REAL-TIME IMAGE DENOISING OF MIXED POISSON-GAUSSIAN NOISE IN FLUORESCENCE MICROSCOPY IMAGES USING IMAGEJ

A PREPRINT

Varun Mannam* *
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46656
vmannam@nd.edu

Yide Zhang
Andrew and Peggy Cherng Department of Medical Engineering
California Institute of Technology
Pasadena, CA 91125
yzhang43@caltech.edu

Yinhao Zhu
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46656
yinhaozh@gmail.com

Evan Nichols
Department of Biological Sciences
University of Notre Dame
Notre Dame, IN 46656
Evan.L.Nichols.73@nd.edu

Qingfei Wang
Department of Biological Sciences
University of Notre Dame
Notre Dame, IN 46656
qwang9@nd.edu

Vignesh Sundaresan
Department of Chemical and Biomolecular Engineering
University of Notre Dame
Notre Dame, IN 46656
vsundare@nd.edu

Siyuan Zhang
Department of Biological Sciences
University of Notre Dame
Notre Dame, IN 46656
szhang8@nd.edu

Cody Smith
Department of Biological Sciences
University of Notre Dame
Notre Dame, IN 46656
csmith67@nd.edu

Paul W Bohn
Department of Chemical and Biomolecular Engineering
Department of Chemistry and Biochemistry
University of Notre Dame
Notre Dame, IN 46656
pbohn@nd.edu

Scott Howard*
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46656
showard@nd.edu

November 10, 2021

ABSTRACT

Fluorescence microscopy imaging speed is fundamentally limited by the measurement signal-to-noise ratio (SNR). To improve image SNR for a given image acquisition rate, computational denoising techniques can be used to suppress noise. However, analytical techniques to estimate a denoised image from a single frame are either computationally expensive or rely on simple noise statistical models. These models assume Poisson or Gaussian noise statistics, which are not appropriate for many fluorescence microscopy applications that contain quantum shot noise and electronic Johnson–Nyquist

*This material is based upon work supported by the National Science Foundation (NSF) under Grant No. CBET-1554516 and by the Department of Energy Office of Science under Grant DE-SC0019312. *Corresponding author: vmannam@nd.edu or showard@nd.edu. This version of the manuscript is submitted to the Optica Journal. ©2021 Optica Publishing Group.

noise, and therefore a mixture of Poisson and Gaussian noise. In this paper, we show convolutional neural networks (CNNs) trained on mixed Poisson and Gaussian noise images to overcome the limitations of existing image denoising methods. The trained CNN is presented as an open-source ImageJ plugin that performs instant image denoising (within tens of milliseconds) with superior performance ($\sim$ 8.1 dB SNR improvement) compared to the conventional fluorescence microscopy denoising methods. The method is validated on external datasets with out-of-distribution noise and contrast from the training data and consistently achieves high performance (> 8dB) denoising in less time than other fluorescence microscopy denoising methods.

Keywords Fluorescence microscopy · Noise in imaging systems · Image enhancement · Pattern recognition, neural networks · Neural networks · Mixed Poisson-Gaussian (MPG) noise.

1 Introduction

In modern biology, fluorescence microscopy plays a vital role in functional and structural imaging [1]. However, the imaging speed of fluorescence microscopy is limited by the inherent noise in the system [2]. Fluctuations in the photon emission events contribute to shot noise which follows Poisson statistics per time period. Additionally, microscope electronic components contribute Johnson–Nyquist noise, which follows a Gaussian distribution per time period. The combination of both of these contributions leads to fluorescence microscopy systems containing a mixed Poisson-Gaussian (MPG) noise. Due to this inherent noise in the system, the acquired image has a limited signal-to-noise ratio (SNR) that hinders the underlined ground truth information about the biological sample. Therefore, a robust denoising approach is required to address these two distinct noise distributions.

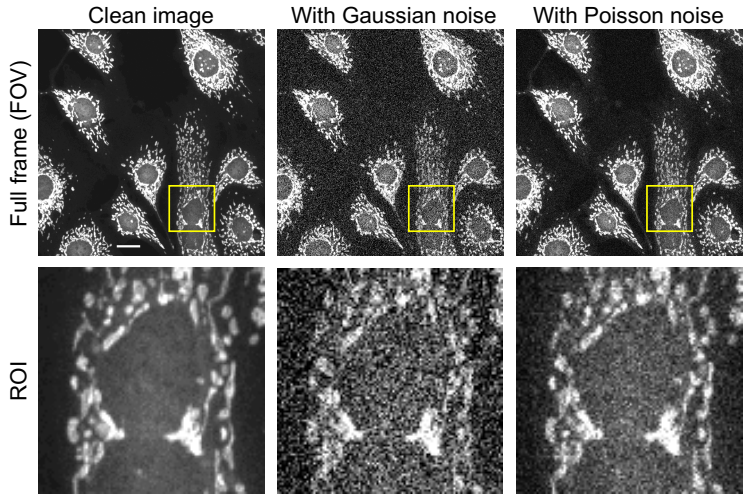

Figure 1: Illustration of different types of noise with a BPAE cell sample. The top row indicates the full frame images and the PSNR of the synthetic noisy image is 25 dB with reference to clean image (obtained by averaging 50 noisy images in the same FOV). The bottom row shows the selected region of interest (ROI) as shown in the yellow box of the corresponding top row images. Full frame and ROI dimensions are 512 $\times$ 512 and 125 $\times$ 125 respectively. Scale bar: 20 μm.

In fluorescence microscopy, the noise is introduced from two sources: quantum noise (from the stream of photons) and thermal noise (from electronic components). However, the statistics of each noise source affect images differently. To illustrate these differences, we simulate the effects of Gaussian noise and Poisson noise individually on a clean image in Figure 1. The clean image represents bovine pulmonary artery endothelial cells (BPAE, Invitrogen FluoCells slide #1 F36924, mitochondria labeled with MitoTracker Red CMXRos, F-actin labeled with Alexa Fluor 488 phalloidin, nuclei labeled with DAPI) imaged with a commercial fluorescence microscope (Nikon A1R-MP laser scanning confocal microscope equipped with a Nikon Apo LWD 40x, 1.15 NA water immersion objective and more details are provided in the Supplementary Note S1) with a 4 mW excitation power and 50 frame averaging to produce a high SNR image. Gaussian noise and Poisson noise are added to the clean image (SNR of 25 dB with reference to the clean image) and shown as “With Gaussian noise," and “With Poisson noise," images, respectively. The “Clean image" is obtained by averaging 50 noisy images in the same field of view (FOV). Here, the Gaussian noise is independent of clean image

pixel intensity, whereas the Poisson noise is pixel intensity-dependent. We can observe this behaviour in the marked ROI center region, i.e., the smooth gray area (Nuclei region). Hence, Poisson noise and Gaussian noise affect the clean image differently, which makes the MPG noise denoising a challenging task.

