This WACV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Unsupervised Denoising for Signal-Dependent and Row-Correlated Imaging Noise

Benjamin Salmon and Alexander Krull
School of Computer, University of Birmingham
brs209@student.bham.ac.uk, a.f.f.krull@bham.ac.uk
https://github.com/krulllab/COSDD

Abstract

Accurate analysis of microscopy images is hindered by the presence of noise. This noise is usually signal-dependent and often additionally correlated along rows or columns of pixels. Current self- and unsupervised denoisers can address signal-dependent noise, but none can reliably remove noise that is also row- or column-correlated. Here, we present the first fully unsupervised deep learning-based denoiser capable of handling imaging noise that is row-correlated as well as signal-dependent. Our approach uses a Variational Autoencoder (VAE) with a specially designed autoregressive decoder. This decoder is capable of modeling row-correlated and signal-dependent noise but is incapable of independently modeling underlying clean signal. The VAE therefore produces latent variables containing only clean signal information, and these are mapped back into image space using a proposed second decoder network. Our method does not require a pre-trained noise model and can be trained from scratch using unpaired noisy data. We benchmark our approach on microscopy datatsets from a range of imaging modalities and sensor types, each with row- or column-correlated, signal-dependent noise, and show that it outperforms existing self- and unsupervised denoisers.

1. Introduction

Microscopy has wide ranging applications in natural sciences. As imaging technology is routinely pushed to its limits, microscopy images become affected by the presence of noise. This poses a challenge in downstream analysis, making denoising an important preprocessing step. Over the years, various denoising approaches have been devised [30], from traditional filter-based methods [7] to supervised deep learning-based methods [56].

Supervised deep learning-based methods excel with respect to the quality of their output, however, they require paired training data, typically consisting of pairs of noisy and clean images. Unlike in applications with consumer images (photographs), such paired data is often unobtainable in microscopy [23, 50, 57]. Consequently, the applicability of supervised denoising methods in many areas of scientific imaging is limited.

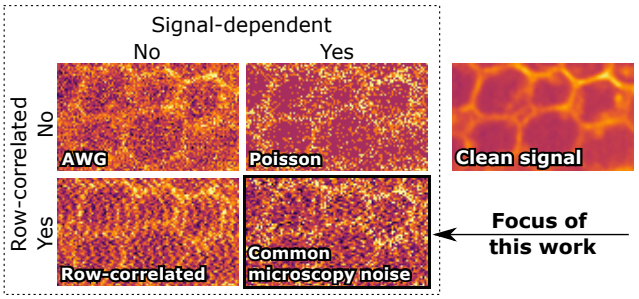


Figure 1. Noise is often assumed to be spatially uncorrelated and/or signal-independent such as additive white gaussian noise (AWG). Such assumptions greatly facilitate self/unsupervised denoising. Unfortunately, real noise in microscopy is usually neither. We present the first unsupervised denoising approach for this challenging scenario.

Self- and unsupervised methods (*e.g.* [2, 12, 27, 44, 45]) have been proposed as a solution to the lack of paired training data. They can be trained directly on the data that is to be denoised and have substantially improved the practicality of denoising in scientific imaging. These methods separate the imaging noise from the underlying signal by making assumptions about the statistical nature of the noise. Typically, they assume the noise is (i) signal-independent (purely additive and occurs separate from the underlying signal) [48], or (ii) spatially uncorrelated (unstructured and occurs separately for each pixel) [22, 27, 28, 45, 47]. Examples of (i) and (ii) are in Fig. 1.

In practice, (i) is often broken by the presence of Poisson shot noise [30]. Moreover, many popular scientific cameras and imaging setups break (ii) by producing row- or column-correlated noise (see Sec. 3.1 **Row-Correlated Noise**). These types of noise cannot be removed with basic

self- or unsupervised methods. Recently, variants of these methods for spatially correlated noise have been proposed, but they are limited to locally correlated noise and can come at the expense of reduced reconstruction quality [3, 44].

In this paper, we present the first unsupervised deep learning-based denoiser capable of reliably removing signal-dependent noise that is correlated along rows or columns of pixels, as it commonly occurs in microscopy data. The approach is illustrated in Fig. 2. Our method requires neither examples of noise-free images, which can be impossible to obtain (*e.g*. [53]), nor pre-trained noise models [28, 45, 48] or hand-crafted priors [7, 14, 53]. Furthermore, we do not rely on blind-spot approaches [3] or subsampling [22, 32] techniques that degrade image quality and limit denoising performance.

Our method is based on the representation learning technique proposed by Chen *et al*. [5], who revealed how a Variational Autoencoder (VAE) [25] prefers to avoid using latent variables to describe structures that could instead be modeled locally by its decoder. For an autoregressive (AR) decoder such as PixelCNN [15, 52], these are any inter-pixel dependencies that lie within its receptive field. Consequently, latent variables will only represent information that the decoder cannot model, *i.e*., dependencies that span beyond its receptive field.

We take advantage of this behavior by designing a decoder that is capable of modeling exactly what we want to be excluded from our latent variables – the imaging noise – while being incapable of modeling the remainder of the data. Specifically, we use an AR decoder that can only model axis-aligned structures because its receptive field spans only a row or column of the image. This trains our model to exclude row- or column-correlated noise from the latent variables, while encouraging it to include the statistics of the underlying signal. We then propose a second network, termed *signal decoder*, that is trained to map these latent variables back into image space, thus producing denoised images. Following Lehtinen *et al*. [33], who showed that noisy images can be used as training targets for denoisers, we train the signal decoder alongside the VAE using the original noisy data as targets.

Using real microscopy data recorded with various imaging modalities, we demonstrate that our denoiser achieves state-of-the-art results compared to other unsupervised methods. Furthermore, we show that the method is not sensitive to the length of the AR decoder's receptive field, as long as it is above a minimum, and that samples from the trained noise model have the same autocorrelation and signal-dependence as real noise.

2. Related Work

In the following, we will focus on methods that do be trained without paired training data.

Self-Supervised Denoising Most self-supervised denoisers use the blind-spot approach [2, 12, 27, 47, 55], where a network learns a denoising function when trained to predict the value of a pixel from surrounding pixels. Another technique trains the network to predict a subset of randomly sampled pixels from another [22]. For photon-counting data, Krull *et al*. [26] proposed removing photons and training a network to predict them with a Poisson distribution. These techniques all exploit a property of spatially uncorrelated noise, which is that the noise in one pixel cannot be predicted from the noise in other pixels.

