

A Poisson-Gaussian Denoising Dataset with Real Fluorescence Microscopy Images

Yide Zhang*, Yinhao Zhu*, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, Scott Howard
 University of Notre Dame
 Notre Dame, IN 46556, USA

{yzhang34, yzhu10, enichol3, qwang9, szhang8, csmith67, showard}@nd.edu

Abstract

Fluorescence microscopy has enabled a dramatic development in modern biology. Due to its inherently weak signal, fluorescence microscopy is not only much noisier than photography, but also presented with Poisson-Gaussian noise where Poisson noise, or shot noise, is the dominating noise source. To get clean fluorescence microscopy images, it is highly desirable to have effective denoising algorithms and datasets that are specifically designed to denoise fluorescence microscopy images. While such algorithms exist, no such datasets are available. In this paper, we fill this gap by constructing a dataset - the Fluorescence Microscopy Denoising (FMD) dataset - that is dedicated to Poisson-Gaussian denoising. The dataset consists of 12,000 real fluorescence microscopy images obtained with commercial confocal, two-photon, and wide-field microscopes and representative biological samples such as cells, zebrafish, and mouse brain tissues. We use image averaging to effectively obtain ground truth images and 60,000 noisy images with different noise levels. We use this dataset to benchmark 10 representative denoising algorithms and find that deep learning methods have the best performance. To our knowledge, this is the first real microscopy image dataset for Poisson-Gaussian denoising purposes and it could be an important tool for high-quality, real-time denoising applications in biomedical research.

1. Introduction

Fluorescence microscopy is a powerful technique that permeates all of biomedical research [15]. Confocal [23], two-photon [9], and wide-field [26] microscopes are the most widely used fluorescence microscopy modalities that are vital to the development of modern biology. Fluorescence microscopy images, however, are inherently noisy because the number of photons captured by a microscopic

*Equal contribution.

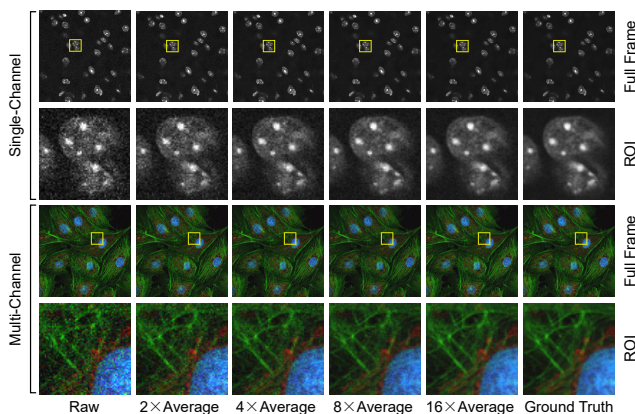


Figure 1. Examples of images with different noise levels and ground truth. The single-channel (gray) images are acquired with two-photon microscopy on fixed mouse brain tissues. The multi-channel (color) images are obtained with two-photon microscopy on fixed BPAE cells. The ground truth images are estimated by averaging 50 noisy raw images.

detector, such as a photomultiplier tube (PMT) or a charge coupled device (CCD) camera, is extremely weak ($\sim 10^2$ per pixel) compared to that in photography ($\sim 10^5$ per pixel [21]). Consequently, the measured optical signal in fluorescence microscopy is quantized due to the discrete nature of photons, and fluorescence microscopy images are dominated by Poisson noise, instead of Gaussian noise that denominates in photography [22]. One way to obtain clean images is to increase the power of the excitation laser or lamp, but the excitation power is not only limited by the dosage of light a biological sample can receive, but also fundamentally limited by the fluorescence saturation rate; i.e., the fluorescence signal will stop to increase when the excitation power is too high [32]. Alternatively, one can get clean images by increasing the imaging time, e.g., pixel dwell time, exposure time, number of line or frame averages; this, however, may cause photodamage to the sample. Moreover, for dynamic or real-time imaging, increasing the

imaging time may be impossible since each image has to be captured within tens of milliseconds. Therefore, developing an algorithm to effectively denoise (reduce the noise in) a fluorescence microscopy image is of great importance to biomedical research. Meanwhile, a high-quality denoising dataset is necessary to benchmark and evaluate the effectiveness of the denoising algorithm.

Most of the image denoising algorithms and datasets are created for Gaussian noise dominated images, with a recent focus on denoising with real noisy images, such as smart phones [1] or digital single-lens reflex camera (DSLR) images [24]. However, there is a lack of a reliable Poisson noise dominated denoising dataset comprising of *real* fluorescence microscopy images. The goal of this work is to fill this gap. More specially, we create a Poisson-Gaussian denoising dataset - the Fluorescence Microscopy Denoising (FMD) dataset - consisting of 12,000 real noisy microscopy images which cover the three most widely used imaging modalities, i.e., confocal, two-photon, and wide-field, as well as three representative biological samples including cells, zebrafish, and mouse brain tissues. With high-quality commercial microscopy, we use image averaging to effectively obtain ground truth images and noisy images with five different noise levels. Some image averaging examples are shown in Figure 1. We further use this dataset to benchmark classic denoising algorithms and recent deep learning models, with or without ground truth. Our FMD dataset is publicly available¹, including the code for the benchmark². To our knowledge, this is the first dataset constructed from real noisy fluorescence microscopy images and designed for Poisson-Gaussian denoising purposes.

