm{pred}}(i,j) = 1 \text{ and } S_{\mathrm{gt}}(i,j) = 1\}\right)}{\mathrm{card}\left(\{(i,j) \mid S_{\mathrm{pred}}(i,j) = 1 \text{ or } S_{\mathrm{gt}}(i,j) = 1\}\right)}$$

The IoU value ranges from 0 to 1, and a higher IoU score indicates that the model segmentation S_{pred} more closely matches the ground truth S_{gt}.

Precision measures the accuracy of positive predictions.

Given two binary segmentation masks, S_{pred} (predicted) from a generated image and S_{gt} (ground truth) from a real image, where each pixel is either 0 (background) or 1 (foreground), precision is defined as follows:

$$\mathrm{Precision}(S_{\mathrm{pred}}, S_{\mathrm{gt}}) = \frac{\mathrm{card}\left(\{(i,j) \mid S_{\mathrm{pred}}(i,j) = 1 \text{ and } S_{\mathrm{gt}}(i,j) = 1\}\right)}{\mathrm{card}\left(\{(i,j) \mid S_{\mathrm{pred}}(i,j) = 1\}\right)}$$

Recall quantifies the model's ability to identify all relevant positive instances.

Given two binary segmentation masks, S_{pred} (predicted) from a generated image and S_{gt} (ground truth) from a real image, where each pixel is either 0 (background) or 1 (foreground), recall is defined as:

$$\mathrm{Recall}(S_{\mathrm{pred}}, S_{\mathrm{gt}}) = \frac{\mathrm{card}\left(\{(i,j) \mid S_{\mathrm{pred}}(i,j) = 1 \text{ and } S_{\mathrm{gt}}(i,j) = 1\}\right)}{\mathrm{card}\left(\{(i,j) \mid S_{\mathrm{gt}}(i,j) = 1\}\right)}$$

The F1-score is the harmonic mean of Precision and Recall, providing a single metric that balances both aspects. Given two binary segmentation masks, S_{pred} (predicted) from a generated image and S_{gt} (ground truth) from a real image, the F1-score is defined as:

$$\mathrm{F1-score}(S_{\mathrm{pred}}, S_{\mathrm{gt}}) = 2 \times \frac{\mathrm{Precision}(S_{\mathrm{pred}}, S_{\mathrm{gt}}) \times \mathrm{Recall}(S_{\mathrm{pred}}, S_{\mathrm{gt}})}{\mathrm{Precision}(S_{\mathrm{pred}}, S_{\mathrm{gt}}) + \mathrm{Recall}(S_{\mathrm{pred}}, S_{\mathrm{gt}})}$$

4.7 *μPIX* architecture

Based on Pix2Pix, our architecture consists of a generator and a discriminator. The generator follows a U-Net architecture with an EfficientNet-B0 backbone [45], implemented using the TensorFlow Segmentation Model library [61], while the discriminator employs a PatchGAN classifier [44], [62].

4.7.1 Generator encoder

The encoder of our U-Net generator is based on the EfficientNet architecture, which is optimized for both accuracy and computational efficiency. EfficientNet employs a multi-objective neural architecture search to balance model accuracy with floating-point operations per second (FLOPS). The encoder is composed of several stages, each utilizing Mobile Inverted Bottleneck Convolutions (MBConv) [63] with varying kernel sizes, enabling the model to efficiently capture hierarchical features. These blocks integrate three key components: the expansion layer, depthwise separable convolution [64], and the squeeze-and-excitation (SE) module [65]. The expansion layer increases the number of channels before processing, allowing the network to learn richer representations. Depthwise separable convolution reduces computational cost by applying spatial convolutions independently to each input channel, followed by a pointwise 1×1 convolution to fuse channel information. The SE module dynamically recalibrates feature maps by adaptively weighting channels based on their importance. Starting with an initial convolutional layer, the architecture sequentially applies MBConv blocks with different dilation factors and depths, progressively reducing spatial resolution while expanding the number of channels. This progressive structure is specifically designed to maximize performance while minimizing computational cost. Furthermore, the integration of squeeze-and-excitation optimization in every MBConv block enhances the model's capacity to focus on the most salient features, allowing the encoder to learn rich, discriminative feature representations. *μPIX* uses an EfficientNet-B0 backbone consisting of 7 stacked MBConv blocks with varying expansion factors and convolution kernel sizes. Additionally, we opted for a non-pretrained version of EfficientNet-B0 in the generator. This decision helps maintain a balanced training dynamic between the generator and the discriminator, preventing a situation where a pretrained generator might overpower the discriminator due to its stronger initial performance, which could hinder effective adversarial learning [47].

