WED-PM-159
21/06/2023

Zero-Shot Noise2Noise: Efficient Image Denoising without any Data

Youssef Mansour and Reinhard Heckel

Technical University of Munich
Munich Center for Machine Learning

Motivation

- Networks trained on datasets achieve SOTA denoising performance, but building a dataset is difficult
- Dataset free methods require heavy compute, have poor performance, or do not generalize to different noise distributions
- We propose a novel dataset free algorithm that performs well on different noise models and levels, and is fast to execute even on a CPU

Setup

- Zero-Shot: only noisy test image is given
- Blind: no information on noise distribution or level

Related work

- Supervised: clean-noisy image pairs

- Self-Supervised: only noisy images

- Zero-Shot: no data available

BM3D[1] Deep Image Prior (DIP)[2] Self2Self (S2S)[3]

Drawbacks of existing zero-shot methods

- BM3D: works well only for Gaussian noise and requires noise level as input

- DIP: poor performance & early stopping iteration is critical

- S2S:
  - Long denoising time (1.25 hrs for one 256 x 256 img)
  - Works bad in regime of low noise levels
  - Relies heavily on ensembling

❖ Goal: reach a good trade-off between performance and compute

Method

1) Convolve the noisy test image with two fixed filters, which yields two downsampled images

2) Train a lightweight network with regularization to map one downsampled image to the other

Elements

Downsampling scheme

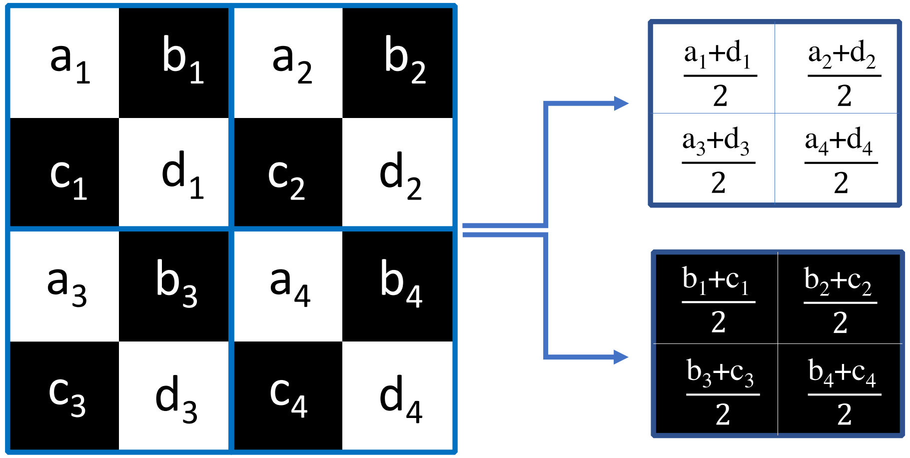
Loss function

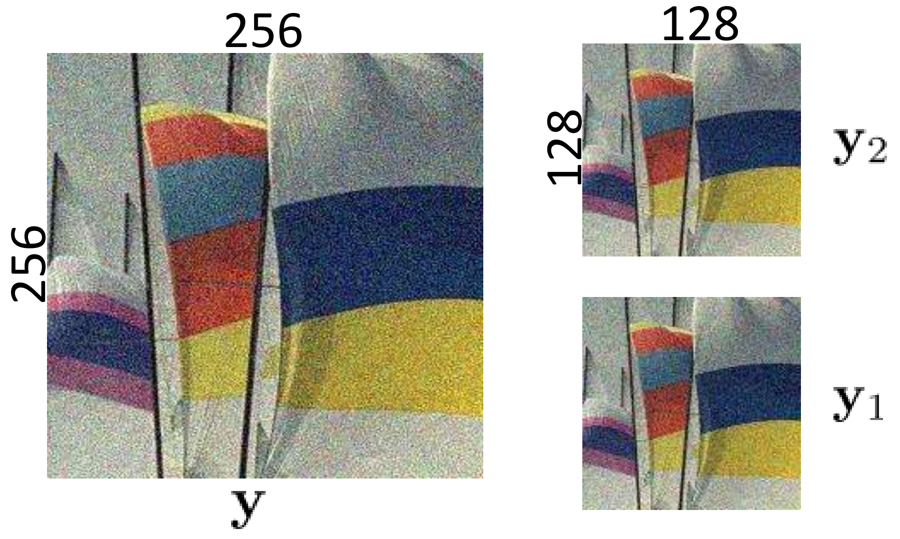
Lightweight network

Downsampling Scheme

motivated by Neighbour2Neighbour [5]

- Assuming that nearby pixels of clean image are highly correlated, while noise is independent and unstructured
- Downsample a noisy image into a pair of smaller images, which is an approximation of two noisy observations of the same clean image





$$\mathbf{k_1} = \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \end{bmatrix} \quad \mathbf{k_2} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

2D depthwise convolution, with stride 2

$$\mathbf{y}_1 = \mathbf{y} \circledast \mathbf{k_1} \qquad \mathbf{y}_2 = \mathbf{y} \circledast \mathbf{k_2}$$