2. Related Work

There are consistent efforts in constructing denoising dataset with real images to better capture the real-world noise characteristics and evaluate denoising algorithms, such as RENOIR [4], Darmstadt Noise Dataset [24], Smartphone Image Denoising Dataset [1], and PolyU Dataset [28]. Those datasets contain real images taken from either DSLR or smartphones with different ISOs and different number of scenes. The dominating noise in those images is Gaussian or Poisson-Gaussian in real low-light conditions. However, there is no dedicated dataset for Poisson noise dominated images, which are inherently different from Gaussian denoising datasets. This work is dedicated for fluorescence microscopy denoising where the images are corrupted by Poisson-Gaussian noise; in particular, Poisson noise, or shot noise, is the dominant noise source.

Image averaging is the most used method to obtain ground truth images when constructing denoising dataset.

The main efforts are spent on image pre-processing, such as image registration to remove the spatial misalignment of an image sequence with the same field of view (FOV) [3, 1], intensity scaling due to the changes of light strength or analog gain [24], and methods to cope with clipped pixels due to over exposure or low-light conditions [4]. The images captured by commercial microscopes in our dataset turns out to be well aligned, and the analog gain is carefully chosen to avoid clipping and to utilize the full dynamic range.

There are two main approaches to denoise an image corrupted by Poisson-Gaussian noise. One way is to directly apply an effective denoising algorithm, such as the PURE-LET method [17], which is designed to handle the Poisson-Gaussian denoising problem based on the statistics of the noise model. Another approach is using a nonlinear variance-stabilizing transformation (VST) to convert the Poisson-Gaussian denoising problem into a Gaussian noise removal problem, which is well studied with a considerable amount of effective denoising algorithms to choose from, such as NLM, BM3D, KSVD, EPLL, and WNNM [6, 8, 2, 33, 11] etc. The VST-based denoising process generally involves three steps. First, the noisy raw images are transformed using a VST designed for the noise model. In our case, we use the generalized Anscombe transformation (GAT) that is designed for Poisson-Gaussian noise [19]. The VST is able to remove the signal-dependency of the Poisson component, whose noise variance varies with the expected pixel value, and results in a modified image with signal-independent Gaussian noise only and a constant (unitary) noise variance. Next, a Gaussian denoising algorithm is applied to the transformed image. And finally, the Gaussian-denoised data is transformed back via an inverse VST algorithm, such as the exact unbiased inverse transformation [19], and the estimation of the noise-free image is obtained.

Recently there is an increasing interest in deep learning based methods for image denoising, where fully convolutional networks (FCNs) [16] are used for this image-to-image regression problem. With residual learning and batch normalization, DnCNN [30] reports better performance than traditional denoising methods such as BM3D. Further development towards blind image denoising includes incorporating non-uniform noise level map in the input of FFDNet [31], or noise estimation network as in CBDNet [12], or utilizing the non-local self-similarity in UDNet [13] and [25]. These methods all require clean images to supervise the training. There are also progress on denoising methods without paired clean images [7] using generative adversarial networks to learn the noise model. In [14], a Noise2Noise model is trained without clean images at all and outperforms VST+BM3D by almost 2dB on synthetic Poisson noise.

We perform intensive study of the noise statistics of the

¹<http://tinyurl.com/y6mwqcjs>

²<https://github.com/bmmi/denoising-fluorescence>

FMD dataset and show that the noise is indeed Poisson-dominated for two-photon and confocal microscopy, and has larger Gaussian component for wide-field microscopy. We then benchmark 10 representative denoising algorithms on the FMD dataset, and show better denoising performance with deep learning models than with traditional methods on the real noisy images.

3. Noise Modeling in Fluorescence Microscopy

The microscopy imaging system is modeled with a Poisson-Gaussian noise model [10, 19]. The model is composed of a Poisson noise component that accounts for the signal-dependent uncertainty, i.e., shot noise, and an additive Gaussian noise component which represents the signal-independent uncertainty such as thermal noise. Specifically, let z_i , $i = 1, 2, \dots, N$, be the measured pixel values obtained with a PMT or a CCD, and

$$z_i = y_i + n_i = y_i + n_p(y_i) + n_g, \quad (1)$$

where y_i is the ground truth and n_i is the noise of the pixel; the noise n_i is composed of two mutually independent parts, n_p and n_g , where n_p is a signal-dependent Poisson noise component that is a function of y_i , and n_g is a signal-independent zero-mean Gaussian component. Denoting $a > 0$ as the conversion or scaling coefficient of the detector, i.e., a single detected photon corresponds to a measured pixel value of a , and $b \geq 0$ as the variance of the Gaussian noise, we can describe the Poisson and Gaussian (normal) distributions as

$$(y_i + n_p(y_i))/a \sim \mathcal{P}(y_i/a), \quad n_g \sim \mathcal{N}(0, b). \quad (2)$$

Note that a is related to the quantum efficiency of the detector. Assuming that the Poisson and Gaussian processes are independent, the probability distribution of z_i is the convolution of their individual distributions, i.e.,

$$p(z_i) = \sum_{k=0}^{+\infty} \left(\frac{\left(\frac{y_i}{a}\right)^k}{k!} e^{-\frac{y_i}{a}} \times \frac{1}{\sqrt{2\pi b}} e^{-\frac{(z_i - ak)^2}{2b}} \right). \quad (3)$$

The denoising problem of a microscopy image is then to estimate the underlying ground truth y_i given the noisy measurement of z_i .