4.7.2 Generator decoder

The decoder part of our U-Net generator follows a progressive upsampling approach to restore the spatial resolution of feature maps. It consists of five upsampling stages followed by two consecutive 3×3 convolutional layers with ReLU activation. At each stage, the upsampled feature maps are concatened with skip connections from the encoder, allowing network to recover fine-grained spatial details that where lost in the downsampling process. This concatenation is performed along the channel axis to preserve multiscale contextual information. The number of filters used at each decoding stage follow the sequence (256, 128, 64, 32, 16) ensuring a gradual reduction in feature complexity as spatial resolution increases. The final layer of the decoder applies a 3×3 convolution with a single output channel (for grayscale image generation) and a hyperbolic tangent activation function. This ensures that the output pixel values are mapped to the [-1,1] range. This choice ensures that the generated images are consistent with the normalized pixel values of real images, where the pixel values are centered around zero. By maintaining this symmetry, the model avoids biases that could arise from a non-centered range, improving training stability and enabling smoother convergence in the adversarial learning process [44], [66].

4.7.3 Discriminator

The discriminator in our architecture is based on a convolutional PatchGAN model, which distinguishes between real and generated image patches. The core idea behind PatchGAN is to classify patches of the input images as either real (from the true target image) or fake (from the generated image), rather than evaluating the entire image as a whole. While alternative patch sizes such as 1×1 and 70×70 could have been considered, original results on Pix2Pix [44] have shown that smaller patch sizes, such as 1×1, focus primarily on color diversity but lack spatial consistency, making them less effective for capturing detailed textures. On the other hand, larger patch sizes, like 70×70, tend to enforce sharper outputs, but they may introduce unrealistic artifacts due to the mismatch in scale between local and global image features. Based on these insights, we opted for a patch size of 16×16, as it strikes a balance between capturing fine-grained details and avoiding overly localized artifacts. The input to the discriminator consists of image pairs: a real image and a generated image provided by the generator. These two are concatenated along the channel axis and then passed through a series of convolutional layers, each

designed to progressively extract more abstract features from the image. Finally, the output is passed through a sigmoid activation function, which gives a probability score for each patch, indicating whether it is real or fake. The discriminator is trained using the BCE loss, which compares the predicted patch labels with the ground truth.

4.8 Denoising algorithms on CSBDeep benchmark dataset

To comprehensively evaluate our approach, we compared *μPIX* against several established denoising algorithms, including both non-deep learning-based and deep learning-based methods.

For non-deep learning-based denoising, we applied NLM using Python *scikit-image restoration* package, with a patch size of 7, a patch distance of 11 and a smoothing factor of 100. BM3D was implemented via the Python *bm3d* package using a standard noise standard deviation of 800 for optimal noise reduction. For the Low-pass filtering, we applied a Fourier Transform-based low pass filter using the *scipy fftpack* library. A cutoff frequency of 0.1 was used to mask out high-frequency components. For Total Variation Denoising (TVD), we utilized the *denoise_tv_chambolle* function from the *scikit-image restoration*library. The method applies a regularization parameter, weight, to control the strength of the denoising process (weight was set to 1,000).

For deep learning-based denoising, we used the pre-trained "Denoise Nuclei" model provided in the Cellpose3 distribution. For RCAN, we used a popular PyTorch implementation [67]. The model was trained for 10 epochs with a learning rate of $1e^{-4}$ and a batch size of 4 on an A40-48 GPU. The training was conducted using the MSE loss function with default parameters for the network architecture, including 64 feature maps, 10 residual groups, and 20 residual channel attention blocks. The CARE model was trained using the Python *csbdeep* package using the default parameters [68].

4.9 *μPIX* training parameters and inference

Although we rely on the well-known Pix2Pix architecture, we have developed unique training strategies. Since determining optimal stopping criteria for generative adversarial networks is challenging, we implemented a custom early stopping method. At the end of each epoch, the ratio between the MSE and the SSIM was evaluated on the validation set. If this ratio reached a new minimum, the model was saved, and training continued for a set number of epochs defined by the patience hyperparameter (set to 20 by default). If no improvement was observed within the patience period, training was halted. Moreover, adaptive learning rate scheduling has been introduced for both adversarial and discriminator losses. If the validation performance does not improve for a specified number of epochs (patience), set to 20 by default, the learning rate is reduced by a factor of 10 %, with a minimum learning rate of $1e^{-5}$ to ensure stable training.