To address spatially correlated noise, Structured Noise2Void (SN2V) [3] extended the blind-spot approach to also mask pixels containing noise that is correlated with the noise in the pixel being predicted. Lee *et al*. [32] also extended the blind-spot approach by subsampling pixels to break up noise structures, making noise effectively spatially uncorrelated and ready for traditional blind-spot denoising. N2V2 [21] constrained a blind-spot network's ability to reconstruct high-frequency content.

There are also methods that add two independent noise samples to already-noisy images, training networks to denoise by mapping between these doubly-noisy pairs [39, 41]. While theoretically applicable to signal dependent, spatially correlated noise, practical implementation has not been demonstrated. The obstacle lies in obtaining a suitable noise model to sample from, particularly when signal dependent and spatially correlated noise must be described by a covariance matrix ([41]).

Diffusion-based Denoising In recent years, diffusion models have been proposed as priors for solving inverse problems, such as denoising, in a Bayesian way [6, 11, 51]. However, such methods require clean training data and a known mathematical forward model of the corruption process during inference, both of which are unavailable in many microscopy applications. Very recently, methods have been proposed to remove at least the requirement for clean training data by assuming simple additive Gaussian noise [1] or focusing on inpainting problems [8].

GAN-based Denoising Another approach of denoising trains a conditional generative adversarial network [37] to model the distribution of clean images given noisy images. This can be done using a dataset of unpaired noisy and clean images, provided that the generator is forced to output an estimate of the clean signal underlying its noisy input, and not a random clean image. Various methods have been proposed to ensure this, including minimizing the perceptual difference between the generator's input and output [58], or maximizing their mutual information [54]. While these methods have more relaxed requirements than supervised denoisers, obtaining a dataset of clean images can be a challenge in microscopy, even if it is unpaired. This is because the clean images must follow the same distribution as the

signal content in the noisy training data. That means, *e.g.*, the same cell type, imaging conditions, *etc*.

VAE-based Denoising These methods train a latent variable model of noisy data using a VAE and a pre-trained explicit noise model. Through the use of the noise model, the VAE is trained to represent clean signals with its latent variables. In DivNoising [45], the noise model assumes the noise is spatially uncorrelated, *i.e.*, noise is generated in each pixel independently. Later, HDN [44] was proposed as an extension to DivNoising that used a VAE with a hierarchy of latent variables [49]. It was found that short-range spatially correlated noise structures were modeled by only the bottom levels of the hierarchy and could be removed by preventing these latent variables from using information from the encoded input. This technique is known as HDN_{36}.

Next, Autonoise [48] was proposed to remove spatially correlated but signal-independent noise by replacing the pixel-independent noise model in HDN with a CNN-based AR noise model [52]. This noise model was pre-trained using samples of pure noise which can be obtained by, *e.g.*, imaging without light. The method was therefore able to remove noise of any spatial correlation, but could not be applied to signal-dependent noise.

3. Background

Before we come to describe our method, we will first formalise the type of noise that we aim to remove, then give some background on VAEs for representation learning.

3.1. Image Formation and Imaging Noise

We can express image formation as a two step process. The clean image, or as we also refer to it, the *signal*, $\mathbf{s} = (s_{1,1}, \ldots, s_{N,M})$, is drawn from a distribution $p(\mathbf{s})$ and then subjected to noise, producing the noisy image $\mathbf{x} = (x_{1,1}, \ldots, x_{N,M})$ as drawn from the noise distribution $p(\mathbf{x}|\mathbf{s})$. Here, $s_{i,j}$ and $x_{i,j}$ correspond to the respective pixel values at position (i, j), and N and M are the number of rows and columns in the image, respectively.

Generally, we assume that the imaging noise is zero-centered, that is, the expected value of a noisy image equals the signal; $\mathbb{E}_{p(\mathbf{x}|\mathbf{s})}[\mathbf{x}] = \mathbf{s}$. Note that zero-centered noise is a weak assumption that is widely shared in literature (*e.g.* [12, 27, 31]). Another way of looking at this assumption is that the sensor or camera is assumed to behave linearly so that an increase in $s_{i,j}$ (*e.g.* by increasing the light intensity) will, on average, result in a proportional increase in $x_{i,j}$.

Additive White Gaussian Noise The most basic traditionally used noise model is additive white Gaussian (AWG) noise, see *e.g.* [4, 7, 36]. We can write the probability distribution for noisy observations in an AWG noise model as

$$p(\mathbf{x}|\mathbf{s}) = \prod_{i=1}^{N}\prod_{j=1}^{M} p(x_{i,j}|s_{i,j}), \tag{1}$$

formulated as a product over pixels. In general, we refer to any type of noise model that can be factorised according to Eq. (1) as spatially uncorrelated. In AWG noise, $p(x_{i,j}|s_{i,j})$ corresponds to a normal distribution centered at $s_{i,j}$ with variance fixed for all $s_{i,j}$. In general, we call noise with a variance that does not depend on the underlying signal as signal-independent. Noise for which this is not the case will be referred to as signal-dependent.

Denoising approaches have assumed noise to be both spatially uncorrelated and signal-independent. Unfortunately, real imaging noise rarely follows these assumptions.

Poisson Shot Noise Most imaging systems are affected by Poisson shot noise [30]. A Poisson noise model follows Eq. (1), with $p(x_{i,j}|s_{i,j})$ corresponding to a Poisson distribution with rate $\lambda = s_{i,j}$. A popular variant on this model is the Poisson-Gaussian noise model (*e.g.* [59]), where $p(x_{i,j}|s_{i,j})$ is assumed to be a combination of Poisson shot noise and Gaussian read-out noise.

Poisson and Poisson-Gaussian noise models are spatially uncorrelated but signal-dependent, since the variance of the Poisson component in each pixel depends on the pixel's signal $s_{i,j}$.

Row-Correlated Noise While Poisson-Gaussian noise models are a popular choice in describing noise distributions, many imaging systems can produce noise that does not conform to Eq. (1), but is instead correlated along rows or columns of pixels. For example, Scanning Transmission Electron Microscopy (STEM) [42] is prone to line artifacts caused by the slow reaction time of read-out electronics [34], and Laser Scanning Confocal Microscopy (LSCM) [40] has been shown to produce artifacts in the scanning direction [20]. Depending on their settings, Electron Multiplying Charge-Coupled Device (EM-CCD) [9] cameras can produce horizontally correlated read-out noise [3]. The scientific Complementary Metal-Oxide-Semiconductor (sCMOS) [38] cameras that are popular in optical microscopy [35] have separate amplifiers for each column of pixels, leading to correlated noise within each column [60]. In addition to microscopy, the detectors used in infrared imaging systems, *e.g.* microbolometers, also commonly use separate column amplifiers and suffer from similar noise structures [10]. Examples of noisy images from these modalities along with plots of their spatial autocorrelation and signal-dependence can be found in Fig. 3.