To denoise a fluorescence microscopy image, one can use algorithms that are specifically designed for Poisson-Gaussian denoising. A more common approach is using VST to stabilize the variance such that the denoising task can be tackled by a well-studied Gaussian denoising method. As a representative VST method, GAT transforms the measured pixel value z_i in the image to

$$f(z_i) = \frac{2}{a} \sqrt{\max\left(az_i + \frac{3}{8}a^2 + b, 0\right)}, \quad (4)$$

which stabilizes its noise variance to approximately unity, i.e., $\text{Var}\{f(z_i)\} \approx 1$. A Gaussian denoising algorithm, such as NLM and BM3D, can then be applied to $f(z_i)$ because its noise can be considered as a signal-independent Gaussian process with zero mean and unity variance. Once the denoised version of $f(z_i)$, denoted as $D(z_i)$, is obtained, an inverse VST is used to estimate the signal of interest y_i . However, simply applying an algebraic inverse f^{-1} to D will generally result in a biased estimate of y_i . An asymptotically unbiased inverse can mitigate the bias, but the denoising accuracy will be problematic for images with low signal levels, a common property of fluorescence microscopy images [29]. To address this problem, we use the exact unbiased inverse transformation, which can estimate the signal of interest accurately even at low signal levels [19]. In practice, since the exact unbiased inverse requires tabulation of parameters, one can employ a closed-form approximation of it [18], i.e.,

$$\tilde{I}(D) = \frac{1}{4}D^2 + \frac{1}{4}\sqrt{\frac{3}{2}}D^{-1} - \frac{11}{8}D^{-2} + \frac{5}{8}\sqrt{\frac{3}{2}}D^{-3} - \frac{1}{8}. \quad (5)$$

The closed-form approximation ensures the denoising accuracy while reducing the computational cost, and the estimated noise-free signal is $\tilde{y}_i = \tilde{I}[D(z_i)]$.

To evaluate and benchmark the performances of different denoising algorithms, a ground truth and images with various noise levels are needed, which can be obtained by averaging a series of noisy raw fluorescence microscopy images taken on the same FOV. In this work, the raw images are the immediate outputs of microscopy detectors, without any preprocessing. The averaging is performed after ensuring that no image shift larger than a half-pixel can be detected by an image registration algorithm. Since for different raw images, their Poisson-Gaussian random processes are independent, the average of S noisy raw images, v_i^S , can be written as

$$\begin{aligned} v_i^S &= \frac{1}{S} \sum_{j=1}^S z_i^j = \frac{a}{S} \sum_{j=1}^S \frac{y_i + n_p^j(y_i)}{a} + \frac{1}{S} \sum_{j=1}^S n_g^j \\ &\sim \frac{a}{S} \mathcal{P}\left(\frac{Sy_i}{a}\right) + \frac{1}{S} \mathcal{N}(0, Sb), \end{aligned} \quad (6)$$

where n_p^j and n_g^j are the noise realizations of the j -th noisy image. Based on the properties of Poisson and Gaussian distributions, the mean and variance of the averaged image, v_i^S , can be written as

$$\mathbb{E}[v_i^S] = y_i, \quad \text{Var}[v_i^S] = \frac{a}{S}y_i + \frac{b}{S}. \quad (7)$$

As the number of noisy images used for averaging increases, the noise of ground truth estimation, $\sqrt{\text{Var}[v_i^S]}$, decreases, while the ground truth signal, $\mathbb{E}[v_i^S]$ is invariant;

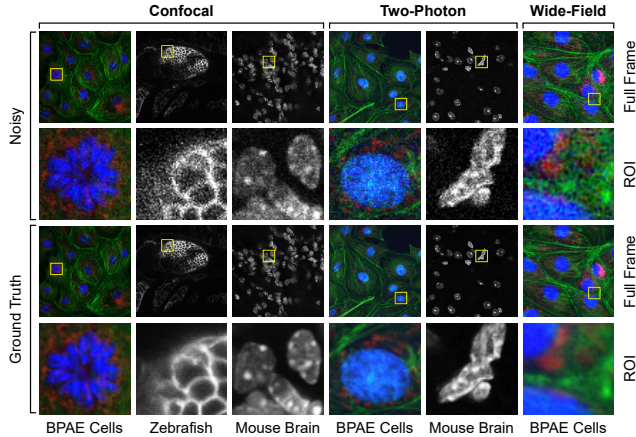


Figure 2. Examples of raw fluorescence microscopy images and their estimated ground truth from our FMD dataset. Shown here are FOVs from different microscopy modalities on different biological samples.

therefore, image averaging is equivalent to increasing the signal-to-noise ratio (SNR) of estimating the ground truth. We make $S = 1, 2, 4, 8, 16$ to create images with five different noise levels, and $S = 50$ to generate the ground truth. As demonstrated in [3] and also shown in Section 4.3, for fluorescence microscopy images, little image quality improvement can be seen after including around 40 images in averaging.

4. Dataset

In this Section, we describe the experimental setup that we used to acquire the fluorescence microscopy images. We then discuss how the raw images are utilized to estimate ground truth as well as images with different noise levels. Finally we present the statistics as well as the estimated noise levels of our dataset.

4.1. Image Acquisition Setup

Our FMD dataset covers the three main modalities of fluorescence microscopy: confocal, two-photon, and wide-field. All images were acquired with high-quality commercial fluorescence microscopes and imaged with real biological samples, including fixed bovine pulmonary artery endothelial (BPAE) cells [labeled with MitoTracker Red CMXRos (mitochondria), Alexa Fluor 488 phalloidin (F-actin), and DAPI (nuclei); Invitrogen FluoCells F36924], fixed mouse brain tissues (stained with DAPI and cleared), and fixed zebrafish embryos [EGFP labeled *Tg(sox10:megfp)* zebrafish at 2 days post fertilization]. All animal studies were approved by the university’s Institutional Animal Care and Use Committee.