To improve the peak signal-to-noise ratio (PSNR) (which is a way to quantify the SNR in a given image with reference to ground truth) in fluorescence microscopy images, one can increase the sample dosage (either the excitation power or the pixel dwell time, which is also known as the integration time) in the system. However, this method could lead to photobleaching or photodamage and cannot be used for imaging real-time dynamics in living animals. Furthermore, lower dosage leads to noisy (with low PSNR) images and thereby posing a challenge for applications like cell classification and segmentation [3]. Denoising algorithms are available improve image SNR, but are either designed specifically for Gaussian or Poisson noise (thereby not suitable for the other distribution or MPG noise), or are capable of denoising MPG noise but require long computation time to achieve relatively low PSNR improvement. Therefore, there is a need for a computational fast single-frame denoising method that works for all experimental noise distributions: Poisson, Gaussian, and MPG.

Reiterating the mathematical notation of different noise distributions using 1st and 2nd order moments (or simply mean and variance) [4], the Gaussian noise distribution is represented as standard normal distribution $n_g \sim \mathcal{N}(0, b)$ where the mean is zero and the variance is b. Similarly, the Poisson noise distribution can be represented as $n_p \sim \mathcal{N}(y_i, ay_i)$ where the true pixel value (y_i) is the mean and Poisson variance is ay_i (conversion factor or quantum efficiency of a times y_i). Assuming that the Poisson and Gaussian processes are independent, the MPG noise distribution is represented as $n_{\text{mixed}} \sim \mathcal{N}(y_i, ay_i + b)$ where the mean is the true pixel value (y_i) and the variance is the addition of two independent noise variances ($ay_i + b$). Hence, MPG noise distribution is similar to Poisson noise with an additional Gaussian noise variance (b).

In order to increase the SNR of an image, additional information must be included to estimate the original, denoised image. If the noise distribution is known *a priori* to be Gaussian, simple filtering (by assuming noise exceeds signal at high spatial frequencies by using mean, median, bilateral filters) or machine-learning based approaches such as non-local means (NLM) [5], block matching 3D filtering (BM3D) [6], K-means singular value decomposition (K-SVD) [7], expected patch log-likelihood (EPLL) [8], and weighted nuclear norm minimization (WNNM) [9] can all be used to leverage knowledge of the noise distribution and to leverage information from similar features within the same image. If the noise distribution is known to be Poisson, the above methods can be used by (1) converting the Poisson noise into Gaussian noise using a variance stabilization transform (VST) (e.g., generalized Anscombe transform (GAT [10])), (2) denoising using popular Gaussian denoising algorithms, and (3) converting the denoised results back using an inverse VST [10]. Such popular Poisson denoising methods include Poisson NLM (PNLM) [11], non-local principle component analysis (NL-PCA) [12], VST+NLM and VST+BM3D which apply VST, NLM or BM3D Gaussian denoising methods, and inverse VST, sequentially, on an image. If the noise is from a MPG distribution, more complicated methods are needed to estimate the relative contributions of each process at each pixel (e.g., PURE-LET [13]).

Overall, each of these image denoising methods have trade-offs between denoising performance, computation speed, and applicability to a specific noise distribution. In this paper, we show how a convolutional neural network (CNN) trained on mixed Poisson-Gaussian noise can in fact achieve high-performance denoising, quickly, on images with Poisson and/or Gaussian noise distributions.

Recently, CNN based fluorescence microscopy image denoising as demonstrated improved denoising performance and speed [14]. CNN-based image denoising methods can be classified into two categories: supervised and self-supervised. Supervised methods require training the network on a dataset prior to performing denoising on experimental images and include content-aware image restoration (CARE) [14], denoising convolutional neural networks (DnCNN) [15], and Noise2Noise (N2N) [16]. Self-supervised methods train the network on the same image it is performing denoising on and Noise2Void (N2V) [17], probabilistic Noise2Void (PN2V) [18], Noise2Fast [19], and other methods [20, 21]. Supervised CNN training methods typically provide higher performance than the self-supervised methods due to the inclusion of additional information in target images. In this paper, we utilize a supervised CNN model that is capable of removing MPG noise from a single frame nearly instantly (within a few ms). By producing and utilizing a large, multiple species, and multiple microscope modality training dataset [22], the trained model outperforms existing image denoising methods. More importantly, our method demonstrates superior denoising performance compared to the existing image denoising approaches including ML-based methods regardless of the noise distribution presented in the fluorescence microscopy images.

To address the fundamental problem of denoising microscopy images containing MPG noise sources, we train and evaluate well-known CNN models on MPG noise microscopy data from the FMD dataset [23] in Sections 2 and 3. In Section 4, the trained ML model is validated by evaluating performance in a diverse set of microscopy images, including the W2S fluorescence microscopy dataset [24], with out-of-training-distribution PSNR and contrast. The method

consistently produces high-quality denoised images and is packaged as an ImageJ plugin to reduce the barrier-to-entry for applying denoising CNNs to microscopy data.

2 Methods

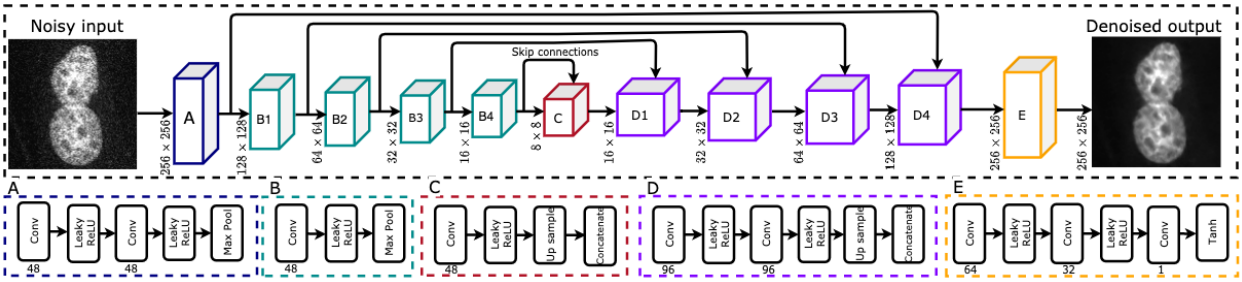

Figure 2: The Noise2Noise architecture we use to train on the FMD dataset. We divide the ML model mainly into two sections, namely encoder (from block A to C: contrast path) and decoder (from block C to E: expansion path). At every block, input dimensions are provided, and the corresponding operations in each block are represented in the sub-blocks along with the number of output channels in the convolution provided in **bold**. Skip connections are useful for adding features from encoder to decoder which might be suppressed in the latent space (at C block). Noisy input and Denoised output image dimensions are 256 $\times$ 256.

As described in Section 1, the fundamental stochastic nature of MPG noise requires introducing some *a priori* knowledge to the denoising process. Analytical approaches use knowledge of the statistical properties of noise, but are computationally slow and not well suited for MPG sources. Deep CNNs, however, are data-driven and expressive enough to overcome the complex statistical nature of MPG noise. To perform denoising using CNNs, a network architecture and training datasets must be chosen. In this paper, we choose to focus on the Noise2Noise [16] and DnCNN network [15] architectures for two reasons. First, they have demonstrated superior performance to other CNN architectures [17]. Second, robust performance in diverse microscopy environments requires training on large datasets. Supervised training of the Noise2Noise and DnCNN models can be achieved in reasonable time ($\approx$ 4 hours for the entire FMD dataset on a single Nvidia 1080-ti GPU) while training other networks (e.g., Noise2Void) on the dataset can take a prohibitively long time ($\approx$ 3.50 hours for a single image training on the GPU). Noise2Void is a popular denoising method that is capable of self-supervised learning, so we will therefore compare supervised trained model performance to self-supervised Noise2Void performance in Section 4.4.1. We chose to train these two models on our FMD dataset [22, 23] that contains 12,000 raw noisy images from confocal [25], two-photon [26], and widefield [27] microscopes. The FMD dataset also helps to benchmark various denoising techniques on the fluorescence images. The creation of the training and testing datasets from the FMD dataset is explained in the Supplementary Note S1.