We propose to describe such noise as

$$p(\mathbf{x}|\mathbf{s}) = \prod_{i=1}^{N}\prod_{j=1}^{M} p(x_{i,j}|\mathbf{s}, x_{i,1}, \ldots, x_{i,j-1}), \tag{2}$$

where $p(x_{i,j}|\mathbf{s}, x_{i,1}, \ldots, x_{i,j-1})$ is the distribution of pos-

sible noisy pixel values $x_{i,j}$ conditioned on the signal, as well as all "previous" values in the same row, i. While this model can describe interactions between pixels within the same row, pixels in different rows are (conditionally) independent (given a signal $\mathbf{s}$). Note that in this formulation, a pixel $x_{i,j}$ can depend on not only the signal in pixel (i, j), as with shot noise, but on the entire image for more complex interactions. In this work, we address the type of noise described in Eq. (2), which we believe is a good model for many real scientific imaging data.

3.2. Latent Variable Models and VAEs

A latent variable model, with parameters θ, defines a probability distribution, $p_\theta(\mathbf{x}) \propto p_\theta(\mathbf{x}|\mathbf{z})p_\theta(\mathbf{z})$, over observed variables $\mathbf{x}$ via latent variables $\mathbf{z}$. This can be used to represent a data generation process in which a value $\mathbf{z}$ is first sampled from the prior $p_\theta(\mathbf{z})$ and a value $\mathbf{x}$ is sampled from the conditional distribution $p_\theta(\mathbf{x}|\mathbf{z})$.

A VAE can be used to simultaneously optimize the model's parameters and approximate the posterior, $p_\theta(\mathbf{z}|\mathbf{x})$, via maximization of a lower bound on the marginal log-likelihood,

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z}))] - KL(q_\phi(\mathbf{z}|\mathbf{x}))\|p_\theta(\mathbf{z})), \tag{3}$$

where the second term on the RHS is the Kullback-Leibler divergence [29] from the prior to an approximate posterior $q_\phi(\mathbf{z}|\mathbf{x})$. In Eq. (3), the first term is known as the reconstruction error, the second term as the regularization error, $q_\phi(\mathbf{z}|\mathbf{x})$ as the encoder, and $p_\theta(\mathbf{x}|\mathbf{z})$ as the decoder.

3.3. VAEs and the Division of Labor

When designing a latent variable model for image data, the decoder can be made autoregressive, where the distribution of each pixel is conditioned on the value of previous pixels in row-major order, as well as on $\mathbf{z}$,

$$p_\theta(\mathbf{x}|\mathbf{z}) = \prod_{i=1}^{N}\prod_{j=1}^{M} p_\theta(x_{i,j}|\mathbf{z}, x_{<(i,j)}). \tag{4}$$

We refer to $x_{<(i,j)}$ as the full AR *receptive field*, and its shape is shown in Fig. 2b. This modification is intended to make the model more expressive. However, the decoder is now powerful enough to model the entire data distribution locally, to the extent that it is unclear which aspects of the data should be encoded in $\mathbf{z}$ and which should be modeled by the decoder.

When using a VAE in practice, there seems to be a preference to model as much content as possible with the decoder. He *et al*. [18] argued that this behavior is caused by the approximate posterior lagging behind the true posterior in the early stages of training, causing the parameters to get stuck in a local optimum where the decoder models the data without using the latent variables.

Alternatively, Chen *et al*. [5] reasoned that with most practical VAEs, ignoring the latent variables leads a tighter lower bound on the marginal log-likelihood. By only using the latent variables to express what the decoder cannot model, the true posterior is brought closer to the prior and can be more closely matched by the relatively inflexible approximate posterior. The authors then demonstrated how this behavior can be used to control the division of labor in a VAE. Specifically, they designed an AR decoder that is capable of modeling information that they do not want captured in the latent variables, but is incapable of modeling the information that they do want captured in the latent variables.

In this paper, we design an AR receptive field that only captures the correlations common in scientific imaging noise. This allows us to train a latent variable model of noisy image data where latent variables explain the signal content and the decoder models only the noise generation process. We then use the approximate posterior to sample latent variables, each representing one of the clean signals that could possibly underlie a given noisy image.

4. Method

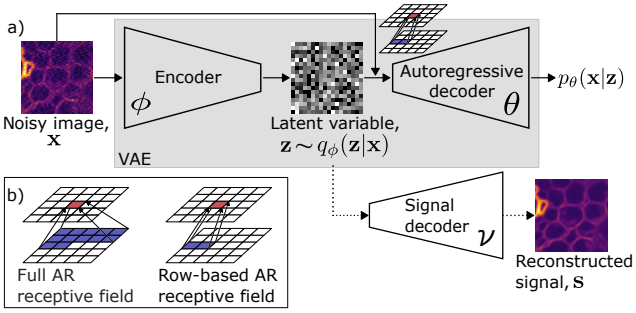


Figure 2. **a):** A Variational Autoencoder [5] (solid arrows) is trained to model the distribution of noisy images $\mathbf{x}$. The autoregressive (AR) decoder models the noise component of the images, while the latent variable models only the clean signal component $\mathbf{s}$. In a second step (dashed arrows), our novel *signal decoder* is trained to map latent variables into image space, producing an estimate of the signal underlying $\mathbf{x}$. **b):** To ensure that the decoder models only the imaging noise and the latent variables capture only the signal, we modify the AR decoder's receptive field. In a full AR receptive field (Eq. (4)), each output pixel (red) is a function of all input pixels located above and to the left (blue). In our decoder's row-based AR receptive field (Eq. (5)), each output pixel is a function of input pixels located in the same row, which corresponds to the row-correlated structure of imaging noise.

We propose a VAE-based unsupervised image denoiser for noise that is both signal-dependent and correlated along rows or columns of pixels. It is trained using only noisy images and does not require a pre-trained noise model. Moreover, it does not require pixel blind-spots or subsampling.

Instead, we restrict the receptive field of the AR decoder in a hierarchical VAE [49] to a row or column of pixels, allowing the AR decoder to model the correlations of noise content but not the correlations of underlying signal. Following the insights of Chen *et al*. [5] (Sec. 3.3), the AR decoder will therefore learn to model noise, leaving the underlying signal to be encoded in the latent variable $\mathbf{z}$. We then propose a method for taking these latent variables and mapping them back into image space, obtaining estimates of clean signals. An outline of our method can be found in Fig. 2.