To acquire noisy microscopy images for denoising purposes, we kept an excitation laser/lamp power as low as

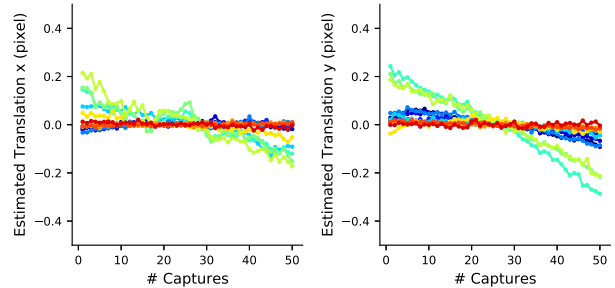


Figure 3. Estimated translation along x and y axes, both within a half-pixel (0.5). The estimation is performed on the 20-th FOV of each imaging configuration. Each line in a plot shows the estimation of one of the 12 configurations (different modalities on different samples).

possible for all imaging modalities. Specifically, the excitation power was low enough to generate a very noisy image, and yet high enough such that the image features were discernible. We also manually set the detector/camera gain to a proper value to avoid clipping and to fully utilize the dynamic range. Although pixel clipping could be inevitable because distinct biological structures with various optical properties could generate extremely bright fluorescence signals that could easily saturate the detector, we were able to maintain a very low number of clipped pixels (less than 0.2% of all pixels) in all imaging configurations. A table summarizing the percentages of clipped pixels to all pixels in the images is presented in the supplementary material. The details of the fluorescence microscopy setups, including a Nikon A1R-MP laser scanning confocal microscope and a Nikon Eclipse 90i wide-field fluorescence microscope, can also be found in the supplementary material.

For any imaging modality, each sample was imaged with 20 different FOVs, and each FOV was repeatedly captured for 50 times as 50 noise realizations. The acquired images were preprocessed and used for noisy image and ground truth estimation as described in Section 4.2. Figure 2 shows some example images of a single FOV from different imaging modalities and different samples.

4.2. Noisy Image and Ground Truth Estimation

Image registration The approach to estimate ground truth by averaging a sequence of captures usually comes with the issue of spatial misalignment, which is typical in photos taken by smartphones and DSLR. We use intensity-based image registration to register a sequence of image with the same FOV against the mean image of the sequence, but find that the estimated global translations in both x and y axis are less than a half-pixel (0.5), as shown in Figure 3. Translation in sub-pixel smooths out noisy images, and thus destroys the realness of Poisson noise which is the main characteristic of our dataset. In short, the image

sequence obtained by the commercial fluorescence microscopes is already well aligned; thus image registration is not performed.

Different noise levels As described in Section 4.1, the raw images are acquired with a low excitation power thus a relatively high noise level (low SNR) to increase the difficulty of denoising task. Meanwhile, the raw images with high noise levels allow us to create images with lower noise levels by image averaging. Particularly, we obtain averaged images with four extra noise levels by averaging S ($S = 2, 4, 8, 16$) raw images, respectively, within the same sequence (FOV) of 50. We sequentially select each image within the sequence; for each selected image, $S - 1$ images next to it are circularly selected; the S selected images in total are used for averaging. Using this circular averaging method, we are able to obtain the same number of averaged images as the number of raw images in the sequence, i.e., 50; meanwhile, the newly generated 50 raw images can be considered as 50 different noise realizations. In this way, the amount of noisy images in the dataset can be increased to five-fold ($S = 1, 2, 4, 8, 16$). Some example images with different noise levels are shown in Figure 1. As also shown in Table 2, the peak signal-to-noise ratio (PSNR) of the averaged images increases as the number of raw images used for averaging increases.

Ground truth estimation We estimate the ground truth by averaging all 50 captures on the same FOV, similar to the approaches employed in [3] and [17]; hence in the FMD dataset, each FOV has only one ground truth that is shared by all noise realizations from that FOV. As demonstrated in [3] and also shown in Section 4.3, the image quality or noise characteristics of a fluorescence microscopy image will see little improvement after including around 40 images in the average; therefore, we choose 50 captures as our criterion to obtain the ground truth. As shown in Equations (6) and (7), the ground truth y_i for images with different noise levels z_i^j is the same, and image averaging is equivalent to sampling from a Poisson-Gaussian distribution with a higher SNR. Regardless of the number of images used for averaging, the mean stays the same and equals to the ground truth. Figure 1 shows two ground truth images as well as their corresponding noise realizations.

4.3. Dataset Statistics and Noise Estimation

Taking the combination of each sample (the BPAE cells are considered as three samples due to its fluorophore composition) and each microscopy modality as a configuration, the FMD dataset includes 12 different imaging configurations that are representative of almost all fluorescence microscopy applications in practice. For each configuration, we capture 20 different FOVs of the sample, and for each

Modality	Samples	a	b
CF	BPAE (Nuclei)	1.39×10^{-2}	-2.16×10^{-4}
CF	BPAE (F-actin)	1.37×10^{-2}	-1.85×10^{-4}
CF	BPAE (Mito)	1.21×10^{-2}	-1.54×10^{-4}
CF	Zebrafish	9.43×10^{-2}	-1.60×10^{-3}
CF	Mouse Brain	1.94×10^{-2}	-2.68×10^{-4}
TP	BPAE (Nuclei)	3.31×10^{-2}	-8.39×10^{-4}
TP	BPAE (F-actin)	2.55×10^{-2}	-5.43×10^{-4}
TP	BPAE (Mito)	2.10×10^{-2}	-4.57×10^{-4}
TP	Mouse Brain	3.38×10^{-2}	-9.16×10^{-4}
WF	BPAE (Nuclei)	2.29×10^{-4}	2.35×10^{-4}
WF	BPAE (F-actin)	1.94×10^{-3}	1.91×10^{-4}
WF	BPAE (Mito)	3.55×10^{-4}	1.95×10^{-4}

Table 1. Estimation of noise parameters (a , b) of the FMD dataset. The shown a and b are average estimation values of 20 raw noisy images from 20 different FOVs (one raw image from each FOV). CF, confocal; TP: two-photon; WF: wide-field.

