

Lightweight Data-Free Denoising for Detail-Preserving Biomedical Image Restoration

Tomáš Chobola^{1,2}[0009-0000-3272-9996]✉, Julia A. Schnabel^{1,2,3}[0000-0001-6107-3009], and Tingying Peng^{1,2}[0000-0002-7881-1749]

¹ School of Computation, Information and Technology, Technical University of Munich, Munich, Germany

² Helmholtz AI, Helmholtz Munich - German Research Center for Environmental Health, Neuherberg, Germany

³ School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

{tomas.chobola,tingying.peng}@helmholtz-munich.de

Abstract. Current self-supervised denoising techniques achieve impressive results, yet their real-world application is frequently constrained by substantial computational and memory demands, necessitating a compromise between inference speed and reconstruction quality. In this paper, we present an ultra-lightweight model that addresses this challenge, achieving both fast denoising and high quality image restoration. Built upon the Noise2Noise training framework—which removes the reliance on clean reference images or explicit noise modeling—we introduce an innovative multistage denoising pipeline named Noise2Detail (N2D). During inference, this approach disrupts the spatial correlations of noise patterns to produce intermediate smooth structures, which are subsequently refined to recapture fine details directly from the noisy input. Extensive testing reveals that Noise2Detail surpasses existing dataset-free techniques in performance, while requiring only a fraction of the computational resources. This combination of efficiency, low computational cost, and data-free approach make it a valuable tool for biomedical imaging, overcoming the challenges of scarce clean training data—due to rare and complex imaging modalities—while enabling fast inference for practical use.

Keywords: Image denoising · Self-supervised · Lightweight.

1 Introduction

Image denoising is one of the fundamental tasks in computational imaging, particularly in medical and biological contexts, where it seeks to eliminate visual artifacts to enhance applications like medical diagnosis and microscopy imaging [9,13,8]. State-of-the-art denoising methods, however, often grapple with achieving a balance between efficiency and performance in these fields. Typically, these approaches depend on large datasets of paired noisy and clean images to train image reconstruction models. While such techniques yield impressive results, their

utility in biomedical imaging is curtailed by the challenges of acquiring such datasets which are not only costly and time-intensive but often impractical due to the inherent scarcity and complexity of obtaining noise-free biological samples, such as high-quality tissue scans or cellular images. Moreover, models trained on generic or unrelated datasets struggle to adapt to the unique noise distributions and intricate details of biomedical images, where even subtle deviations can compromise diagnostic accuracy. This is a significant limitation in clinical and biological research, where retaining fine details—such as cellular structures or pathological markers—is paramount for precise analysis and reliable diagnosis. Consequently, there is a pressing need for denoising techniques that are both data-efficient and adaptable, capable of delivering robust performance in medical and biological imaging without the harsh requirement of extensive paired training data.

To address these issues, several unsupervised denoising methods have been proposed as potential alternatives. Methods such as Noise2Noise [6], Noise2Void (N2V) [4], Neighbor2Neighbor (Ne2Ne) [3], Zero-Shot Noise2Noise (ZS-N2N) [10], and others [2,14,15,16] aim to utilize the intrinsic characteristics or patterns within the data, bypassing the requirement for noise-free reference images, or they harness the architecture of deep neural networks to extract low-level image details. However, these methods often suffer from significant limitations, such as high computational and memory requirements, suboptimal denoising performance, or reliance on prior knowledge of the noise distribution—information that is often difficult to obtain in real-world scenarios.

In this work, we build on Noise2Noise [6] and Zero-Shot Noise2Noise [10] to develop a lightweight denoising method termed Noise2Detail (N2D) that eliminates the need for noise model knowledge, extensive pre-training or clean reference datasets. Our approach uses a simple three-layer convolutional network designed for efficient execution on both GPUs and CPUs, enhancing its practicality in resource-constrained biomedical settings like clinical diagnostics or research labs. The restoration process begins with a simple network generating a *partially restored image* after being trained on two downsampled versions of the noisy input, though background artifacts persist due to limited model capacity. Next, a multi-stage denoising step using pixel-shuffled images refines the background by breaking spatial noise correlations, albeit with a slight trade-off in foreground sharpness. Finally, we re-use the network and fine-tune its weights to sharpen these critical foreground details, guided by the original noisy image. This balance of efficiency and performance makes it an effective tool for biomedical applications requiring rapid, precise denoising. We provide the source code here: <https://github.com/ctom2/noise2detail>. The contributions of this paper are as follows:

1. **Lightweight Denoising Solution.** We propose a compact denoising model with just three convolutional layers, ensuring minimal computational cost. This enables rapid, high-quality image restoration, making it highly practical for biomedical imaging applications where computational resources are limited and fast inference is essential, such as in smart microscopy.