All models we developed were trained using Adam Optimizer, an initial learning rate of $1e^{-2}$ and $\beta_1 = 0.8$ for both generator and discriminator for faster initial convergence. We empirically chose a loss weight imbalance of 10 between BCE and MSE and a batch size of 128. We trained our models for a maximum of 100 epochs or until the early stoping is triggered. Our models were trained on a NVIDIA A6000 GPU with 48 GB of memory. Depending of the dataset size, the training time takes from 5 h (CSBDeep dataset) to 7 h (our rejuvenation dataset) to converge.

At inference, only the generator component is used and the pixel range of the generated images was remapped to $[0, 255]$ for consistency with the original input image format.

4.10 *μPIX* software

The complete *μPIX* software package is based on Tensorflow 2.14. The source code, pre-trained models, installation instructions, and detailed usage guidelines, is available on our GitLab repository: https://gitlab.lis-lab.fr/sicomp/mupix. We provide dedicated use cases for both inference and training, along with ready-to-use demonstration notebooks to facilitate experimentation. To further support users, we have included video tutorials covering installation, model deployment, inference and training procedures. In addition, a reproducibility notebook is provided, allowing users to replicate the key results presented in this study.

Acknowledgments: We acknowledge the France-Bioimaging Infrastructure (ANR-24-INBS-0005 FBI BIOGEN). We would like to thank the GDR Imabio consortium (CNRS) for funding the start of this project with an exploratory master's fellowship and the IBDM management and CNRS for funding Gabriel Bon. Special thanks to Sham Tlili for providing us with gastruloids samples and stimulating discussions. We thank all IBDM members who join us to contribute to image-blind evaluation.

Research ethics: The conducted research is not related to either human or animals use.

Informed consent: Informed consent was obtained from all individuals included in this study.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission. C.M. and F.D. conceived the project,

G.B. performed experiments, built the dataset with support of D.S. and C.M, analyzed the data and performed the deep learning model conception, training, validation and software development with the support of F.D, C.M. and F.D. wrote the manuscripts with inputs from all authors.

Use of Large Language Models, AI and Machine Learning Tools: None declared.

Conflict of interest: Authors state no conflicts of interest.

Research funding: France-Bioimaging Infrastructure (ANR-24-INBS-0005 FBI BIOGEN), GDR Imabio consortium (CNRS).

Data availability: The source code and data are available on https://gitlab.lis-lab.fr/sicomp/mupix.

References

[1] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 60—5, 2005.

[2] S. Huang and S. Wan, "A total variation denoising method based on median filter and phase consistency," *Sens. Imag.*, vol. 21, no. 1, p. 19, 2020.

[3] M. Lebrun, "An analysis and implementation of the BM3D image denoising method," *Image Process. Line*, vol. 2, pp. 175—213, 2012, [Online]. Available: https://www.ingentaconnect.com/content/doaj/21051232/2012/00000002/00000001/art00011.

[4] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast interscale wavelet denoising of poisson-corrupted images," *Signal Process.*, vol. 90, no. 2, pp. 415—27, 2010.

[5] T. Blu and F. Luisier, "The sure-let approach to image denoising," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2778—86, 2007.

[6] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast haar-wavelet denoising of multidimensional fluorescence microscopy data," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009, pp. 310—3.

[7] Lanteri, H., C. Aime, H. Beaumont, and P. Gaucherel, "Blind deconvolution using the Richardson-Lucy algorithm," in *Optics in Atmospheric Propagation and Random Phenomena*, vol. 2312, A. Kohnle and A. D. Devir, Eds., International Society for Optics and Photonics. SPIE, 1994, pp. 182-92.

[8] L. Chen, *et al*., "Deep Richardson-Lucy deconvolution for low-light image deblurring," 2023. [Online]. Available: https://arxiv.org/abs/2308.05543.

[9] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction wiener filter," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 4, pp. 1218—34, 2006.

[10] C. Bled and F. Pitié, "Pushing the limits of the wiener filter in image denoising," 2023. [Online]. Available: https://arxiv.org/abs/2303.16640.

[11] A. H. Klemm, A. W. Thomae, K. Wachal, and S. Dietzel, "Tracking microscope performance: A workflow to compare point spread function evaluations over time," *Microsc. Microanal.*, vol. 25, no. 3, pp. 699—704, 2019.

[12] Y. Li, *et al*., "Incorporating the image formation process into deep learning improves network performance," *Nat. Methods*, vol. 19, no. 11, pp. 1427—37, 2022.

[13] K. Yanny, K. Monakhova, R. W. Shuai, and L. Waller, "Deep learning for fast spatially varying deconvolution," *Optica*, vol. 9, no. 1, pp. 96—9, 2022.