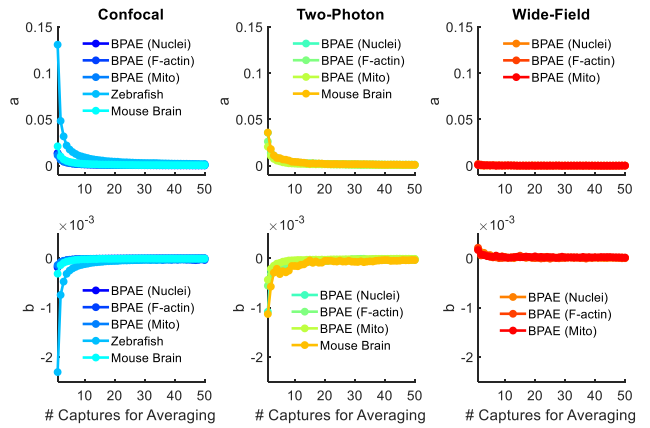


Figure 4. Estimated noise parameters (a and b) of averaged images obtained with different raw image numbers in the average. The estimation is performed on the second FOV of each imaging configuration.

FOV, we acquire 50 raw images. Meanwhile, the 50 raw images in a FOV can be extended to five-fold using the circular averaging method described in Section 4.2. Therefore, in total, the dataset has $12 \times 20 = 240$ FOVs or ground truth images, $240 \times 50 = 12,000$ raw images, and $12,000 \times 5 = 60,000$ noisy images as noise realizations.

While there are blind denoising methods (e.g., DnCNN) that are able to denoise an image without any additional information, most denoising algorithms such as NLM and BM3D, however, require an estimate of the noise levels presented in the image. In this work, we employ the noise estimation method in [10] to estimate the Poisson-Gaussian noise parameters, a and b , described in Section 3. The estimated values of a and b not only are needed in the benchmark of various denoising algorithms, they also reflect the characteristics of the noise presented in an images. Specif-

ically, since Poisson-Gaussian noise is a mixture of both Poisson and Gaussian noises, which are parameterized by a and b , respectively, an image with a large estimate value of a but a small b may be considered as a Poisson noise dominated image, while a small a with a large b can indicate that the image is Gaussian noise dominated. In fluorescence microscopy, however, it is unlikely to have a Gaussian noise dominated image due to the low signal levels; most fluorescence microscopy images are Poisson noise, or shot noise, dominated, with certain types of microscopes, such as wide-field ones, have a considerable amount of Gaussian noise involved [5, 20]. Note that the noise estimation program from [10] could generate a negative b value when the Gaussian noise component is small relative to the pedestal level (offset-from-zero of output). This, however, does not mean that the image has a “negative” Gaussian noise variance. More details can be found in [10]. In practice, when b is estimated to be negative, we make it zero in the subsequent PURE-LET and VST-based algorithms.

We evaluate the noise characteristics of our FMD dataset by estimating the noise parameters of raw noisy image (1 in each FOV, 240 in total). The estimated a and b are then grouped according to their corresponding imaging configurations (20 FOVs in each configuration, 12 configurations in total) and averaged. The results are presented in Table 1. For confocal and two-photon microscopy, the estimated a are comparably large while the b are negative; hence confocal and two-photon images are Poisson noise dominated. For wide-field microscopy, however, the a are much smaller than above, possibly due to the much lower sensitivity of CCD cameras used in wide-field microscopy compared to the PMTs used in confocal and two-photon microscopy; meanwhile, the b are now all positive, which indicates that wide-field images have a mixed Poisson-Gaussian noise with a considerable amount of Gaussian noise presented. We further evaluate the effect of image averaging on its noise characteristics. Figure 4 shows the estimated a and b values when different number of images, S , are included in the average. The results are in good agreement with the theory in Equation 7 and the observations in Table 1, as the estimated parameters follow the trend of a/S and b/S , and their initial values ($S = 1$) are close to the ones in Table 1. Figure 4 also shows that the values of a and b exhibit little change when the number of captures used for averaging is more than 40; this confirms the observation reported in [3] that the image quality or noise characteristics of a fluorescence microscopy image will see little improvement after including around 40 images in the average.

5. Benchmark

In this Section we benchmark several representative denoising methods, including deep learning models, on our fluorescence microscopy images with real Poisson-

Gaussian noise. We show that deep learning models perform better than traditional methods on the FMD dataset.

5.1. Setup

The FMD dataset is split to training and test sets, where the test set is composed of images randomly selected from the 19-th FOV of each imaging configuration and noise levels (the rest 19 FOVs are for training and validation purposes). The mixed test set consists of 4 images randomly selected from the 19-th FOV of 12 imaging configurations (combination of microscopy modalities and biological samples), organized in different noise levels. Thus we have 5 mixed test sets each of which have 48 noisy images with a specific noise level corresponding to 1 (raw), 2, 4, 8, and 16 times averaging. We also test the denoising algorithms on all 50 images from the same FOV (19-th) of a specific imaging configuration, also organized in different noise levels, with denoising results shown in the supplementary material.