The first ML model we evaluate is based on a deep CNN with the Noise2Noise architecture [16] trained to denoise fluorescence images with MPG noise. The ML model is similar to the originally published Noise2Noise architecture [16] with the addition of a non-linear activation (tanh) layer after the final convolutional layer [23], adaptive learning (one cycle policy) during the training phase, and carefully tuned hyper-parameters such as batch size and learning rate for achieving best fluorescence microscopy image denoising performance.

In the Noise2Noise CNN model, the input and target are both noisy images within the same FOV (which is the case when the ground truth is absent or its extraction is difficult). This ML model is beneficial compared to our second ML model which is based on the DnCNN architecture [15] that requires ground truth since getting the clean image (typically acquired with the increased dosage) is usually difficult, especially when imaging living species. Figure 2 shows the complete Nosie2Noise architecture as our first ML model. Here we divide the ML architecture into five blocks. Block A to C represent an encoder structure, and Block C to E represent a decoder structure. In the encoder, the image feature size (256 $\times$ 256) is reduced by selecting only essential features of the image (with more channels to select different features), and at block C, the images with feature map size (8$\times$8) are less noisy when compared to the noisy input image. In the decoder, this noise-free image is restored to the original noisy image feature size (256 $\times$ 256). In this work, both input and target are noisy images. At the i-th pixel of the input and target images, the noisy measurements are $z_1 = y + n_1$ and $z_2 = y + n_2$, where y is the real value and n_1 and n_2 are MPG noise values corresponding to the input and target, respectively. The Noise2Noise model estimates the real value y instead of the noisy target z_2 since n_2 is random. The Noise2Noise model can predict the clean image since training happens on the noisy image pair as input and target and this ML model works well as long as the clean image is the average of noisy target images within a FOV.

The other ML model is based on a deep CNN with the DnCNN architecture trained to denoise the fluorescence images with MPG noise. The DnCNN ML model can estimate the residual (only noise) for a given noisy input image and the denoised image is obtained by subtracting the estimated residual from the noisy input image. Unlike the supervised DnCNN ML model which requires clean images as the target, the Noise2Noise ML model utilizes the noisy image as the target. The training and testing of the proposed ML models are explained in the Supplementary Note S2.

While the proposed ML approach is fast and accurate, the complexity of implementing the ML model is a barrier to entry. To address this barrier, we integrated our ML models as plugins to ImageJ, one of the most widely used biomedical image processing packages [28]. Two ImageJ plugins are presented for two ML models: Noise2Noise and DnCNN. We have developed our ImageJ plugins based on the ImageJ-tensorflow library [29]. Here the library provides the minimal set of functions in ImageJ to load the pre-trained ML model weights as a graph and perform a set of convolutional operations on the input image (loaded by ImageJ) which results in the output of the trained ML model. To accommodate more image formats, we have extended the ImageJ-tensorflow library to handle 16-bit and 32-bit image data-types (in addition to the original 8-bit images) for both training and inference. The required image pre-processing (e.g., pixel value normalization) is performed automatically by the plugin, allowing users to load and process images without manually pre-processing. Both ML models were trained in Keras with TensorFlow as the backend. ImageJ plugins simplify applying denoising CNNs by integrating into biomedical researcher's existing workflow. Additionally, the plugins support graphical processing unit (GPU) computing, resulting in a significant speed increase compared to conventional ImageJ image denoising plugins [30, 14, 31] that run on a central processing unit (CPU). Also, recent implementation methods such as DeepImageJ plugin [31] and ZeroCost4Microscopy Jupyter python code [32] can be used with our pre-trained ML model weights (saved in .zip format) to perform image denoising. The ImageJ plugin design and code are explained in the Supplementary Note S5. The following section describes performance and analysis of the trained denoising ML models when applied to mixed Poisson-Gaussian noise in fluorescence microscopy images.

3 Results and Discussion

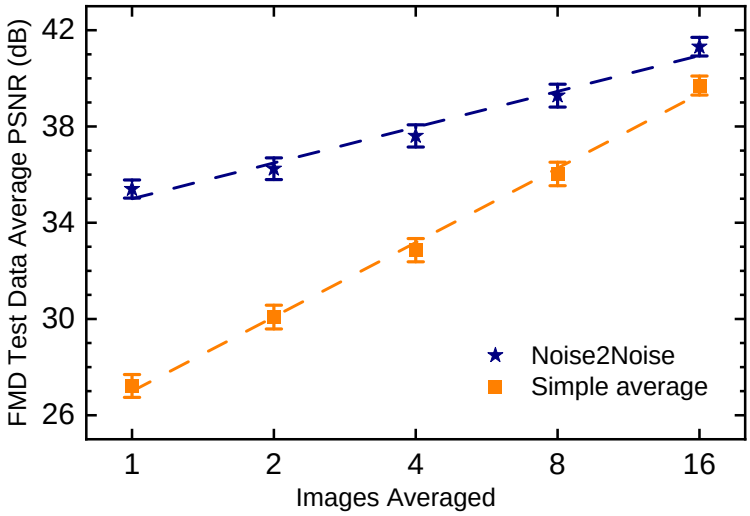

Figure 3: Average PSNR of the test-mix dataset which contains 48 images of the MPG noise samples from different microscopes in the FMD dataset. Here the error bars indicate one standard error below and above the average value, respectively.

The Noise2Noise model is trained with different initial learning rates (ILRs). The training and testing loss (mean square error) are provided in the Supplementary Note S2. We evaluate the trained ML model performance on test-mix data by selecting four images in each microscopy modality (48 images), as seen in Figure 3. Results show an average PSNR improvement of **8.13 dB** in the denoised images compared to the average PSNR of the raw images (Images Averaged = 1) in the test-mix data.

For deeper understanding along with more detailed analysis on these results, we perform a simulation that maps the fluorescence microscopy modality and corresponding noise characteristics of the acquired image by estimating the noise variance parameters (a and b) using the algorithm provided in [33] on the FMD dataset. The results are tabulated in Table 1.

Modality	**Sample**	**Estimated** a	**Estimated** b
CLSM	BPAE (Nuclei)	1.76×10^{-2}	-2.06×10^{-4}
CLSM	BPAE (F-actin)	1.48×10^{-2}	-0.65×10^{-4}
CLSM	BPAE (Mito)	2.01×10^{-2}	-1.43×10^{-4}
CLSM	Zebrafish	9.11×10^{-2}	-2.62×10^{-4}
CLSM	Mouse Brain	1.99×10^{-2}	-0.68×10^{-4}
TPEF	BPAE (Nuclei)	3.41×10^{-2}	-7.01×10^{-4}
TPEF	BPAE (F-actin)	2.66×10^{-2}	-0.88×10^{-4}
TPEF	BPAE (Mito)	4.15×10^{-2}	-2.10×10^{-4}
TPEF	Mouse Brain	3.36×10^{-2}	-5.67×10^{-4}
WF	BPAE (Nuclei)	7.88×10^{-4}	2.03×10^{-4}
WF	BPAE (F-actin)	19.15×10^{-4}	2.36×10^{-4}
WF	BPAE (Mito)	9.59×10^{-4}	2.53×10^{-4}

Table 1: Estimation of noise parameters (a and b) using the FMD dataset. The estimated a and b are average values of raw noisy images from 20 different FOVs (50 images within each FOV). CLSM: confocal laser scanning microscopy; TPEF: two-photon excitation microscopy; WF: widefield microscopy.