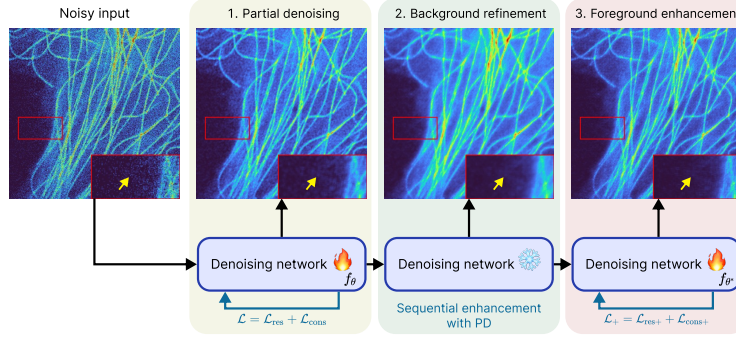


Fig. 1: Overview of our proposed Noise2Detail denoising framework. Initially, f_θ is trained to denoise an image using a single-image Noise2Noise loss, but residual artifacts persist due to the limitations of the training (see the yellow arrow in the zoomed in insets). We then freeze the weights of the network, apply a sequential refinement to remove the background artifacts, and then retrain f_{θ^*} to enhance the foreground details, yielding the final denoised image.

2. **Data-Efficient Framework.** Our three-step framework performs denoising without requiring clean reference data—an important advantage in biomedical imaging, where such data is often scarce or may not exist at all.
3. **Diverse Biomedical Data Evaluation.** We rigorously evaluate our method on CT scans with synthetic noise and a fluorescence microscopy dataset affected by real noise, demonstrating its ability to restore degraded images while preserving fine structural details.

2 Method

2.1 Revisiting Noise2Noise

In supervised image denoising, models f_θ are typically trained to map a noisy input y to an estimate $f_\theta(y)$ of the clean image x , requiring paired noisy and clean data, where $y = x + e$ and e is additive noise. Conversely, Noise2Noise [6] demonstrates that two noisy versions of the same image, $y_1 = x + e_1$ and $y_2 = x + e_2$, with e_1 and e_2 as independent noise instances, can substitute for noisy-clean pairs. The insight is that a model trained to predict y_2 from y_1 cannot learn the random noise but instead captures the consistent, correlated structure of the clean signal x , resulting in:

$$\arg \min_{\theta} \mathbb{E} [\|f_\theta(y_1) - x\|_2^2] = \arg \min_{\theta} \mathbb{E} [\|f_\theta(y_1) - y_2\|_2^2]. \quad (1)$$

With infinite paired noisy samples, Noise2Noise theoretically matches the performance of training with clean data [6], as the unpredictable noise is ignored in favor of the learnable signal. In practice, finite data causes a slight performance

drop, and acquiring two pixel-aligned noisy images of the same scene is often impractical or costly, especially in applications like biomedical imaging, limiting its actual applications.

2.2 Zero-Shot Denoising and Its Limitation

To address the limitations of Noise2Noise training, ZS-N2N [10] extends Ne2Ne [3] to operate on single images through spatial downsampling. Given a noisy image $y \in \mathbb{R}^{H \times W \times C}$, two subsampled views $D_1(y), D_2(y) \in \mathbb{R}^{H/2 \times W/2 \times C}$ are generated via stride-2 convolutions with diagonal averaging kernels:

$$k_1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, k_2 = \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \end{bmatrix}. \quad (2)$$

These kernels preserve local averages while creating pseudo-pairs for self-supervised training. The denoising network f_θ then learns a noise residual through a symmetric loss:

$$\mathcal{L}_{\text{res}} = \frac{1}{2} (\|D_1(y) - f_\theta(D_1(y)) - D_2(y)\|_2^2 + \|D_2(y) - f_\theta(D_2(y)) - D_1(y)\|_2^2), \quad (3)$$