Autoregressive Receptive Fields for Noise Our AR decoder has a one-dimensional receptive field that is sufficient for modeling row- or column-correlated noise (Eq. (2)) as it occurs in microscopy (see Sec. 3.1 **Row-Correlated Noise**) by spanning pixels in the same row or column as $x_{i,j}$. See Fig. 2b for a visual representation.

To remove row-correlated noise, the first step in our denoising process is to train a VAE to model noisy image data with the objective function in Eq. (3), where $p_\theta(\mathbf{x}|\mathbf{z})$ is factorized as,

$$p_\theta(\mathbf{x}|\mathbf{z}) = \prod_{i=1}^{N}\prod_{j=1}^{M} p_\theta(x_{i,j}|\mathbf{z}, x_{i,1}, \ldots, x_{i,j-1}). \quad (5)$$

To remove column-correlated noise, the factorization over pixels is perpendicular.

We find that this factorization is insufficient for modeling signal content, which is highly correlated in all directions. Consequently, the VAE learns to encode signal content in its latent variables. An experimental investigation of the effects of changing receptive field size can be found in Sec. 5.2, and details on how we construct this receptive field with our AR decoder architecture can be found in the supplementary material.

Decoding the Signal Once our VAE has been trained, its latent space will represent clean signals, *i.e*., each $\mathbf{z}$ contains all the information about an $\mathbf{s}$. We would now like to use the model for denoising by inferring possible clean signals $\mathbf{s}$ for a given noisy image $\mathbf{x}$. Unfortunately, unlike previous methods [44, 45, 48], we cannot directly sample clean images from our encoder. Rather, samples from $q_\phi(\mathbf{z}|\mathbf{x})$ will directly *correspond* to a clean signal $\mathbf{s}$. We denote the signal corresponding to a value of $\mathbf{z}$ as $\mathbf{s}(\mathbf{z})$. For an experimental validation of this deterministic relationship, please refer to the supplementary material. To obtain denoised images, we approximate $\mathbf{s}(\mathbf{z})$ with an additional regression network, termed *signal decoder*, $f_\nu(\mathbf{z}) \approx \mathbf{s}(\mathbf{z})$. In the following, we describe how $f_\nu(\mathbf{z})$ is trained despite not having access to training pairs $(\mathbf{z}, \mathbf{s}(\mathbf{z}))$.

In fact, training the signal decoder only requires pairs $(\mathbf{z}_k, \mathbf{x}_k)$, where $\mathbf{x}_k$ is a noisy image and $\mathbf{z}_k \sim q_\phi(\mathbf{z}|\mathbf{x}_k)$. If we assume that the approximate posterior is accurate, such that $q_\phi(\mathbf{z}|\mathbf{x}) \approx p_\theta(\mathbf{z}|\mathbf{x})$, these pairs can be equivalently thought of as samples from the joint distribution, $(\mathbf{z}_k, \mathbf{x}_k) \sim p_\theta(\mathbf{z}, \mathbf{x})$.

Least squares regression analysis tells us that optimizing the signal decoder by minimizing the squared $L2$ norm,

$$\mathcal{L}(\nu) = \|f_\nu(\mathbf{z}_k) - \mathbf{x}_k\|_2^2, \quad (6)$$

will train it to approximate $\mathbb{E}_{p_\theta(\mathbf{x}|\mathbf{z})}[\mathbf{x}]$ for any $\mathbf{z}$ [17]. By definition, $\mathbf{z}$ contains no more or less information about $\mathbf{x}$ than $\mathbf{s}(\mathbf{z})$, so the signal decoder will equivalently be approximating $\mathbb{E}_{p_\theta(\mathbf{x}|\mathbf{s}(\mathbf{z}))}[\mathbf{x}]$. Recalling that imaging noise is zero-centered (Sec. 3.1), the expected value of a noisy image given an underlying signal is *that signal*. Therefore, $f_\nu(\mathbf{z}) \approx \mathbb{E}_{p_\theta(\mathbf{x}|\mathbf{s}(\mathbf{z}))}[\mathbf{x}] = \mathbf{s}(\mathbf{z})$. This is similar to Noise2Noise [33], where the regressor for $\mathbf{s}$ is trained using noisy targets $\mathbf{x}$.

Even though the signal decoder would naturally be trained in a second stage after the main VAE is finished, in practice we co-trained it alongside the main VAE. At every training step, the sampled latent variable is fed to both decoders, but only the loss from the AR decoder is allowed to backpropagate to the encoder. This method of training is simply for convenience and we did not observe any changes in performance compared to a signal decoder that is trained separately, after the main VAE.

Inference With the VAE and signal decoder trained, we can denoise an image $\mathbf{x}$ in a two-step process. We first sample a latent variable $\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})$, then obtain a clean image by decoding $\hat{\mathbf{s}} = f_\nu(\mathbf{z})$. Similarly to [44, 45, 48], the result constitutes a random possible solution $\hat{\mathbf{s}} \sim p(\mathbf{s}|\mathbf{x})$. To obtain a consensus solution, we follow [44, 45, 48] in averaging a large number of such samples. In the following experiments, all the results are the mean of 100 samples, both for our method and the baseline HDN_{36}. Please refer to the supplementary material for the inference time and PSNR of the mean of 1, 10 and 1000 samples.

5. Experiments

5.1. Benchmarking Denoising Performance

Datasets We tested the performance of our proposed denoiser on real noisy images captured by five different imaging modalities that commonly suffer from row-correlated noise. Two modalities have noisy image datasets with known ground truth for quantitative evaluation. The method for obtaining ground truth can be found in the source publications. The first is the EMCCD sensor, for which we have three spinning disk confocal fluorescence microscopy datasets: *Convallaria A* [3], *Convallaria B* [46] and *Mouse Actin* [46]. The second is LSCM, for which we have three fluorescence datasets: *Mouse Nuclei* [46], *Actin Confocal* [16] and *Mito Confocal* [16]. Note that *Mouse Nuclei* contains spatially uncorrelated noise, but was included to demonstrate that the proposed method is still applicable