From Table 1, the estimated Gaussian noise variance (b) for the confocal and two-photon microscope is negative, which is due to an offset in the noisy image. More details can be found in [33] section II (A). For these cases, we set the estimated Gaussian variance to zero. From Table 1, we also observe that the estimated Poisson noise variance (a) in confocal and two-photon microscopy is dominating over the Gaussian noise variance ($a \gg b$). In the case of widefield microscopy, both noise parameters exist ($a, b > 0$), which indicates the mixture of Poisson-Gaussian noise. Also, the Gaussian variance (b) is in the range of $\sim$ 2e-4, whereas the Poisson variance (a) varies depending on the signal intensity. This behavior is observed in the widefield microscope, where the a value of BPAE (F-actin) is much larger when compared to the a value of BPAE (Mitochondria) and BPAE (Nuclei). The estimation of these noise parameters, i.e., the Poisson variance (a) and the Gaussian variance (b), shows a clear map between different noises in each microscope modality. Hence, we broadly divide the microscopy modalities into two categories: first, confocal and two-photon microscopy with Poisson-dominated noise; second, widefield microscopy with MPG noise. For each class of noise type, we show the trained model performance on a single noisy image and the statistics on the complete test-mix dataset.

In this work, we first select a single test image with Poisson-dominated noise, which is captured by a confocal microscope, and another test image with MPG noise, which is captured by a widefield microscope. Here these images are considered as specific images that do not reflect the complete test dataset.

Figure 4 illustrates a zebrafish embryos image with Poisson-dominated noise [EGFP labeled Tg (sox10:megfp) zebrafish at two days post fertilization] captured with a commercial confocal microscope (More details are provided in the Supplementary Note S1). Similarly, Figure 5 shows an image of a BPAE cell (slide #1, F36924 that contains Nuclei, F-actin, and Mitochondria) with MPG noise captured by a commercial widefield microscope. In these figures, the target intensity images were generated by taking the average of 50 noisy images in the same FOV. We observe that our ML-based Noise2Noise denoising method reduces both Poisson-dominated and MPG noise significantly by providing a PSNR improvement of 10.16 dB and 10.5 dB in zebrafish and BPAE cell, respectively. From these images, it is clear that the denoised images also look similar to the respective target images in both cases (SSIM values are 0.8384 and 0.8320, respectively).

Second, the PSNR distribution of the raw images from the test-mix data and the corresponding denoising results using our ML models are provided in Figure 6. Figure 6 shows the trained Noise2Noise ML model (with an ILR of 5e-4) image denoising performance with complete statistics on the test-mix dataset raw images (48) from our FMD dataset. Also, for image denoising, we train other ML model (requires a clean target), which is a DnCNN ML model that can estimate the residual (only noise) for a given noisy input image. Here, the denoised image is obtained by subtracting the estimated residual from the noisy input data. Figure 6 also shows complete statistics on the test-mix data using the DnCNN architecture ML model with the same ILR of 5e-4. The ML model with Noise2Noise architecture has a better PSNR when compared to the ML model trained with the DnCNN architecture. Hence, the Noise2Noise model demonstrates superior denoising performance compared to the other ML model, and it does not require the clean target image during the training phase. From the raw data distribution, it is evident that there exist two distinct groups of data corresponding to Poisson-dominated noise and MPG noise. The average PSNR of the raw images (including Poisson-dominated noise and MPG noise images) and the corresponding denoised images are 27.22 dB and 35.40 dB, respectively. From Figure 3, the PSNR improvement using the Noise2Noise method is comparable to that of an average PSNR of 8 noisy raw images in the same FOV (with PSNR of 36.02 dB) and thereby reducing the acquisition time approximately by eight fold.

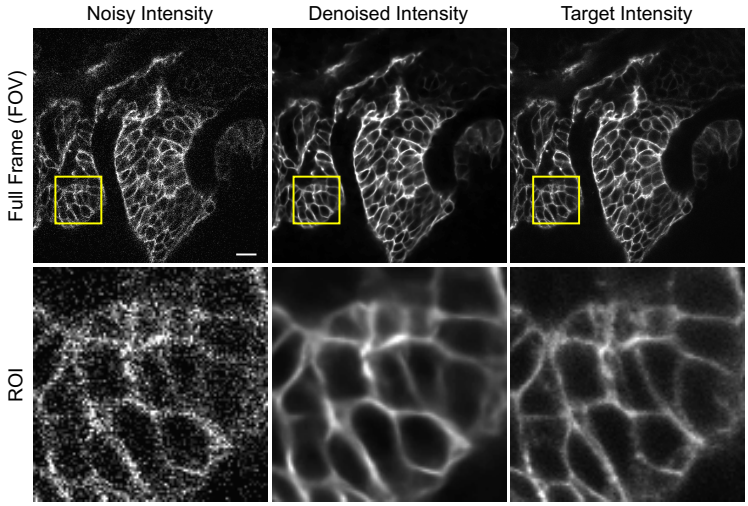

Figure 4: Image denoising results using the Noise2Noise ImageJ plugin on a fixed zebrafish embryos [EGFP labeled Tg(sox10:megfp) zebrafish at two days post fertilization] captured with confocal microscopy (pixel dwell time of 2 μs and the pixel width of 300 nm, with 10% of power and PMT gain of 140, frame rate is 2, excitation and emission wavelengths are 488 nm and 509 nm respectively). The top row indicates the full-frame (of size 512 $\times$ 512) of noisy input, denoised output, target, and the bottom row indicates the region of interest (ROI: marked in the yellow square of size 100 $\times$ 100) from the respective top row images.