Considering GPU memory constraint for training fully convolutional networks [30, 14] on large images, we crop the raw images of size 512×512 to four non-overlapping patches of size 256×256 . We evaluate the computation time on Intel Xeon CPU E5-2680, and additionally on Nvidia GeForce GTX 1080 Ti GPU for deep learning models.

The 10 benchmarked algorithms in this work can be divided into three categories. The first category is for the methods that are specifically designed for Poisson-Gaussian denoising; we benchmark PURE-LET [17], an effective and representative Poisson-Gaussian denoising algorithm. The second category is for using well-studied Gaussian denoising methods in combination with VST and inverse VST; we combine GAT and the exact unbiased inverse transformation with classical denoising algorithms including NLM [6], BM3D [8], KSVD and its two variants KSVD(D) (over-complete DCT dictionary) and KSVD(G) (global or given dictionary) [2], EPLL [33], and WNNM [11]. The last category is for deep learning based methods; we benchmark DnCNN [30] and Noise2Noise [14]. Note that the estimation of noise parameters a (scaling coefficient) and b (Gaussian noise variance) are required for the algorithms in the first and second categories to work. The estimation is performed according to Section 4.3 and then the images as well as the estimated parameters are sent to the denoising algorithms.

For benchmarking deep learning methods, unlike previous work [1] that directly tests with the pre-trained models, we re-train these models with the same network architecture and similar hyper-parameters on the FMD dataset from scratch. Specifically, we compare two representative models, one of which requires ground truth (DnCNN) and the other does not (Noise2Noise).

Methods	Number of raw images for averaging					Time
	1	2	4	8	16	
Raw	27.22 / 0.5442	30.08 / 0.6800	32.86 / 0.7981	36.03 / 0.8892	39.70 / 0.9487	-
VST+NLM [6]	31.25 / 0.7503	32.85 / 0.8116	34.92 / 0.8763	37.09 / 0.9208	40.04 / 0.9540	137.10 s
VST+BM3D [19]	32.71 / 0.7922	34.09 / 0.8430	36.05 / 0.8970	38.01 / 0.9336	40.61 / 0.9598	5.67 s
VST+KSVD [2]	32.02 / 0.7746	33.69 / 0.8327	35.84 / 0.8933	37.79 / 0.9314	40.36 / 0.9585	341.21 s
VST+KSVD(D) [2]	31.77 / 0.7712	33.45 / 0.8292	35.67 / 0.8908	37.69 / 0.9300	40.32 / 0.9579	67.96 s
VST+KSVD(G) [2]	31.98 / 0.7752	33.64 / 0.8327	35.83 / 0.8930	37.82 / 0.9312	40.44 / 0.9584	58.82 s
VST+EPLL [33]	32.61 / 0.7876	34.07 / 0.8414	36.08 / 0.8970	38.12 / 0.9349	40.83 / 0.9618	288.63 s
VST+WNNM [11]	32.52 / 0.7880	34.04 / 0.8419	36.04 / 0.8973	37.95 / 0.9334	40.45 / 0.9587	451.89 s
PURE-LET [17]	31.95 / 0.7664	33.49 / 0.8270	35.29 / 0.8814	37.25 / 0.9212	39.59 / 0.9450	2.61 s
DnCNN [30]	34.88 / 0.9063	36.02 / 0.9257	37.57 / 0.9460	39.28 / 0.9588	41.57 / 0.9721	3.07 s [†]
Noise2Noise [14]	35.40 / 0.9187	36.40 / 0.9230	37.59 / 0.9481	39.43 / 0.9601	41.45 / 0.9724	2.94 s [†]

Table 2. Denoising performance using the mixed test set, which includes confocal, two-photon, and wide-field microscopy images. PSNR (dB), SSIM, and denoising time (seconds) are obtained by averaging over 48 noise realizations in the mixed test set for each of 5 noise levels. Results of DnCNN and Noise2Noise are obtained by training on dataset with all noise levels. All 50 captures of each FOV (except the 19-th FOV which is reserved for test) are included in the training set, with 1 (DnCNN) or 2 (Noise2Noise) samples of which randomly selected from each FOV when forming mini-batches during training for 400 epochs. [†]Note that test time for deep learning models on GPU is faster in orders of magnitude, i.e. 0.62 ms for DnCNN and 0.99 ms for Noise2Noise on single GPU in our experiment.

5.2. Results and Discussion

The benchmark denoising results on the mixed test set is shown in Table 2, including PSNR, structural similarity index (SSIM) [27] and denoising time. From the table, BM3D (in combination with VST) is still the most versatile traditional denoising algorithm regarding its high PSNR and relatively fast denoising speed. PURE-LET, though its PSNR is not the highest, is the fastest denoising method among all the benchmarked algorithms thanks to its specific design for Poisson-Gaussian denoising. Finally, deep learning models outperform the other 8 methods by a significant margin in all noise levels, both in terms of PSNR and SSIM, even though they are *blind* to noise levels. This is different from the observation made before in [1, 24], probably because the nature of Poisson dominated noise is different from Gaussian noise while most of the denoising methods are developed for Gaussian noise model. Even if we applied the VST before Gaussian denoising, the transformed noise may still be different from a pure Gaussian one. More importantly, here the models are re-trained with our FMD dataset instead of pre-trained on other datasets.