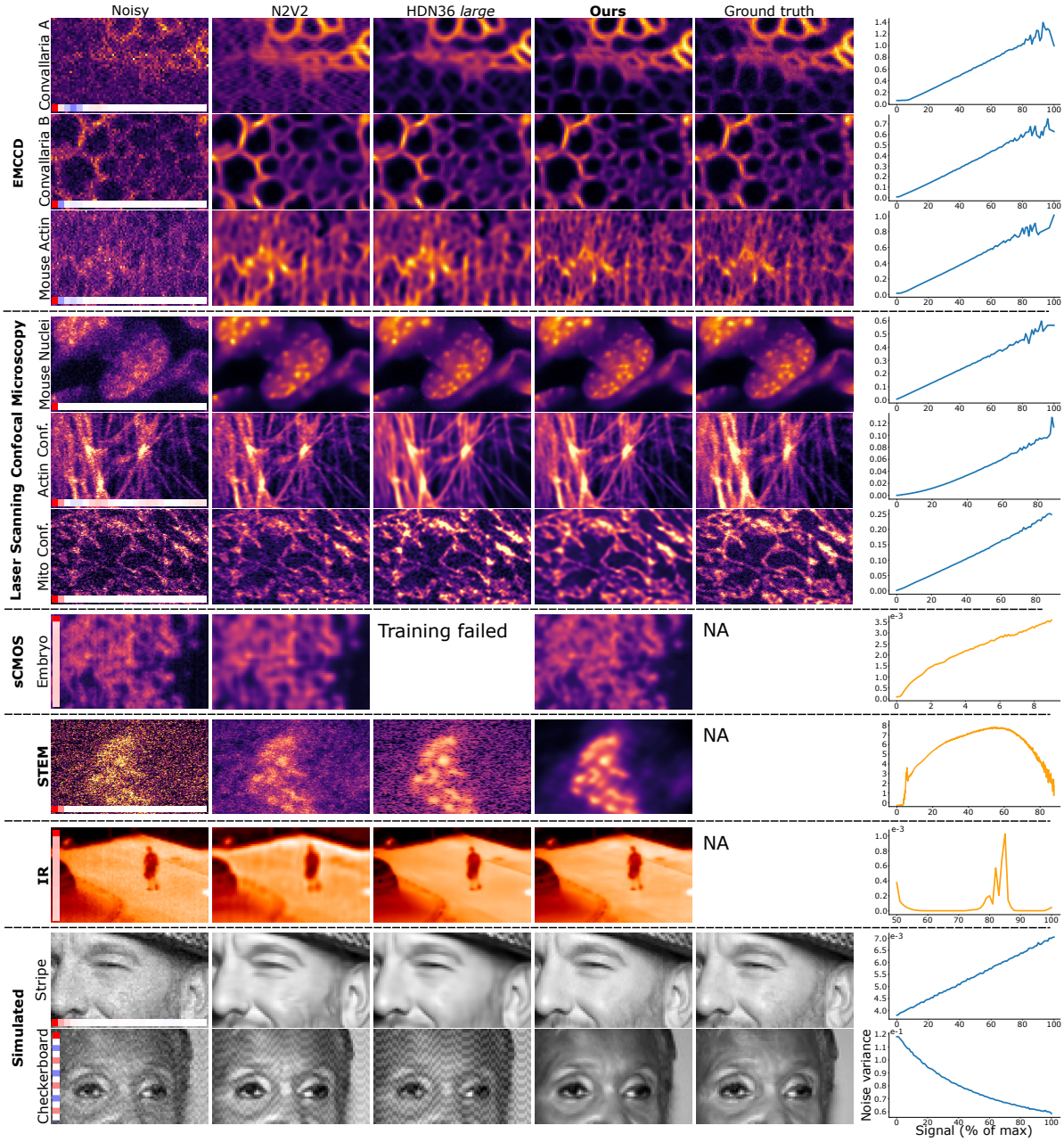


Figure 3. Visual results from our method and two unpaired baselines on all datasets. The spatial autocorrelation of the noise is overlaid on each noisy image, with red indicating a positive correlation and blue indicating a negative correlation. The direction of the correlation is given by the orientation of the autocorrelation bar. Additionally, the signal dependence of the noise in each dataset is shown in the right-hand column. The horizontal axis of these graphs is the clean signal intensity as a percentage of the maximum, while the vertical axis is the variance of noisy pixel values recorded for these signal intensities. Ground truth must be used to calculate signal dependence, so orange lines are used where denoised images from our method are used as pseudo-ground truth.

to spatially uncorrelated noise without modification. For details on the size of the datasets and the train/test splits, see original publications. The remaining modalities have only noisy images. These are the sCMOS sensor for which we have one fluorescence dataset: *Embryo* [13], the microbolometer with one infrared imaging dataset: *IR* [43], and a STEM dataset: *STEM* [19].

In addition to real data, we created two datasets by corrupting the Flickr Faces HQ thumbnails dataset [24] with simulated noise. For *FFHQ - Stripes*, the images were cor-

Table 1. Comparison to baseline methods using Peak Signal-to-Noise Ratio (PSNR), higher is better. The best results for methods that do not require paired images are printed in **bold**, and best results overall are underlined. HDN is a variant of HDN_{36} that is intended for spatially uncorrelated noise, but was applied to two datasets with spatially correlated noise by the original authors. These extra results are presented in brackets. Note that the *Mouse Nuclei* dataset contains spatially uncorrelated noise and that HDN_{36} failed to train on the *Mouse Actin* dataset. CARE and N2N are supervised denoisers requiring paired images, unlike the other baselines.

	EMCCD			LSCM			Simulated	
	Conv. A [3]	Conv. B [46]	Mouse Actin [46]	Mouse Nuclei* [46]	Actin Conf. [16]	Mito Conf. [16]	FFHQ Stripe [24]	FFHQ Checkerb. [24]
AP-BSN [32]	22.30	24.10	29.55	35.41	26.89	24.94	28.22	13.19
SN2V [3]	30.29	31.67	32.80	36.62	23.30	26.41	29.52	19.11
N2V2 [21]	29.36	36.72	34.23	35.88	26.71	25.89	33.08	29.22
HDN_{36} (HDN) [44]	31.41	37.21 (37.39)	- (34.12)	(36.87)	26.84	26.51	32.64	29.61
HDN_{36} *large* (HDN *large*) [44]	31.80	37.92	34.14	(38.12)	27.17	26.15	34.54	25.51
Ours *small*	34.42	38.99	36.50	39.56	27.35	27.49	32.87	33.51
Ours *large*	**37.49**	**44.10**	**39.23**	**42.98**	**27.41**	**27.50**	**35.66**	**36.27**
CARE [56]	31.56	36.71	34.20	36.58	29.44	27.55	36.46	36.89
N2N [33]	28.22	36.85	34.60	37.33	-	-	36.18	36.43

rupted by a combination of additive white Gaussian noise, Poisson shot noise, and additive white Gaussian noise that had undergone a horizontal Gaussian blur. For *FFHQ - Checkerboard*, the images were corrupted by a combination of a vertical checkerboard pattern and Gaussian noise with an inverse signal dependence. Details on the simulated noises are in the supplementary materials.