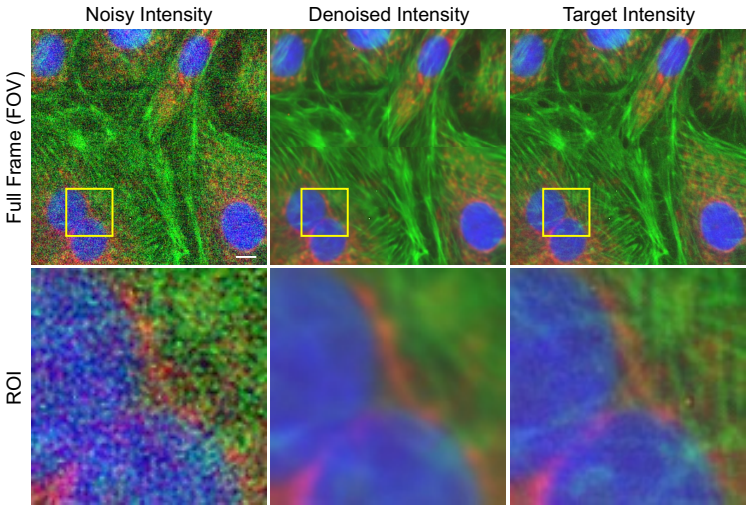

Figure 5: Image denoising results using the Noise2Noise ImageJ plugin on the BPAE cell [labeled with MitoTracker Red CMXRos (mitochondria), Alexa Fluor 488 phalloidin (F-actin), and DAPI (nuclei) with excitation/emission wavelengths are 561/599, 488/512 and 405/461, respectively; Invitrogen FluoCells prepared slide#1: F36924] captured with the widefield microscopy (exposure time of 200 ms, frame rate of 10, CCD gain of 46 and the pixel width is 170 nm). The top row indicates the full-frame (of size 512 $\times$ 512) of noisy input, denoised output, target, and the bottom row indicates the region of interest (ROI: marked in the yellow square of size 100 $\times$ 100) from the respective top row images. We imaged three times for the same FOV, each time with a different filter block (DAPI for nuclei, FITC for F-actin, TRITC for mitochondria), to acquire the multi-channel (color) fluorescence image of the cells.

Finally, we compare the performance of our ML based image denoising methods with the existing image denoising methods. This PSNR improvement is more substantial than existing image denoising methods like PURE-LET, VST+NLM, and VST+BM3D, as shown in Figure 7. Figure 7 shows the existing denoising methods performed on images presented with either Poisson or Gaussian noise. In contrast, our ML based image denoising methods enable superior denoising performance regardless of the noise distribution and the microscopy modality. The PURE-LET, VST+NLM and VST+BM3D image denoising methods perform well only for the images with Poisson-dominated

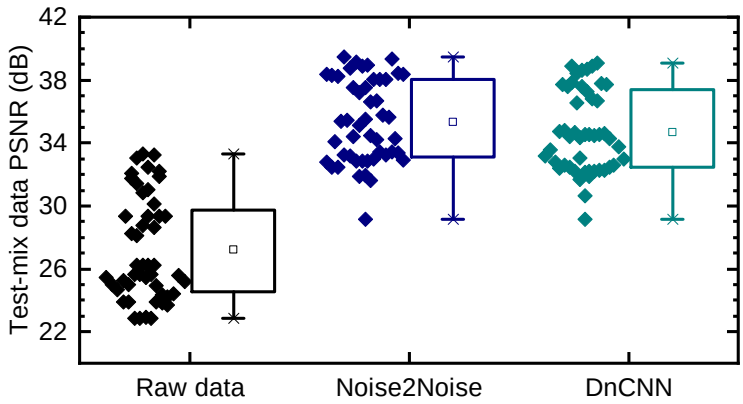

Figure 6: Test-mix dataset (which contains 48 images of the MPG noise samples from different microscopes in the FMD dataset) raw images PSNR box-plot to compare Noise2Noise and DnCNN ML models image denoising performance at a fixed Initial learning rate (ILR: 5e-4). Box-plot contains min, max, 25th, and 75th percentile values respectively. Here the square box indicates the mean value. A scattered plot (left to box plot) indicates the true PSNR values used for the statistics.

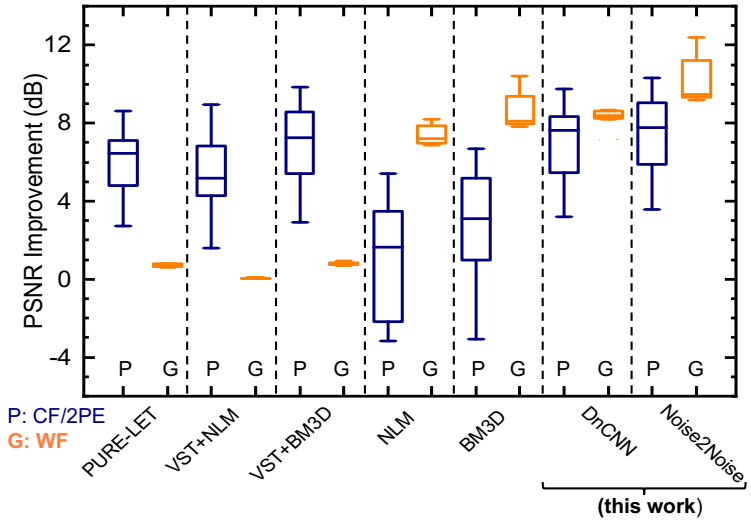

Figure 7: PSNR improvement (denoised image PSNR - raw image PSNR) using different image denoising methods on the raw data with Poisson and MPG noise statistics independently. Here PURE_LET, VST+NLM, VST+BM3D, NLM and BM3D are existing image methods and DnCNN and Noise2Noise are our demonstrated supervised trained CNN methods. Blue and Orange color statistics indicate the box plot results of the PSNR improvement for the Poisson-dominated and MPG noises, respectively. Supervised CNN based methods outperform compared to existing image methods independent of noise characteristics.

noise, while the NLM and BM3D image denoising methods perform well only for the images with Gaussian noise. In contrast, the ML based image denoising methods (Noise2Noise and DnCNN) perform well for images with both Poisson-dominated noise and MPG noise.

Our Noise2Noise ML model image denoising method shows significant improvements over existing methods and can provide better results for almost any fluorescence microscopy image. To increase the impact and integrate it into existing workflows, we package our method as an ImageJ plugin, a popular image analysis tool used by biologists. Our developed ImageJ plugin has two advantages. First, the plugin can perform image denoising independent of microscopy modality since the ML model was trained with MPG noise images with superior performance when compared to the existing methods. Second, the plugin can denoise "instantly" (i.e., in real-time) for any fluorescence microscopy noisy image. More details about the ImageJ image denoising plugin are provided in the Supplementary Note S5. In addition, Supplementary Note S3 shows the qualitative and quantitative comparison of the existing image denoising ImageJ

plugins (existing and our ML-based image denoising methods) available with the trained Nosie2Noise ML model on a test image from the FMD dataset. The image denoising time in our ImageJ plugin using a central processing unit (CPU) and a graphics processing unit (GPU) are 960 ms and 80 ms, respectively. Therefore, the image denoising speed is faster by 12-fold when using GPU compared to CPU. Overall, the ImageJ plugin can be used to denoise fluorescence images that are obtained with low laser power or at a fast acquisition rate. The source code for image denoising including ML models training and testing in Keras and creation of our ImageJ plugin are provided[2].

4 Model Validation

Model validation outside of the training FMD dataset images and noise distributions is crucial to demonstrate the method's reliability and utility. To evaluate the validity, we evaluate model performance (a) on the W2S fluorescence microscopy dataset to compare performance to other analytic and ML-based image denoising methods outside of the FMD training dataset, (b) on the W2S dataset to evaluate denoising performance on samples outside of the training dataset the PSNR distribution and on a flavin-based autofluorescent bacteria images to evaluate performance outside of the training dataset contrast distribution, and (c) discuss the risk and limitations of applying the trained CNN model to higher-dimensional data.