To mitigate overfitting, the additional consistency loss enforces agreement between denoised and downsampled outputs:

$$\mathcal{L}_{\text{cons}} = \frac{1}{2} (\|D_1(y) - f_\theta(D_1(y)) - D_2(y - f_\theta(y))\|_2^2 + \|D_2(y) - f_\theta(D_2(y)) - D_1(y - f_\theta(y))\|_2^2). \quad (4)$$

The total loss $\mathcal{L} = \mathcal{L}_{\text{res}} + \mathcal{L}_{\text{cons}}$ ensures stable training, and then the final denoised image can be obtained as $\bar{x} = y - f_\theta(y)$. However, averaging diagonal pixels poorly represents true noise distribution and retains spatial correlation of real noise [18], causing artifacts and reduced performance. We use this as the first stage of our pipeline, with f_θ supporting later quality-enhancing steps.

2.3 Refining Denoised Outputs with Pixel-Shuffle Techniques

To mitigate the challenges above, we propose a two-stage refinement framework that breaks noise correlations while preserving authentic image structures from the input noisy image.

Noise Correlation Elimination. A key component of our proposed Noise2Detail method is the use of pixel-shuffle downsampling (PD) [18,5] to disrupt spatial noise patterns. For stride s , PD decomposes y into s^2 non-overlapping sub-images $\text{PD}_s(y) = \{p_{s,i}(y)\}_{i=1}^{s^2}$, that then can be again rearranged into the original image structure through an inverse operator PD^{-1} .

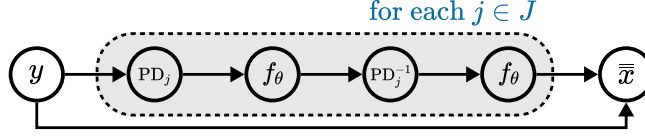


Fig. 2: Illustration of the sequential enhancement process with a trained network f_θ . For every $j \in J$, f_θ is applied twice—initially to denoise the pixel-shuffled downsampled image, and then to correct the grid-like artifacts caused by disrupted spatial correlations when reconstructing the original pixel layout.

Thus, a naive and straightforward approach would be taking the noisy input image y , applying PD to break the spatial correlations, denoise it using the trained model f_θ , and then reshuffle the pixels back to their original positions. However, this approach introduces two critical issues: (1) The discontinuity between pixels in the denoised image produces grid-like artifacts as shown in [11], degrading the quality of the outputs. (2) While PD effectively breaks the spatial correlations between noisy pixels, it inherently also disrupts the spatial correlations of the signal pixels. This rearrangement can break the original structures, leading to over-smoothed results, as the model can no longer rely on local pixel relationships.

To address those challenges, we introduce a sequential denoising refinement process. After computing the initial PD-denoised estimate \tilde{x}_j , we leverage the pre-trained model f_θ to further denoise the unshuffled image as follows:

$$\tilde{x}_j = PD_j^{-1}(PD_j(y) - f_\theta(PD_j(y))), \quad (5)$$

$$\hat{x}_j = \tilde{x}_j - f_\theta(\tilde{x}_j). \quad (6)$$

However, while this refinement step mitigates unwanted artifacts in the denoised image, the use of PD breaks the spatial correlations between pixels, resulting in a smoothing effect that blurs fine details and sometimes eliminates them completely. This happens because downsampling reduces the number of pixels available for the model to predict detailed structures, limiting its ability to reconstruct the full resolution of intricate features present in the original image. To overcome this, we blend the original partially denoised output \bar{x} with the refined outputs \hat{x}_j , effectively merging preserved structural details with smoother regions into a unified representation:

$$\bar{\bar{x}} = \frac{1}{1 + |J|} \left(\bar{x} + \sum_{j \in J} \hat{x}_j \right). \quad (7)$$

We illustrate the refinement process in Figure 2.

Detail Recovery. To restore lost foreground details, we introduce a detail enhancement process. We repurpose the network f_θ and retrain it with a modified

residual and consistency loss, as defined in Equations 8 and 9, respectively. Here, the network f_{θ^*} uses the partially denoised image \bar{x} as input and targets the original noisy image y . Since the network cannot predict the random noise in y from the denoised \bar{x} , it focuses on refining the structural details, effectively enhancing the foreground features.

$$\mathcal{L}_{\text{res}+} = \frac{1}{2} (\|f_{\theta^*}(D_1(\bar{x})) - D_2(y)\|_2^2 + \|f_{\theta^*}(D_2(\bar{x})) - D_1(y)\|_2^2), \quad (8)$$

$$\mathcal{L}_{\text{cons}+} = \frac{1}{2} (\|f_{\theta^*}(D_1(\bar{x})) - D_2(f_{\theta^*}(\bar{x}))\|_2^2 + \|f_{\theta^*}(D_2(\bar{x})) - D_1(f_{\theta^*}(\bar{x}))\|_2^2). \quad (9)$$

With the trained network f_{θ^*} , we then obtain the final denoised image as follows, $\hat{x} = f_{\theta^*}(\bar{x})$. The complete denoising framework is visualized in Figure 1 with an image from BioSR dataset [12].