Baselines and Architecture We compare our method with other deep learning-based denoisers that can be applied to spatially correlated, signal-dependent noise and do not require paired images. These are the self-supervised denoisers Structured Noise2Void (SN2V) [3], Noise2Void2 (N2V2) [21] and Asymmetric PD BSN (AP-BSN) [32], and the unsupervised denoiser Hierarchical DivNoising36 (HDN_{36}) [44]. We use these as baseline methods along with Content-Aware Image Restoration (CARE) [56], which is trained using pairs of noisy and clean images, and Noise2Noise (N2N) [33], which is trained using two noisy acquisitions of the same signal, on datasets with suitable training images available. For details on how each baseline was implemented, refer to the supplementary material.

Of the denoisers not requiring paired images, the performance of HDN_{36} is the best, but the model implemented in the baseline's publication uses significantly fewer parameters than ours (7 million to 25 million). We therefore evaluate an additional version, termed HDN_{36} *large*, with a similar number of parameters made by increasing the number of latent dimensions from 32 to 64.

It should be noted that the *Convallaria B* and *Mouse Actin* datasets had been treated as spatially uncorrelated by Prakash *et al*. [44] when HDN was tested. We also report those results in Tab. 1. It should also be noted that the noise in the *Mouse Nuclei* dataset is spatially uncorrelated, so was denoised by HDN and its higher parameter version, HDN *large*, which was made the same way as HDN_{36} *large*.

Our method requires a choice of orientation and size for the AR decoder's receptive field. Orientation was determined by examining the spatial autocorrelation of noise samples, with these plots reported in Fig. 3, and following the ablation study in Sec. 5.2, we always used a receptive field length of 40 pixels.

As for our model's encoder, latent variables are produced by a hierarchical VAE [49] with 14 levels to its hierarchy. In addition to the full-sized version, we evaluate a smaller version that requires approximately 6GB (as opposed to 20GB) of GPU memory to train. This was achieved by reducing the number of latent variables from 14 to 6 and reducing the number of latent dimensions from 64 to 32. We refer to these models as Ours *small* and Ours *large*, for the lower memory and higher memory versions respectively. See the supplementary materials for full architecture and training details.

Discussion Quantitative results measuring the Peak Signal-to-Noise Ratio (PSNR) in dB are reported in Tab. 1, and qualitative results are reported in Fig. 3. Note that HDN_{36} failed to train with the *Mouse Actin* and the *Embryo* dataset. Out of the methods that do not require paired images, Ours *large* achieved the highest PSNR across all datasets, even beating the supervised CARE and N2N on four of six microscopy datasets. Ours *small* then had the second highest PSNR for an unpaired method on all datasets except *FFHQ - Stripe* .

Turning to the qualitative results in Fig. 3, we see that Ours *large* denoised images from each dataset without leaving behind any artifacts, whereas HDN_{36} *large* could not remove the correlated component of the noise from the *STEM* or the *FFHQ - Checkerboard* dataset. SN2V left artifacts in *Convallaria A* and *FFHQ - Checkerboard*. It is unclear if the high frequency features that are visible in the *Mito Confocal* output of HDN_{36} *large* and N2V2, but not in our output, are true signal or artifacts of remaining noise, as the ground truth also contains a low level of noise.

5.2. Ablation Study - Receptive Field Size

As stated in Sec. 4, to model the noise correlations addressed in this paper, the receptive field of a VAE's AR decoder must span pixels in the same row or column as the pixel being predicted. In Fig. 4, the effect of receptive

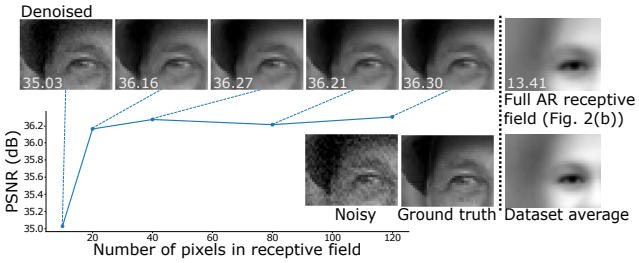


Figure 4. The *FFHQ - Checkerboard* dataset wass denoised 5 times by varying the number of pixels covered by the AR decoder's receptive field. The images show denoising results for each receptive field size with PSNR overlaid. Additionally, an image denoised using a full AR receptive field is included on the right. In this situation, the signal decoder is given completely uninformative inputs and learns to output the mean of the training dataset.

field size (number of pixels) is investigated by denoising the *FFHQ - Checkerboard* dataset using a range of receptive field lengths, from 10 to 120 pixels, and measuring the effect on PSNR. We also include the effect of training a model with a full AR receptive field, as shown in Fig. 2b.

This study shows that the AR decoder is able to model this spatially correlated noise, and therefore have it removed by the encoder, when the receptive field spans 40 pixels. Moreover, the AR decoder does not seem to model more of the signal as the receptive field grows. If it did, we would expect a steady drop in PSNR as denoised images lose signal content. However, as the image on the right of Fig. 4 shows, the signal will be modeled by an AR decoder with a full receptive field. Latent variables in this situation carry no information about $\mathbf{x}$, causing the signal decoder to minimize Eq. (6) by predicting the mean of the training set. This is expected, as when the latent variable $\mathbf{z}$ has no information about $\mathbf{x}$, $\mathbb{E}_{p_\theta(\mathbf{x}|\mathbf{z})}[\mathbf{x}] = \mathbb{E}_{p_\theta(\mathbf{x})}[\mathbf{x}]$.

5.3. Ablation Study - Noise Reconstruction

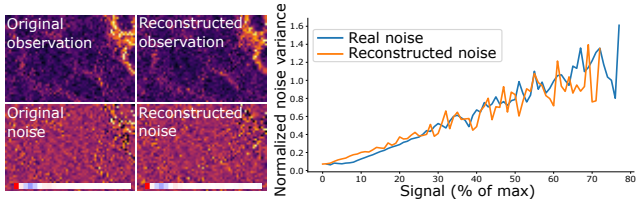


Figure 5. A noisy image from the *Convallaria A* dataset was encoded and decoded to produce a reconstructed observation and an artificial noise sample. The sampled noise has spatial autocorrelation and signal dependence that match those of the real noise, indicating that the AR decoder has learnt an accurate noise model.