4.1 Denoising performance on the W2S dataset

To validate the trained Noise2Noise ML model on images that differ from the FMD dataset, we evaluate denoising plugin performance on 360 widefield fluorescence microscopy images of varying PSNR from the W2S dataset [24]. We also investigate whether the FMD trained ML plugin model applied to the W2S dataset introduces performance degradation or artifacts compared to three of the best performing self-supervised (no external training) ML methods: N2V [17], VST+BM3D [10] and BM3D [6].

The average PSNR improvement and computation time for the different methods applied to the W2S dataset is given in Table 2. For Noise2Noise, the trained ImageJ plugin presented here was used; for the VST+BM3D method (in sequential steps: apply VST transform [10], use the BM3D method [6] and apply inverse VST transformation [10]) for image denoising a and BM3D method [6] MATLAB code was used and for the N2V method an ImageJ plugin was used [34] with two sets of training parameters to explore performance versus computation time trade-offs ("fast" with 10 epochs and 200 steps, "default with 300 epochs and 200 steps). In summary, the supervised Noise2Noise CNN plugin that was trained on the FMD dataset yields higher PSNR improvements compared to self-supervised ML methods. Additionally, the Noise2Noise method only requires convolutions to perform denoising inference, while self-supervised require re-training (or self-training) on every image. The resulting denoising computation time is therefore significantly shorter for the pre-trained Noise2Noise plugin when computed using either a CPU (Intel Xeon E5-2680 CPU) or GPU (Nvidia GeForce GTX 1080 Ti). While the Noise2Noise plugin, VST+BM3D, and BM3D image denoising methods were performed on the entire dataset to find the mean PSNR improvement over the entire W2S dataset, the N2V method requires a long training time per-image, therefore estimates of the mean improvement (and standard error of the estimate) are presented.

Method	**Training**	**ΔPSNR (dB)**	**Time/frame (s) CPU/GPU**
Noise2Noise (this work)	Supervised, FMD	8.25	3.07/0.08
VST+BM3D	Self-supervised	4.37	5.67/None
BM3D	Self-supervised	4.74	4.65/None
Noise2Void (fast settings)	Self-supervised	4.6 (0.7)*	15600/400
Noise2Void (default settings)	Self-supervised	4.8 (0.7)*	478800/12000

Table 2: Average PSNR improvement (ΔPSNR= denoised image PSNR - input image PSNR) and single-frame processing computation time (Intel Xeon E5-2680 CPU and Nvidia GeForce GTX 1080 Ti GPU) of Noise2Noise, VST+BM3D, BM3D and Noise2Void image denoising methods when applied to the 360 wide-field fluorescence microscopy images in the W2S dataset. *Estimated mean (and standard error of the mean) from 10 random noisy images from the W2S dataset.

[2] `https://github.com/ND-HowardGroup/Instant-Image-Denoising`

A qualitative comparison of denoising performance is given in Figure 8 for a representative image from the W2S dataset. Here the target image is an average of 400 noisy images in the same FOV and is provided by the W2S dataset. Additional examples are available in Supplementary Note S4 for single-channel and multi-channel test images from the W2S dataset. From these examples, the images qualitatively exhibit lower noise and faithfully represent the target image.

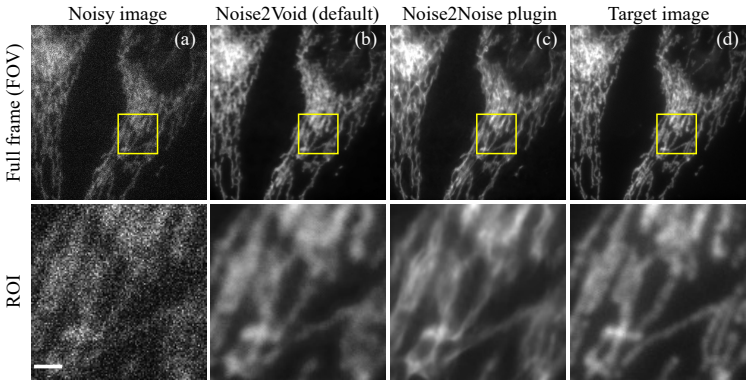

Figure 8: Comparison of the ML-based image denoising plugins in ImageJ. (a) Noisy image, (b) denoised by Noise2Void ImageJ plugin (default training), (c) denoised by the Noise2Noise ImageJ plugin, and (d) ground truth image of a test mage from the W2S dataset. The top row indicates the full-frame (of size 512 $\times$ 512) and the bottom row indicates the region of interest (ROI: marked in the yellow square of size 120 $\times$ 120) from the respective top row images. Scale bar: 10 μm.

4.2 Out-of-distribution validation

Machine learning models are optimized for inference data distributions that are similar to the training data distribution. The FMD training dataset is comprised of images with a PSNR of the range of 22-40 dB with contrast ratios (defined as $(I_{max} - I_{min})/(I_{max} + I_{min})$) of approximately 0.94. To evaluate model performance outside of the training dataset distribution, we validate the trained Noise2Noise plugin on images of varying PSNR and contrast in this section.

The FMD training dataset includes MPG noise from the FMD dataset with PSNR in the range of 22-40 dB; therefore, the model is expected to perform best when for input images with similar noise levels and input images with lower PSNR outside of the training dataset range can expect lower levels of improvement or artifacts [35]. The trained Noise2Noise model presented here, however, is designed to properly denoise images with out-of-distribution noise using "bias-free" (i.e., no bias term) convolution operations [35]. By not including a bias term in the convolution or batch-norm regularization layers, artifacts originating from denoising images with lower PSNR than the training dataset are avoided. In this section, we evaluate denoising performance on the W2S dataset for images with out-of-distribution PSNR.

To validate that the bias-free Noise2Noise model can appropriately denoise low PSNR out-of-distribution images, we evaluated denoising performance on such data using the method described in [35]: Gaussian noise is synthetically added to images from the W2S dataset to yield images with PSNR on the range from 5-40 dB, and denoising performance can thereby evaluated as a function of input PSNR. Results are shown in Figure 9 (dark blue symbols and line) show significant PSNR improvements (> 9.6 dB) for the W2S test data outside of the trained PSNR distribution (< 22 dB). For images with the trained noise distribution range (input PSNR of 22-40 dB, light blue shading), the average PSNR improvement is ≈ 7.06 dB. PSNR improvement decreases as the input PSNR value increases within the trained noise distribution, and this behavior is consistent with the results shown in Figure 3.