3 Experiments and Results

3.1 Datasets and Performance Comparison

We conduct experiments on data with synthetic and real noise. Using a 33-slice CT scan of a spleen from the MSD dataset [1], we apply Gaussian and Poisson noise. For real-world data, we use the FMD testing set [17] consisting of 48 images, featuring noisy fluorescence microscopy images of cells, zebrafish, and mouse brain tissues, with clean ground-truth images averaged from 50 noisy observations. We compare our method to state-of-the-art self-supervised denoising methods: DIP [15], N2V [4], Ne2Ne [3], N2S [2], and ZS-N2N [10], using original training setups except for Ne2Ne and N2S, where we apply single-image adaptations from [7] to make the comparisons fair.

3.2 Implementation Details

Our model f uses a compact design with just three convolutional layers. The first two transform input channels into a 48-dimensional feature space, and the final layer converts these back to image space. With only about 22,000 parameters, it is far smaller and more efficient than existing methods, which often have millions of parameters requiring high computational costs. We train parameters θ and θ^* for 2,000 iterations at a 1×10^{-3} learning rate. For refinement, we use $j \in \{2, 4\}$, optimized based on our ablation study in Subsection 3.3. Results and inference times come from testing on an NVIDIA A100 GPU.

3.3 Ablation Study

To analyze the impact of different combinations of PD-processed images in the refinement step on denoising a spleen CT scan degraded by synthetic Poisson noise, and to justify hyperparameter choices, we conducted an ablation study. The results, detailed in Table 1, demonstrate that incorporating images processed with broken spatial correlation significantly improves the performance of our proposed framework, thus supporting our design approach. Specifically, the highest performance is achieved when combining images for $j \in \{2, 4\}$, while including additional image for $j \in \{2, 4, 8\}$ leads to a performance drop due to over-smoothing, which reduces detail visibility.

Table 1: Performance comparison based on PSNR with indicators for included variables under synthetic Poisson noise $\lambda \in [10, 50]$.

\bar{x}	\hat{x}_2	\hat{x}_4	\hat{x}_8	PSNR \uparrow (dB)
✓				30.75
✓	✓			31.48
✓	✓	✓		31.56
✓	✓	✓	✓	31.23

3.4 Results

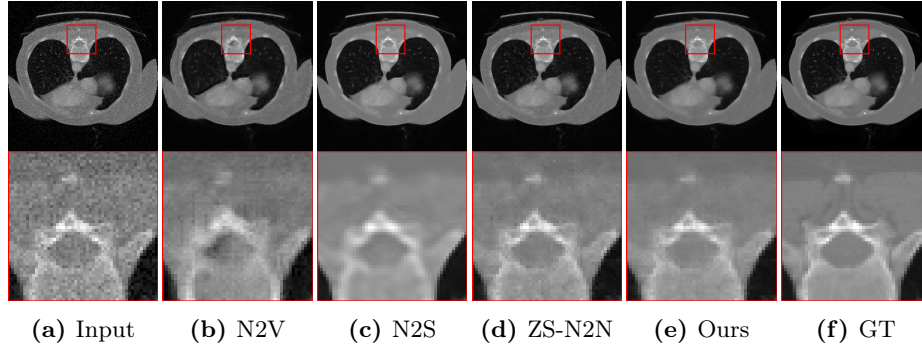
Our quantitative analysis of results for removing synthetic and real microscopy noise is shown in Tables 2 and 3. Table 2 lists the average inference times for 512×512 pixel images from the FMD dataset. Our proposed approach N2D achieves the best performance for synthetic noise removal and second best performance for real microscopy noise, outperformed only by N2S. However, the inference time of our method is significantly lower than N2S. This combination of strong performance and low computational cost emphasizes the strength of our method, positioning it as a competitive alternative to larger and expensive methods. Figures 3 and 4 visualize the denoising results, demonstrating that our method uniquely delivers high-quality results without hallucinating background artifacts, and maintaining high levels of detail fidelity.