If the VAE's AR decoder is modeling only the noise component of images, encoding an image with the VAE and sampling from the AR decoder should yield an image with the same underlying signal but a different random sample of noise. If the decoder's model of the noise is accurate, the sampled noise should exhibit the same autocorrelation and signal-dependence characteristics as the noise in the original image. An investigation into this is reported in Fig. 5, where a noisy image from the *Convallaria A* dataset was encoded by a trained VAE, then a latent variable was sampled and used by the VAE's AR decoder to sample a reconstruction of the original image. The reconstructed noise exhibits spatial autocorrelation and signal dependence very similar to the real noise, indicating that our AR decoder has learnt an accurate model of the noise.

6. Conclusion

We proposed an unsupervised VAE-based denoising algorithm for signal-dependent noise that is correlated along rows or columns of pixels. It is trained without any clean data or a noise model. By engineering the receptive field of the AR decoder, the VAE's latent variables are encouraged to represent the signal content of an image while discarding the noise. We then presented a novel signal decoder that is trained to map this latent variable to an estimate of the clean image. The algorithm outperforms existing self- and unsupervised denoisers.

Our method is suited to noise with correlations that run parallel to the image axes. This commonly occurs in microscopy and lab users often cannot find suitable methods to remove it. We release our code open source and strongly believe that the scientific imaging community will apply and adapt our methods in a variety of applications.

While we have achieved our results using 1-dimensional receptive fields, some imaging modalities produce noise that is correlated in multiple directions, therefore requiring a differently shaped receptive field. We found that extending the receptive field to cover both a row and a column of pixels allows the AR decoder to model some aspects of the signal, making the technique unsuitable for removing noise correlated in two dimensions. However, we believe that techniques other than shaping the AR decoder's receptive field could be discovered to limit it's modeling capabilities. Furthermore, the method is limited to zero-centered noise by the signal decoder. This precludes, *e.g.*, removing salt-and-pepper noise or restoring images with saturated pixel intensities. A direction for future work could therefore be alternative methods for mapping latent variables back into image space. We hope that future work will further improve the theoretical understanding of the method and allow us to utilize its full potential.

Acknowledgements We thank Aleš Leonardis (University of Birmingham) for a helpful discussion on the writing of this paper. Computations in this paper were performed using the University of Birmingham's BlueBEAR HPC service, which provides a High Performance Computing service to the University's research community. See http://www.birmingham.ac.uk/bear for more details.

References

[1] Asad Aali, Marius Arvinte, Sidharth Kumar, and Jonathan I Tamir. Solving inverse problems with score-based generative priors learned from noisy data. In *2023 57th Asilomar Conference on Signals, Systems, and Computers*, pages 837–843. IEEE, 2023. 2

[2] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pages 524–533. PMLR, 2019. 1, 2

[3] Coleman Broaddus, Alexander Krull, Martin Weigert, Uwe Schmidt, and Gene Myers. Removing structured noise with self-supervised blind-spot networks. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 159–163. IEEE, 2020. 2, 3, 5, 7

[4] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 2, pages 60–65. Ieee, 2005. 3

[5] Xi Chen, Diederik P Kingma, Tim Salimans, Yan Duan, Prafulla Dhariwal, John Schulman, Ilya Sutskever, and Pieter Abbeel. Variational lossy autoencoder. In *International Conference on Learning Representations*, 2017. 2, 4, 5

[6] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*. 2

[7] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising with block-matching and 3d filtering. In *Image processing: algorithms and systems, neural networks, and machine learning*, volume 6064, pages 354–365. SPIE, 2006. 1, 2, 3

[8] Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alex Dimakis, and Adam Klivans. Ambient diffusion: Learning clean distributions from corrupted data. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[9] Donal J Denvir and Emer Conroy. Electron multiplying ccds. In *Opto-Ireland 2002: Optical Metrology, Imaging, and Machine Vision*, volume 4877, pages 55–68. SPIE, 2003. 3

[10] Benoit Dupont, Antoine Dupret, Eric Belhaire, and Patrick Villard. Fpn sources in bolometric infrared detectors. *IEEE Sensors Journal*, 9(8):944–952, 2009. 3

[11] Michael Elad, Bahjat Kawar, and Gregory Vaksman. Image denoising: The deep learning revolution and beyond—a survey paper. *SIAM Journal on Imaging Sciences*, 16(3):1594–1654, 2023. 2

[12] Minho Eom, Seungjae Han, Pojeong Park, Gyuri Kim, Eun-Seo Cho, Jueun Sim, Kang-Han Lee, Seonghoon Kim, He Tian, Urs L Böhm, et al. Statistically unbiased prediction enables accurate denoising of voltage imaging data. *Nature Methods*, pages 1–12, 2023. 1, 2, 3

[13] Adam K Glaser, Kevin W Bishop, Lindsey A Barner, Etsuo A Susaki, Shimpei I Kubota, Gan Gao, Robert B Serafin, Pooja Balaram, Emily Turschak, Philip R Nicovich, et al. A hybrid open-top light-sheet microscope for versatile multi-scale imaging of cleared tissues. *Nature methods*, 19(5):613–619, 2022. 6

[14] 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. 2

[15] Ishaan Gulrajani, Kundan Kumar, Faruk Ahmed, Adrien Ali Taiga, Francesco Visin, David Vazquez, and Aaron Courville. Pixelvae: A latent variable model for natural images. *arXiv preprint arXiv:1611.05013*, 2016. 2

[16] Guy M Hagen, Justin Bendesky, Rosa Machado, Tram-Anh Nguyen, Tanmay Kumar, and Jonathan Ventura. Fluorescence microscopy datasets for training deep neural networks. *GigaScience*, 10(5):giab032, 2021. 5, 7

[17] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, pages 18–22. Springer, 2009. 5

[18] Junxian He, Daniel Spokoyny, Graham Neubig, and Taylor Berg-Kirkpatrick. Lagging inference networks and posterior collapse in variational autoencoders. In *International Conference on Learning Representations*, 2019. 4

[19] Trond R Henninen, Marta Bon, Feng Wang, Daniele Passerone, and Rolf Erni. The structure of sub-nm platinum clusters at elevated temperatures. *Angewandte Chemie International Edition*, 59(2):839–845, 2020. 6