[14] T. Chobola, *et al*., "Lucyd: A feature-driven Richardson-Lucy deconvolution network," 2023. [Online]. Available: https://arxiv.org/abs/2307.07998.

[15] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Advances in Neural Information Processing Systems*, vol. 27, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds., Curran Associates, Inc., 2014. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2014/file/1c1d4df596d01da60385f0bb17a4a9e0-Paper.pdf.

[16] J. Zhang, Z. Yue, H. Wang, Q. Zhao, and D. Meng, "Blind image deconvolution by generative-based kernel prior and initializer via latent encoding," in *Computer Vision — ECCV 2024*, A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, and G. Varol, Eds., Cham, Springer Nature Switzerland, 2025, pp. 73—92.

[17] B. Zhang, X. Sun, J. Mai, and W. Wang, "Deep learning-enhanced fluorescence microscopy via confocal physical imaging model," *Opt. Express*, vol. 31, no. 12, pp. 19 048—19 064, 2023.

[18] Emmanouil Xypakis, Valeria deTurris, Fabrizio Gala, Giancarlo Ruocco, and Marco Leonetti, "Physics-informed machine learning for microscopy," *EPJ Web Conf.*, vol. 266, p. 04007, 2022.

[19] N. Gritti, R. M. Power, A. Graves, and J. Huisken, "Image restoration of degraded time-lapse microscopy data mediated by near-infrared imaging," *Nat. Methods*, vol. 21, no. 2, pp. 311—21, 2024.

[20] M. Weigert, *et al*., "Content-aware image restoration: Pushing the limits of fluorescence microscopy," *Nat. Methods*, vol. 15, no. 12, pp. 1090—7, 2018.

[21] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: A generalist algorithm for cellular segmentation," *Nat. Methods*, vol. 18, no. 1, pp. 100—6, 2021.

[22] M. Pachitariu and C. Stringer, "Cellpose 2.0: How to train your own model," *Nat. Methods*, vol. 19, no. 12, pp. 1634—41, 2022.

[23] C. Stringer and M. Pachitariu, "Cellpose3: One-click image restoration for improved cellular segmentation," *bioRxiv*, 2024, [Online]. Available: https://www.biorxiv.org/content/early/2024/02/12/2024.02.10.579780.

[24] J. Chen, *et al*., "Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes," *Nat. Methods*, vol. 18, no. 6, pp. 678—87, 2021.

[25] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void — Learning denoising from single noisy images," 2019. [Online]. Available: https://arxiv.org/abs/1811.10980.

[26] J. Lehtinen, *et al*., "Noise2noise: Learning image restoration without clean data," 2018. [Online]. Available: https://arxiv.org/abs/1803.04189.

[27] J. Lequyer, R. Philip, A. Sharma, W.-H. Hsu, and L. Pelletier, "A fast blind zero-shot denoiser," *Nat. Mach. Intell.*, vol. 4, no. 11, pp. 953—63, 2022.

[28] R. Zhou, M. E. Helou, D. Sage, T. Laroche, A. Seitz, and S. Süsstrunk, "W2s: Microscopy data with joint denoising and super-resolution for widefield to sim mapping," 2020. [Online]. Available: https://arxiv.org/abs/2003.05961.

[29] M. Chatton, "Microscopy image restoration using deep learning on w2s," 2020. [Online]. Available: https://arxiv.org/abs/2004.10884.