4 Conclusion

In this work, we present Noise2Detail, a lightweight denoising framework designed for biomedical imaging, where clean data is scarce and the use of large models are often not feasible. Built on Noise2Noise, our three-layer network employs a sequential process of refinement, background correction, and detail enhancement. It restores images efficiently without requiring noise model assumptions or extensive datasets. Evaluations on medical and biological datasets

Table 2: Performance comparison across different synthetic noise levels with PSNR \uparrow (dB). Best result is in **bold**, second best result is underlined.

Noise	DIP	N2V	Ne2Ne	N2S	ZS-N2N	N2D (ours)
Gaussian $\sigma = 25$	27.31	27.59	19.34	<u>29.18</u>	28.97	29.36
Gaussian $\sigma \in [10, 50]$	26.62	26.94	18.87	<u>28.81</u>	28.47	28.84
Poisson $\lambda \in [10, 50]$	27.08	28.65	19.97	<u>31.50</u>	31.28	31.56
Average	27.00	27.73	19.40	<u>29.83</u>	29.57	29.92

**Fig. 3:** Our method is the only one that maintains the fidelity of the foreground while keeping the background smooth without artifacts.**Table 3:** Performance comparison across subsets of FMD [17] with PSNR \uparrow (dB), and average inference times per image. Best result is in **bold**, second best result is underlined.

Subset	DIP	N2V	Ne2Ne	N2S	ZS-N2N	N2D (ours)
Confocal_BPAAE	31.07	33.58	14.93	<u>34.92</u>	34.82	35.35
Confocal_FISH	23.99	30.26	14.82	31.86	30.46	<u>30.80</u>
Confocal_MICE	29.60	35.06	11.98	36.80	35.95	<u>36.22</u>
TwoPhoton_BPAAE	27.35	32.19	16.79	33.59	32.69	<u>32.95</u>
TwoPhoton_MICE	25.41	32.59	16.62	33.33	32.71	<u>33.14</u>
WideField_BPAAE	28.40	25.98	17.16	<u>25.54</u>	25.83	26.20
Average	27.64	31.61	15.38	32.67	32.08	<u>32.44</u>
Inference time (s)	82	42	413	1250	13	24

confirm its ability to preserve details and remove noise effectively. This scalable method supports rapid denoising, enhancing biomedical research applications.

Acknowledgments. Tomáš Chobola is supported by the Helmholtz Association under the joint research school "Munich School for Data Science - MUDS".