To validate denoising plugin performance for images with out-of-distribution contrast, we evaluate denoising performance for low contrast images (for illustration, defined as $(I_{max} - I_{min})/(I_{max} + I_{min})$). This metric defines contrast on range from 0 to 1, where 1 is ideal high contrast. The FMD dataset and W2S datasets are both high-contrast datasets with average contrast measures of 0.94 and 0.95, respectively. To evaluate denoising performance for low-contrast fluorescence microscopy images, we prepared an experimental dataset of 600 images (time lapse data) of *Myxococcus xanthus* bacteria auto-fluorescence using widefield microscopy. This sample and imaging method was chosen to produce experimental results with low average contrast (0.52), far outside of the training distribution. Sample preparation and microscope configuration is described in [36, 37]: unlabeled bacteria are illuminated with 458 nm laser to excite flavins in the bacteria, resulting in autofluorescence. The microscope is configured as an widefield epi-fluorescence microscopy

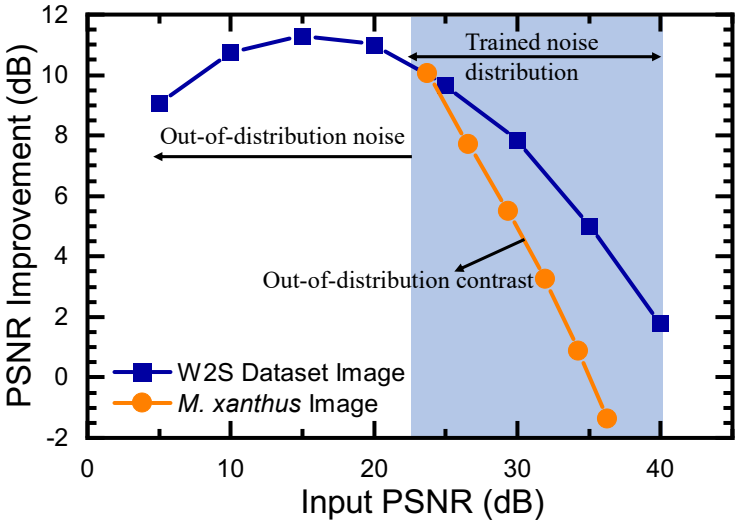

Figure 9: PSNR improvement using the Noise2Noise plugin on test data with acquired PSNR (blue) and contrast (orange) that are out-of-distribution relative to the PSNR and and contrast of the FMD training dataset. Out-of-distribution low PSNR images were obtained by adding synthetic noise to the W2S dataset. Out-of-distribution contrast images taken from the *M. xanthus* autoflourescence dataset and averaged (1, 2, 4, 8, 16, and 32 frames with the same FOV) to obtain images with varying PSNR.

using a single dichroic mirror as both excitation and emission filter, resulting in the excitation light leakage and/or back scattering being detected by the imaging EM-CCD. The resulting contrast is therefore relatively low due to the low emission intensity from unlabeled autofluorescence and the weak background suppression from using a single excitation blocking filter. A representative result is shown in Figure 10 (a), (b) and (c), depicting the out-of-distribution noisy *M. xanthus* image, denoised using the Noise2Noise ML model, target images respectively. The target image is defined as the average of 600 noisy images in the same FOV. In this example, applying the Noise2Noise plugin on a single noisy frame from this dataset increases PSNR from 23.72 dB to 33.78 dB.

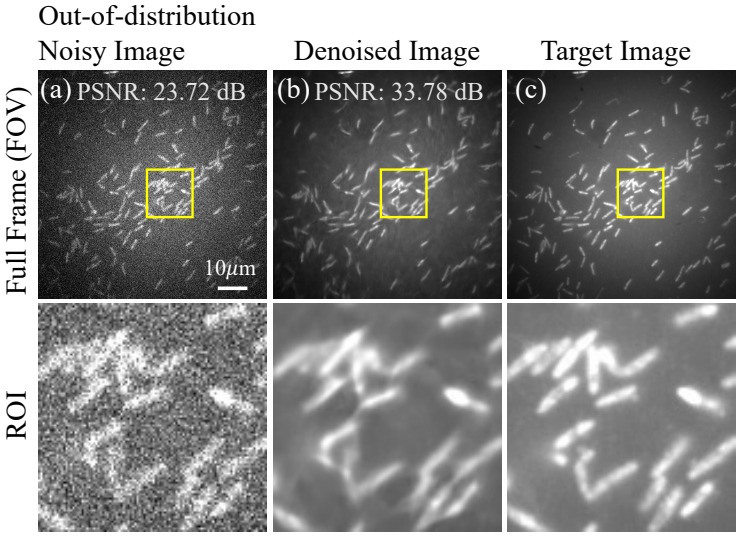

Figure 10: Noise2Noise plugin denoising on *M. xanthus* autoflourescence images captured using a widefield fluorescence microscope. Input noisy image (a), the image denoised using the Noise2Noise plugin (b), and the target image (c) are represented by columns. The full frame FOV is 512 × 512 pixels; the selected ROI (yellow box) is 150×150 pixels. Excitation wavelength 458 nm; emission wavelength 500-580 nm; sample excitation power 0.1 mW; integration time 100 ms; pixel size 150nm.

Denoising plugin performance for these low-contrast images can be evaluated as a function of input image PSNR by averaging sequential frames prior to running the denoising plugin. The resulting denoising improvement is given by Figure 9 (orange symbols). The Noise2Noise plugin achieves high-performance denoising on single-frame, low-PSNR microscopy images even when the contrast is far out-of-distribution. While both low-contrast and high-contrast images show decreased denoising performance as input PSNR increases (Figure 3 and Figure 9), denoising improvement decrease at a faster rate for low-contrast images. This defines the out-of-distribution limit for the trained Noise2Noise CNN: exceedingly low-contrast yet high-PSNR input images may have artifacts introduced in to the image.

In summary, the Noise2Noise plugin demonstrates high-performance denoising on out-of-distribution low-PSNR and low-contrast ratio images taken from datasets external to the training dataset. By using bias-free convolutions, artifacts from low-PSNR images are avoided. The trained CNN performs well on low-PSNR and low-contrast images outside of the dataset, however the CNN introduces artifacts that degrade images in the *M. xanthus* dataset when contrast is low (<0.6) and PSNR $\gtrsim$ 36 dB. However, such high-PSNR images would likely not be candidates for denoising, and therefore applying the Noise2Noise plugin is reasonable for practical experimental fluorescence microscopy images. The Noise2Noise plugin shows exceptional performance (>8 dB) for single-frame, low PSNR images (<25 dB) with typical experimental contrast (at least >0.5).

4.3 Out-of-model data: higher dimensionality data

Modern microscopy techniques commonly acquire data in four or more dimensions (three spatial, one temporal, and possibly more channels) and represent the data as 2D sections per channel at discrete acquisition times. It is therefore important to consider the model in the context of data with dimensionality outside-of-distribution from the training dataset. To address this, the model was trained on 2D sections of 3D spatial data of one to three channels, captured at specific times during dynamic intravital processes (e.g., blood flow and respiration). The noise of each individual channel can be described as MPG noise, and therefore it is appropriate to apply the method to 2D sections of higher dimensional data using the MPG denoising method. Since the trained Noise2Noise model performs denoising solely using information from a single 2D plane of the 3D sample, potential further performance improvements could be achieved by developing and training models using complete high-dimension data (e.g., the entire 3D stack in time [38]). Such an approach relies on significant computational improvements and would effectively amplify signals correlated in 3D space and time, leading to even improved denoising performance.

5 Conclusion

In this paper, we have presented an open-source software tool for fluorescence microscopy image denoising in real-time (within tens of milliseconds) to denoise mixed Poisson-Gaussian noise. The new image denoising technique using the trained ML models (Noise2Noise and DnCNN) allows imaging eight times faster than the fundamental speed limit of fluorescence microscopy while maintaining the same PSNR. The system was trained using the experimentally obtained FMD dataset to develop a neural network that effectively removes MPG noise presented in widefield, confocal, and two-photon fluorescence microscopy images across a variety of cell types. We have shown here that the proposed Noise2Noise ML model has outstanding improvements in PSNR ($\sim$ 8.1 dB) compared to the existing image denoising methods, performs denoising faster, and has been validated on a wide variety of images with ranges of input PSNR and contrast levels commonly seen in fluorescence microscopy.