[30] Y. Wu, *et al*., "Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning," *Nat. Methods*, vol. 16, no. 12, pp. 1323—31, 2019.
[31] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," 2020. [Online]. Available: https://arxiv.org/abs/2006.11239.
[32] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," 2021. [Online]. Available: https://arxiv.org/abs/2105.05233.
[33] B. Xia, *et al*., "Diffir: Efficient diffusion model for image restoration," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 13 049—13 059.
[34] C. Saharia, *et al*., "Palette: Image-to-image diffusion models," 2022. [Online]. Available: https://arxiv.org/abs/2111.05826.
[35] H. Sahak, D. Watson, C. Saharia, and D. Fleet, "Denoising diffusion probabilistic models for robust image super-resolution in the wild," 2023. [Online]. Available: https://arxiv.org/abs/2302.07864.
[36] Y. Wang, *et al*., "Sinsr: Diffusion-based image super-resolution in a single step," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 25 796—25 805.
[37] Y. Xie, M. Yuan, B. Dong, and Q. Li, "Diffusion model for generative image denoising," 2023. [Online]. Available: https://arxiv.org/abs/2302.02398.
[38] C. Yang, L. Liang, and Z. Su, "Real-world denoising via diffusion model," 2023. [Online]. Available: https://arxiv.org/abs/2305.04457.
[39] A. Saguy, *et al*., "This microtubule does not exist: Super-resolution microscopy image generation by a diffusion model," *Small Methods*, p. 2400672, 2024, [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/smtd.202400672.
[40] H. Bachimanchi and G. Volpe, "Diffusion models for super-resolution microscopy: A tutorial," *J. Phys.: Photonics*, vol. 7, no. 1, p. 013001, 2025.
[41] C. Lu, *et al*., "Diffusion-based deep learning method for augmenting ultrastructural imaging and volume electron microscopy," *Nat. Commun.*, vol. 15, no. 1, p. 4677, 2024.
[42] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham, Springer International Publishing, 2015, pp. 234—41.
[43] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014. [Online]. Available: https://arxiv.org/abs/1411.1784.
[44] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967—76.
[45] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2020. [Online]. Available: https://arxiv.org/abs/1905.11946.
[46] C. Li and M. Wand, "Precomputed real-time texture synthesis with Markovian generative adversarial networks," 2016. [Online]. Available: https://arxiv.org/abs/1604.04382.
[47] I. J. Goodfellow, *et al*., "Generative adversarial networks," 2014. [Online]. Available: https://arxiv.org/abs/1406.2661.
[48] V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, "Annotated high-throughput microscopy image sets for validation," *Nat. Methods*, vol. 9, no. 7, p. 637, 2012.
[49] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, *Cell Detection with Star-Convex Polygons*, Granada (Spain), Springer International Publishing, 2018, pp. 265—73.
[50] Hamamatsu, "Photomultiplier tube," 2024. [Online]. Available: https://www.hamamatsu.com/us/en/product/optical-sensors/pmt/pmt_tube-alone/side-on-type/R6357.html.
[51] Hamamatsu1, "H7422p-40," 2024. [Online]. Available: https://www.alldatasheet.com/datasheet-pdf/pdf/62585/HAMAMATSU/H7422P-40.html.
[52] A. Hashmi, *et al*., "Cell-state transitions and collective cell movement generate an endoderm-like region in gastruloids," *eLife*, vol. 11, 2022, [Online]. Available: https://elifesciences.org/articles/59371.
[53] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," 2018. [Online]. Available: https://arxiv.org/abs/1706.08500.
[54] S. Barratt and R. Sharma, "A note on the inception score," 2018. [Online]. Available: https://arxiv.org/abs/1801.01973.
[55] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248—55.
[56] K. Combs, T. J. Bihl, A. Gadre, and I. Christopherson, "A human-factors approach for evaluating ai-generated images," in *Proceedings of the 2024 Computers and People Research Conference, Ser. SIGMIS-CPR '24*, New York, NY, USA, Association for Computing Machinery, 2024.
[57] H. Hudson and T. C. Lee, "Maximum likelihood restoration and choice of smoothing parameter in deconvolution of image data subject to poisson noise," *Comput. Stat. Data Anal.*, vol. 26, no. 4, pp. 393—410, 1998.
[58] N. Ichimura, "Spatial frequency loss for learning convolutional autoencoders," *ArXiv*, vols. abs/1806.02336, 2018.
[59] S. Hong and S.-K. Song, "Kick: Shift-n-overlap cascades of transposed convolutional layer for better autoencoding reconstruction on remote sensing imagery," *IEEE Access*, vol. 8, pp. 107 244—107 259, 2020.
[60] K. E. I. V. K. A, A. Buslaev and A. Parinov, "Albumentations: Fast and flexible image augmentations," *ArXiv e-prints*, 2018.
[61] P. Iakubovskii, "Segmentation models," 2019. Available: https://github.com/qubvel/segmentation_models.
[62] C. Li and M. Wand, "Precomputed real-time texture synthesis with Markovian generative adversarial networks," in *Computer Vision — ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham, Springer International Publishing, 2016, pp. 702—16.
[63] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA, IEEE Computer Society, 2018, pp. 4510—20. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00474.
[64] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA, IEEE Computer Society, 2017, pp. 1800—7. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.195.
[65] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132—41.

[66] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2016. [Online]. Available: https://arxiv.org/abs/1511.06434.
[67] YJN870, "Rcan-pytorch," 2021, https://github.com/yjn870/RCAN-pytorch [accessed: Mar 08, 2025].
[68] T. C. Team, "Care — Channel attention recurrent u-net," 2021, https://github.com/csbdeep/csbdeep [accessed: Mar 08, 2025].

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/mim-2024-0024).