[20] Gerlind Herberich, Reinhard Windoffer, Rudolf E Leube, and Til Aach. Signal and noise modeling in confocal laser scanning fluorescence microscopy. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part I 15*, pages 381–388. Springer, 2012. 3

[21] Eva Höck, Tim-Oliver Buchholz, Anselm Brachmann, Florian Jug, and Alexander Freytag. N2v2–fixing noise2void checkerboard artifacts with modified sampling strategies and a tweaked network architecture. *arXiv preprint arXiv:2211.08512*, 2022. 2, 7

[22] Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14781–14790, 2021. 1, 2

[23] Saeed Izadi, Darren Sutton, and Ghassan Hamarneh. Image denoising in the deep learning era. *Artificial Intelligence Review*, 56(7):5929–5974, 2023. 1

[24] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 6, 7

[25] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *stat*, 1050:1, 2014. 2

[26] Alexander Krull, Hector Basevi, Benjamin Salmon, Andre Zeug, Franziska Müller, Samuel Tonks, Leela Muppala, and Aleš Leonardis. Image denoising and the generative accumulation of photons. In *Proceedings of the IEEE/CVF Win-*

ter Conference on Applications of Computer Vision, pages 1528–1537, 2024. 2

[27] 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. 1, 2, 3

[28] Alexander Krull, Tomáš Vičar, Mangal Prakash, Manan Lalit, and Florian Jug. Probabilistic noise2void: Unsupervised content-aware denoising. *Frontiers in Computer Science*, 2:5, 2020. 1, 2

[29] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951. 4

[30] Romain F Laine, Guillaume Jacquemet, and Alexander Krull. Imaging in focus: an introduction to denoising bioimages in the era of deep learning. *The international journal of biochemistry & cell biology*, 140:106077, 2021. 1, 3

[31] Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. *Advances in Neural Information Processing Systems*, 32, 2019. 3

[32] Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17725–17734, 2022. 2, 7

[33] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. In *International Conference on Machine Learning*, pages 2965–2974. PMLR, 2018. 2, 5, 7

[34] Yougui Liao. Practical electron microscopy and database. *An Online Book*, page 1412, 2006. 3

[35] Biagio Mandracchia, Xuanwen Hua, Changliang Guo, Jeonghwan Son, Tara Urner, and Shu Jia. Fast and accurate scmos noise correction for fluorescence microscopy. *Nature communications*, 11(1):94, 2020. 3

[36] Christopher A Metzler, Ali Mousavi, Reinhard Heckel, and Richard G Baraniuk. Unsupervised learning with stein's unbiased risk estimator. *arXiv preprint arXiv:1805.10531*, 2018. 3

[37] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 2

[38] Butch Moomaw. Camera technologies for low light imaging: overview and relative advantages. *Methods in cell biology*, 114:243–283, 2013. 3

[39] Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. Noisier2noise: Learning to denoise from unpaired noisy data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12064–12072, 2020. 2

[40] Stephen W Paddock. Principles and practices of laser scanning confocal microscopy. *Molecular biotechnology*, 16:127–149, 2000. 3

[41] Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2043–2052, 2021. 2

[42] Stephen J Pennycook and Peter D Nellist. *Scanning transmission electron microscopy: imaging and analysis*. Springer Science & Business Media, 2011. 3

[43] Jan Portmann, Simon Lynen, Margarita Chli, and Roland Siegwart. People detection and tracking from aerial thermal views. In *2014 IEEE international conference on robotics and automation (ICRA)*, pages 1794–1800. IEEE, 2014. 6

[44] Mangal Prakash, Mauricio Delbracio, Peyman Milanfar, and Florian Jug. Interpretable unsupervised diversity denoising and artefact removal. In *International Conference on Learning Representations*, 2021. 1, 2, 3, 5, 7

[45] Mangal Prakash, Alexander Krull, and Florian Jug. Fully unsupervised diversity denoising with convolutional variational autoencoders. *arXiv preprint arXiv:2006.06072*, 2020. 1, 2, 3, 5

[46] Mangal Prakash, Manan Lalit, Pavel Tomancak, Alexander Krul, and Florian Jug. Fully unsupervised probabilistic noise2void. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 154–158. IEEE, 2020. 5, 7

[47] Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2self with dropout: Learning self-supervised denoising from single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1890–1898, 2020. 1, 2

[48] Benjamin Salmon and Alexander Krull. Towards structured noise models for unsupervised denoising. In *European Conference on Computer Vision*, pages 379–394. Springer, 2022. 1, 2, 3, 5

[49] Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. Ladder variational autoencoders. *Advances in neural information processing systems*, 29, 2016. 3, 5, 7

[50] Tzu-An Song, Fan Yang, and Joyita Dutta. Noise2void: unsupervised denoising of pet images. *Physics in Medicine & Biology*, 66(21):214002, 2021. 1

[51] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*. 2

[52] Aaron Van den Oord, Nal Kalchbrenner, Lasse Espeholt, Oriol Vinyals, Alex Graves, et al. Conditional image generation with pixelcnn decoders. *Advances in neural information processing systems*, 29, 2016. 2, 3

[53] Feng Wang, Trond R Henninen, Debora Keller, and Rolf Erni. Noise2atom: unsupervised denoising for scanning transmission electron microscopy images. *Applied Microscopy*, 50(1):1–9, 2020. 2

[54] Wei Wang, Fei Wen, Zeyu Yan, and Peilin Liu. Optimal transport for unsupervised denoising learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2104–2118, 2022. 2

[55] Zejin Wang, Jiazheng Liu, Guoqing Li, and Hua Han. Blind2unblind: Self-supervised image denoising with visible blind spots. In *Proceedings of the IEEE/CVF Conference*

on Computer Vision and Pattern Recognition, pages 2027–2036, 2022. 2

[56] 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. 1, 7

[57] Junshen Xu and Elfar Adalsteinsson. Deformed2self: Self-supervised denoising for dynamic medical imaging. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*, pages 25–35. Springer, 2021. 1

[58] Zhao Yang, Chenggeng Yan, and Hu Chen. Unpaired low-dose ct denoising using conditional gan with structural loss. In *2021 International Conference on Wireless Communications and Smart Grid (ICWCSG)*, pages 272–275. IEEE, 2021. 2

[59] 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 *CVPR*, 2019. 3

[60] Zhaoning Zhang, Yujie Wang, Rafael Piestun, and Zhen-Li Huang. Characterizing and correcting camera noise in back-illuminated scmos cameras. *Optics Express*, 29(5):6668–6690, 2021. 3