Disclosures

The authors declare no conflicts of interest.

Acknowledgments.

We thank Prof. Joshua D. Shrout, Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, for providing *M. xanthus* cells for our experiments. Yide Zhang's research was supported by the Berry Family Foundation Graduate Fellowship of Advanced Diagnostics & Therapeutics (AD&T), University of Notre Dame. The authors further acknowledge the Notre Dame's Center for Research Computing (CRC) for the use of GPUs to train the ML models and the Integrated Imaging Facility (NDIIF) for the use of the Nikon A1R MP confocal microscope and Nikon Eclipse 90i wide-field microscope in NDIIF's Optical Microscopy Core.

References

[1] Jeff W Lichtman and José-Angel Conchello. Fluorescence microscopy. *Nature Methods*, 2(12):910, 2005.

[2] Ernst HK Stelzer. Contrast, resolution, pixelation, dynamic range and signal-to-noise ratio: fundamental limits to resolution in fluorescence light microscopy. *Journal of Microscopy*, 189(1):15–24, 1998.

[3] Varun Mannam, Yide Zhang, Xiaotong Yuan, Cara Ravasio, and Scott S Howard. Machine learning for faster and smarter fluorescence lifetime imaging microscopy. *Journal of Physics: Photonics*, 2(4):042005, 2020.

[4] Yide Zhang, Aamir A Khan, Genevieve D Vigil, and Scott S Howard. Investigation of signal-to-noise ratio in frequency-domain multiphoton fluorescence lifetime imaging microscopy. *Journal of the Optical Society of America A*, 33(7):B1–B11, 2016.

[5] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 60–65, 2005.

[6] K Dabov, A Foi, V Katkovnik, and K. Egiazarian. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[7] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[8] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *IEEE International Conference on Computer Vision*, pages 479–486, 2011.

[9] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2862–2869, 2014.

[10] Markku Makitalo and Alessandro Foi. Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise. *IEEE Transactions on Image Processing*, 22(1):91–103, 2012.

[11] Charles-Alban Deledalle, Florence Tupin, and Loic Denis. Poisson NL means: Unsupervised non local means for poisson noise. In *IEEE International Conference on Image Processing*, pages 801–804, 2010.

[12] Joseph Salmon, Zachary Harmany, Charles-Alban Deledalle, and Rebecca Willett. Poisson noise reduction with non-local PCA. *Journal of Mathematical Imaging and Vision*, 48(2):279–294, 2014.

[13] Florian Luisier, Thierry Blu, and Michael Unser. Image denoising in mixed poisson-gaussian noise. *IEEE Transactions on Image Processing*, 20(3):696–708, 2010.

[14] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, et al. Content-Aware Image Restoration: pushing the limits of fluorescence microscopy. *Nature Methods*, 15(12):1090–1097, 2018.

[15] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[16] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. *arXiv:1803.04189*, 2018.

[17] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2Void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2129–2137, 2019.

[18] Alexander Krull, Tomáš Vičar, Mangal Prakash, Manan Lalit, and Florian Jug. Probabilistic Noise2Void: Unsupervised content-aware denoising. *Frontiers in Computer Science*, 2:5, 2020.

[19] Jason Lequyer, Reuben Philip, Amit Sharma, and Laurence Pelletier. Noise2Fast: Fast self-supervised single image blind denoising. *arXiv:2108.10209*, 2021.

[20] Wesley Khademi, Sonia Rao, Clare Minnerath, Guy Hagen, and Jonathan Ventura. Self-supervised poisson-gaussian denoising. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2131–2139, 2021.

[21] Anna S Goncharova, Alf Honigmann, Florian Jug, and Alexander Krull. Improving blind spot denoising for microscopy. In *European Conference on Computer Vision*, pages 380–393. Springer, 2020.

[22] Varun Mannam, Yide Zhang, Yinhao Zhu, and Scott S. Howard. Fluorescence microscopy denoising (FMD) dataset, 2019.

[23] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11710–11718, 2019.

[24] Ruofan Zhou, Majed El Helou, Daniel Sage, Thierry Laroche, Arne Seitz, and Sabine Süsstrunk. W2S: microscopy data with joint denoising and super-resolution for widefield to SIM mapping. In *European Conference on Computer Vision*, pages 474–491. Springer, 2020.

[25] James Pawley. *Handbook of biological confocal microscopy*, volume 236. Springer Science & Business Media, 2006.

[26] Winfried Denk, James H Strickler, and Watt W Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248(4951):73–76, 1990.

[27] Peter J Verveer, Mark J Gemkow, and Thomas M Jovin. A comparison of image restoration approaches applied to three-dimensional confocal and wide-field fluorescence microscopy. *Journal of Microscopy*, 193(1):50–61, 1999.

[28] Johannes Schindelin, Curtis T Rueden, Mark C Hiner, and Kevin W Eliceiri. The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development*, 82(7-8):518–529, 2015.

[29] Rueden Curtis. ImageJ Tensorflow library. `https://github.com/imagej/imagej-tensorflow`, 2017.

[30] Jérôme Boulanger, Charles Kervrann, Patrick Bouthemy, Peter Elbau, Jean-Baptiste Sibarita, and Jean Salamero. Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE Transactions on Medical Imaging*, 29(2):442–454, 2009.

[31] Estibaliz Gómez-de Mariscal, Carlos García-López-de Haro, Laurène Donati, Michael Unser, Arrate Muñoz-Barrutia, and Daniel Sage. DeepImageJ: A user-friendly plugin to run deep learning models in ImageJ. *bioRxiv*, page 799270, 2019.

[32] Lucas von Chamier, Romain F Laine, Johanna Jukkala, Christoph Spahn, Daniel Krentzel, Elias Nehme, Martina Lerche, Sara Hernández-Pérez, Pieta K Mattila, Eleni Karinou, et al. Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nature Communications*, 12(1):1–18, 2021.

[33] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008.

[34] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. ImageJ Noise2Void image denoising. `https://imagej.net/plugins/n2v`, 2018.

[35] Sreyas Mohan, Zahra Kadkhodaie, Eero P Simoncelli, and Carlos Fernandez-Granda. Robust and interpretable blind image denoising via bias-free convolutional neural networks. *arXiv:1906.05478*, 2019.

[36] Vignesh Sundaresan, Allison R Cutri, Jarek Metro, Chinedu S Madukoma, Joshua D Shrout, Anthony J Hoffman, Katherine A Willets, and Paul W Bohn. Potential dependent spectroelectrochemistry of electrofluorogenic dyes on indium-tin oxide. *Electrochemical Science Advances*, page e2100094, 2021.

[37] Cameron W Harvey, Chinedu S Madukoma, Shant Mahserejian, Mark S Alber, and Joshua D Shrout. Cell division resets polarity and motility for the bacterium *Myxococcus xanthus*. *Journal of Bacteriology*, 196(22):3853–3861, 2014.

[38] Mikhail Papkov, Kenny Roberts, Lee Ann Madissoon, Jarrod Shilts, Omer Bayraktar, Dmytro Fishman, Kaupo Palo, and Leopold Parts. Noise2Stack: Improving image restoration by learning from volumetric data. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 99–108. Springer, 2021.