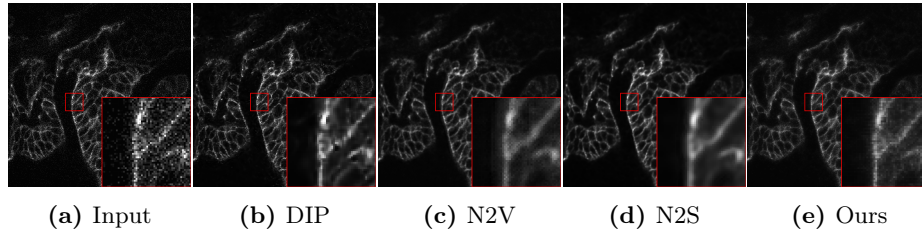


Fig. 4: Examples of restoration given a noisy image of a zebrafish from a confocal microscope taken from FMD [17]. N2V and N2S oversmooth and blur the details while our method maintains the structural fidelity.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., van Ginneken, B., Bilello, M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M.J., Heckers, S.H., Huisman, H., Jarnagin, W.R., McHugo, M.K., Napel, S., Pernicka, J.S.G., Rhode, K., Tobon-Gomez, C., Vorontsov, E., Meakin, J.A., Ourselin, S., Wiesenfath, M., Arbeláez, P., Bae, B., Chen, S., Daza, L., Feng, J., He, B., Isensee, F., Ji, Y., Jia, F., Kim, I., Maier-Hein, K., Merhof, D., Pai, A., Park, B., Perslev, M., Rezaiifar, R., Rippel, O., Sarasua, I., Shen, W., Son, J., Wachinger, C., Wang, L., Wang, Y., Xia, Y., Xu, D., Xu, Z., Zheng, Y., Simpson, A.L., Maier-Hein, L., Cardoso, M.J.: The medical segmentation decathlon. *Nature Communications* **13**(1) (Jul 2022). <https://doi.org/10.1038/s41467-022-30695-9>, <http://dx.doi.org/10.1038/s41467-022-30695-9>
2. Batson, J., Royer, L.: Noise2self: Blind denoising by self-supervision. In: *International Conference on Machine Learning*. pp. 524–533. PMLR (2019)
3. Huang, T., Li, S., Jia, X., Lu, H., Liu, J.: Neighbor2neighbor: Self-supervised denoising from single noisy images. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 14781–14790 (2021)
4. Krull, A., Buchholz, T.O., Jug, F.: Noise2void-learning denoising from single noisy images. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 2129–2137 (2019)
5. Lee, W., Son, S., Lee, K.M.: 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*. pp. 17725–17734 (2022)
6. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T.: Noise2Noise: Learning image restoration without clean data. In: Dy, J., Krause, A. (eds.) *Proceedings of the 35th International Conference on Machine Learning. Proceedings of Machine Learning Research*, vol. 80, pp. 2965–2974. PMLR (10–15 Jul 2018), <https://proceedings.mlr.press/v80/lehtinen18a.html>
7. Lequyer, J., Philip, R., Sharma, A., Pelletier, L.: Noise2fast: fast self-supervised single image blind denoising. *arXiv preprint arXiv:2108.10209* (2021)

8. Li, R., della Maggiora, G., Andriasyan, V., Petkidis, A., Yushkevich, A., Deshpande, N., Kudryashev, M., Yakimovich, A.: Microscopy image reconstruction with physics-informed denoising diffusion probabilistic model. *Communications Engineering* **3**(1) (Dec 2024). <https://doi.org/10.1038/s44172-024-00331-z>, <http://dx.doi.org/10.1038/s44172-024-00331-z>
9. Li, X., Li, Y., Zhou, Y., Wu, J., Zhao, Z., Fan, J., Deng, F., Wu, Z., Xiao, G., He, J., Zhang, Y., Zhang, G., Hu, X., Chen, X., Zhang, Y., Qiao, H., Xie, H., Li, Y., Wang, H., Fang, L., Dai, Q.: Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. *Nature Biotechnology* **41**(2), 282–292 (Sep 2022). <https://doi.org/10.1038/s41587-022-01450-8>, <http://dx.doi.org/10.1038/s41587-022-01450-8>
10. Mansour, Y., Heckel, R.: Zero-shot noise2noise: Efficient image denoising without any data. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 14018–14027 (2023)
11. Pan, Y., Liu, X., Liao, X., Cao, Y., Ren, C.: Random sub-samples generation for self-supervised real image denoising. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 12150–12159 (2023)
12. Qiao, C., Li, D.: BioSR: a biological image dataset for super-resolution microscopy (11 2020). <https://doi.org/10.6084/m9.figshare.13264793.v9>, <https://figshare.com/articles/dataset/BioSR/13264793>
13. Qu, L., Zhao, S., Huang, Y., Ye, X., Wang, K., Liu, Y., Liu, X., Mao, H., Hu, G., Chen, W., Guo, C., He, J., Tan, J., Li, H., Chen, L., Zhao, W.: Self-inspired learning for denoising live-cell super-resolution microscopy. *Nature Methods* **21**(10), 1895–1908 (Sep 2024). <https://doi.org/10.1038/s41592-024-02400-9>, <http://dx.doi.org/10.1038/s41592-024-02400-9>
14. Quan, Y., Chen, M., Pang, T., Ji, H.: Self2self with dropout: Learning self-supervised denoising from single image. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 1890–1898 (2020)
15. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 9446–9454 (2018)
16. Wang, J., Di, S., Chen, L., Ng, C.W.W.: Noise2info: Noisy image to information of noise for self-supervised image denoising. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 16034–16043 (2023)
17. Zhang, Y., Zhu, Y., Nichols, E., Wang, Q., Zhang, S., Smith, C., Howard, S.: A poisson-gaussian denoising dataset with real fluorescence microscopy images. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 11710–11718 (2019)
18. Zhou, Y., Jiao, J., Huang, H., Wang, Y., Wang, J., Shi, H., Huang, T.: When awgn-based denoiser meets real noises. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 34, pp. 13074–13081 (2020)