The training data for deep learning models includes all imaging configurations and noise levels; thus we use one trained model to perform blind denoising on various imaging configurations and noise levels. We confirm that overall the Noise2Noise model has similar denoising performance as DnCNN, but without the need of clean images, and with almost 2dB higher than VST+BM3D in PSNR [14]. It even performs slightly better than DnCNN in the high noise domain, which is desirable in practice.

We investigate the effect of adding batch normalization layers for the Noise2Noise model (i.e. N2N-BN in Fig-

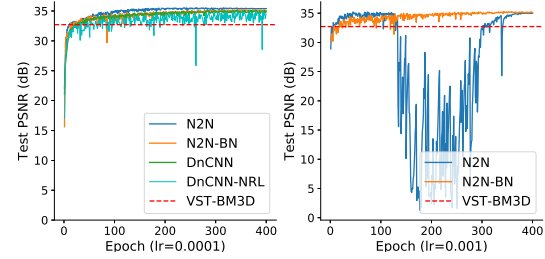


Figure 5. Test PSNR on the mixed test set with raw images during training. Each training epoch contains 18240 ($5 \times 12 \times 19 \times 16$) images of size 256×256 . Given enough training time (e.g. 400 epochs), Noise2Noise eventually outperforms DnCNN and VST-BM3D. Batch normalization helps stabilize training for Noise2Noise, and for DnCNN, residual learning does help improve denoising.

Table 3. PSNR (dB) on raw images in the mixed test set for the models trained with different learning rate.

Learn. Rate	1E-3	5E-4	1E-4	5E-5	1E-5
DnCNN	34.61	-	34.88	34.62	34.01
N2N	34.98	35.19	35.40	35.49	34.65
N2N-BN	35.15	35.07	35.12	35.12	34.60
DnCNN2	33.30	-	34.35	-	33.41

ure 5), which does help stabilize the training process even when the learning rate is relatively large (e.g. 0.001), but does not improve PSNR when the learning rate is well turned (e.g. 0.0001 which is the learning rate for benchmark). We also train DnCNN without residual learning (DnCNN-NRL) where the model directly outputs the denoised image instead of the residual between clean and noisy images, and confirm it is worse than the model with

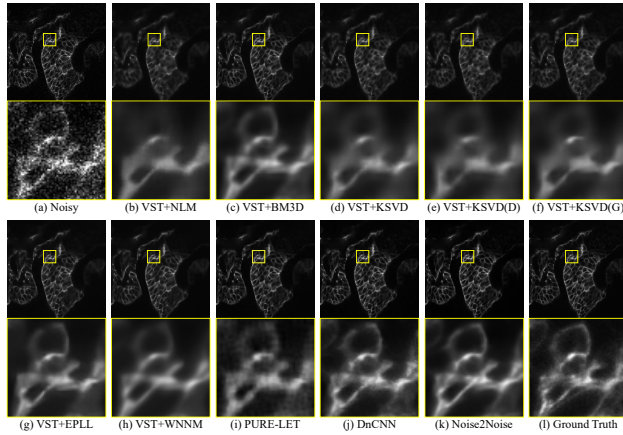


Figure 6. Benchmark results for raw single-channel (gray) images (zebrafish embryo under confocal microscopy). PSNR and SSIM values are in Table 4.

residual learning (DnCNN-RL), as has been reported in [30]. The test performance for the mixed test set with raw images during training is shown in Figure 5 and the PSNR for each case is shown in Table 3.

We also show benchmark results of the 10 algorithms on raw single-channel (gray) and raw multi-channel (color) confocal images in Figures 6 and 7, respectively, where the PSNR and SSIM of the color images are the mean values of that of their three channels.

The denoising time for deep learning models is the time to pass a mini-batch of four 256×256 patches cropped from one 512×512 image through the network. Deep learning models have similar denoising time with that of VST-BM3D and PURE-LET when running on CPU. However, the denoising time can be reduced to less than 1 ms when running on GPU, which potentially enables real-time denoising up to 100 frames per second, which is out of reach of traditional denoising methods. With such a denoising speed and high performance, deep learning denoising methods could dramatically benefit real-time fluorescence microscopy imaging, which allows biomedical researchers to observe the fast and dynamic biological processes in a much improved quality and to see processes that cannot be clearly seen before.

6. Conclusion

In this work, we have constructed a dedicated denoising dataset of real fluorescence microscopy images with Poisson-Gaussian noise, which covers most microscopy modalities. We have used image averaging to obtain ground truth and noisy images with 5 different noise levels. With this dataset, we have benchmarked representative denoising algorithms for Poisson-Gaussian noise including the most recent deep learning models. The benchmark results show

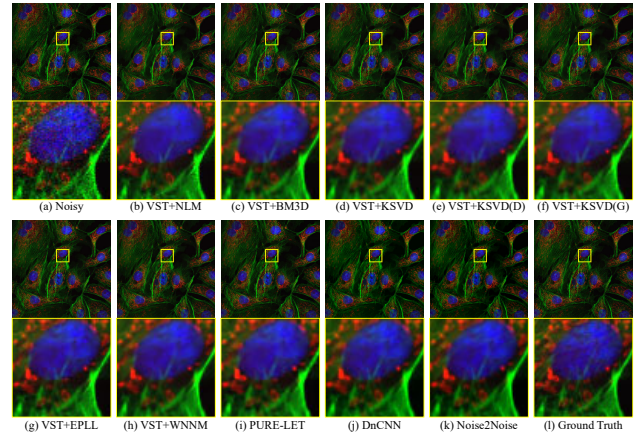


Figure 7. Benchmark results for raw multi-channel (color) images (BPAE cells under confocal microscopy). PSNR and SSIM values are in Table 4.

Table 4. Benchmark results [PSNR (dB) / SSIM] for confocal images of zebrafish embryo (Figure 6) and BPAE cells (Figure 7).

Methods	Zebrafish	BPAE
Raw	22.71 / 0.4441	30.67 / 0.7902
VST+NLM	28.49 / 0.7952	34.74 / 0.9108
VST+BM3D	31.99 / 0.8862	35.86 / 0.9338
VST+KSVD	29.25 / 0.8234	35.72 / 0.9209
VST+KSVD(D)	29.04 / 0.8212	35.47 / 0.9139
VST+KSVD(G)	29.23 / 0.8232	35.63 / 0.9176
VST+EPLL	31.71 / 0.8711	35.72 / 0.9335
VST+WNNM	31.22 / 0.8702	35.89 / 0.9322
PURE-LET	30.59 / 0.8332	35.18 / 0.9262
DnCNN	32.35 / 0.8991	36.15 / 0.9413
Noise2Noise	33.02 / 0.9109	36.35 / 0.9441

that deep learning denoising models trained on our FMD dataset outperforms other methods by a large margin across all imaging modalities and noise levels. We have made our FMD dataset publicly available as a benchmark for Poisson-Gaussian denoising research, which, we believe, will be especially useful for researchers that are interested in improving the imaging quality of fluorescence microscopy.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CBET-1554516. 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 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] A. Abdelhamed, S. Lin, and M. S. Brown. A high-quality denoising dataset for smartphone cameras. In *CVPR*, 2018.
- [2] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.
- [3] N. S. Alexander, G. Palczewska, P. Stremplewski, M. Wojtkowski, T. S. Kern, and K. Palczewski. Image registration and averaging of low laser power two-photon fluorescence images of mouse retina. *Biomedical Optics Express*, 7(7):2671, 2016.
- [4] J. Anaya and A. Barbu. RENOIR - A dataset for real low-light image noise reduction. *Journal of Visual Communication and Image Representation*, 51:144–154, 2018.
- [5] U. Bal. Dual tree complex wavelet transform based denoising of optical microscopy images. *Biomedical Optics Express*, 3(12):3231, 2012.
- [6] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *CVPR*, 2005.
- [7] J. Chen, J. Chen, H. Chao, and M. Yang. Image blind denoising with generative adversarial network based noise modeling. In *CVPR*, 2018.
- [8] 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.
- [9] W. Denk, J. H. Strickler, and W. W. Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248(4951):73–76, 1990.
- [10] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008.
- [11] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In *CVPR*, 2014.
- [12] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang. Toward convolutional blind denoising of real photographs. In *CVPR*, 2019.
- [13] S. Lefkimmiatis. Universal denoising networks: A novel cnn-based network architecture for image denoising. In *CVPR*, 2018.
- [14] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. Noise2noise: Learning image restoration without clean data. In *ICML*, 2018.
- [15] J. W. Lichtman and J.-A. Conchello. Fluorescence microscopy. *Nature Methods*, 2(12):910–919, 2005.
- [16] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.
- [17] F. Luisier, T. Blu, and M. Unser. Image denoising in mixed Poisson-Gaussian noise. *IEEE Transactions on Image Processing*, 20(3):696–708, 2011.
- [18] M. Makitalo and A. Foi. A closed-form approximation of the exact unbiased inverse of the Anscombe variance-stabilizing transformation. *IEEE Transactions on Image Processing*, 20(9):2697–2698, 2011.
- [19] M. Makitalo and A. Foi. Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise. *IEEE Transactions on Image Processing*, 22(1):91–103, 2013.
- [20] W. Meinzel, J.-C. Olivo-Marin, and E. D. Angelini. Denoising of microscopy images: A review of the state-of-the-art, and a New sparsity-based method. *IEEE Transactions on Image Processing*, 27(8):3842–3856, 2018.
- [21] P. A. Morris, R. S. Aspden, J. E. C. Bell, R. W. Boyd, and M. J. Padgett. Imaging with a small number of photons. *Nature Communications*, 6(1):5913, dec 2015.
- [22] S. Nam, Y. Hwang, Y. Matsushita, and S. Joo Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. In *CVPR*, 2016.
- [23] J. Pawley. *Handbook of Biological Confocal Microscopy*. Springer Science & Business Media, 2010.
- [24] T. Plötz and S. Roth. Benchmarking denoising algorithms with real photographs. In *CVPR*, 2017.
- [25] T. Plötz and S. Roth. Neural nearest neighbors networks. In *NIPS*, 2018.
- [26] P. J. Verveer, M. J. Gemkow, and T. 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.
- [27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [28] J. Xu, H. Li, Z. Liang, D. Zhang, and L. Zhang. Real-world noisy image denoising: A new benchmark. *arXiv preprint arXiv:1804.02603*, 2018.
- [29] B. Zhang, J. Fadili, and J. Starck. Wavelets, ridgelets, and curvelets for Poisson noise removal. *IEEE Transactions on Image Processing*, 17(7):1093–1108, 2008.
- [30] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.
- [31] K. Zhang, W. Zuo, and L. Zhang. Ffdnet: Toward a fast and flexible solution for CNN based image denoising. *IEEE Transactions on Image Processing*, 2018.
- [32] Y. Zhang, G. D. Vigil, L. Cao, A. A. Khan, D. Benirschke, T. Ahmed, P. Fay, and S. S. Howard. Saturation-compensated measurements for fluorescence lifetime imaging microscopy. *Optics Letters*, 42(1):155, 2017.
- [33] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *ICCV*, 2011.