Loss Function

- Residual learning

$$\mathcal{D}_{f_{\boldsymbol{\theta}}}(\mathbf{y}) = \mathbf{y} - f_{\boldsymbol{\theta}}(\mathbf{y})$$

$$\mathcal{L}(\boldsymbol{\theta}) = \|\mathcal{D}_{f_{\boldsymbol{\theta}}}(\mathbf{y}_1) - \mathbf{y}_2\|_2^2$$

motivated by Noise2Noise [4]

- Symmetric loss $\mathcal{L}_{\text{res.}}(\boldsymbol{\theta}) = \|\mathcal{D}_{f_{\boldsymbol{\theta}}}(\mathbf{y}_1) - \mathbf{y}_2\|_2^2 + \|\mathcal{D}_{f_{\boldsymbol{\theta}}}(\mathbf{y}_2) - \mathbf{y}_1\|_2^2$

- Consisteny loss: to prevent overfitting/early stopping

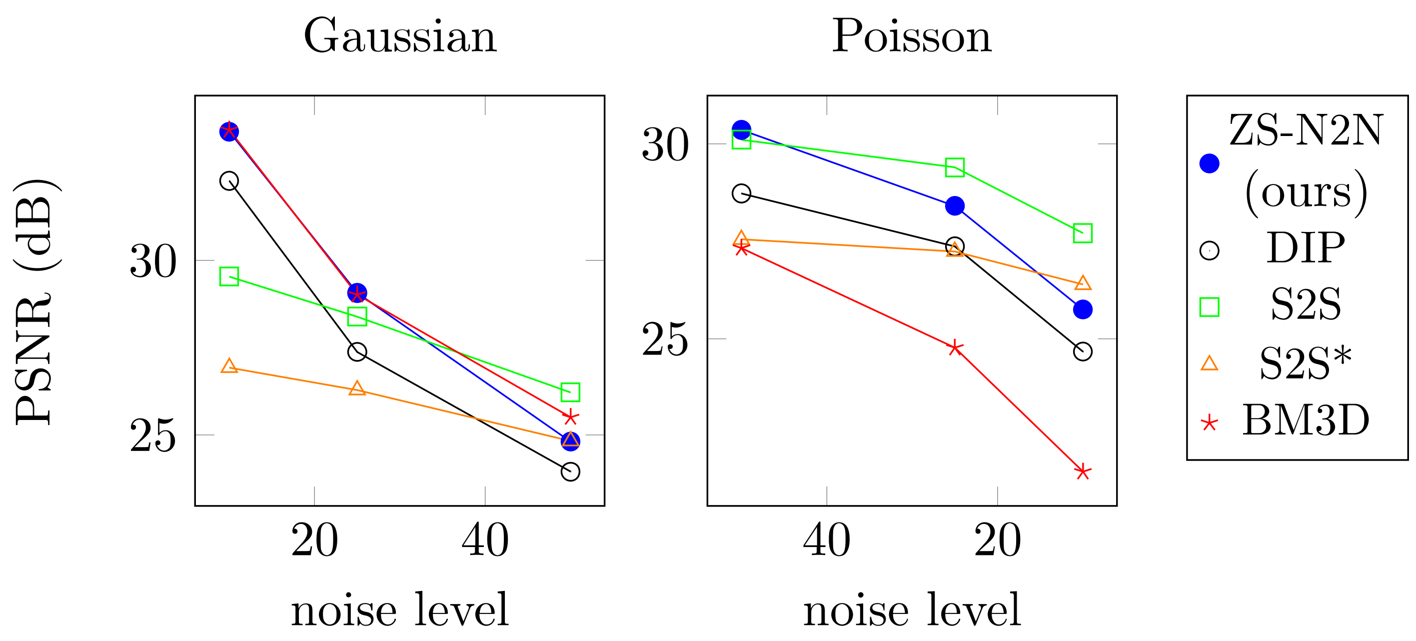
$$\mathcal{L}_{\text{cons.}}(\boldsymbol{\theta}) = \|\mathcal{D}_{f_{\boldsymbol{\theta}}}(\mathbf{y}_1) - \mathcal{D}_{f_{\boldsymbol{\theta}}}(\mathbf{y})_1\|_2^2 + \|\mathcal{D}_{f_{\boldsymbol{\theta}}}(\mathbf{y}_2) - \mathcal{D}_{f_{\boldsymbol{\theta}}}(\mathbf{y})_2\|_2^2$$

- Total loss $\mathcal{L}(\boldsymbol{\theta}) = \mathcal{L}_{\text{res.}}(\boldsymbol{\theta}) + \mathcal{L}_{\text{cons.}}(\boldsymbol{\theta})$

Lightweight Network $f_{\boldsymbol{\theta}}$

- 2 Layers: 3x3 CNN, ReLU, 3x3 CNN, ReLU, 1x1 CNN $\sim$ 20k parameters

Artificial noise



- BM3D's performance drops on Poisson noise
- DIP is worse than other baselines
- S2S requires ensembling for good performance and fails on low noise levels
- Our method generalizes best

Natural noise

Dataset	ZS-N2N	DIP	S2S	S2S*	BM3D
PolyU	36.92	**37.07**	37.01	33.12	36.11
SIDD	34.07	**34.31**	33.98	30.77	28.19

Real-world camera noise

RGB

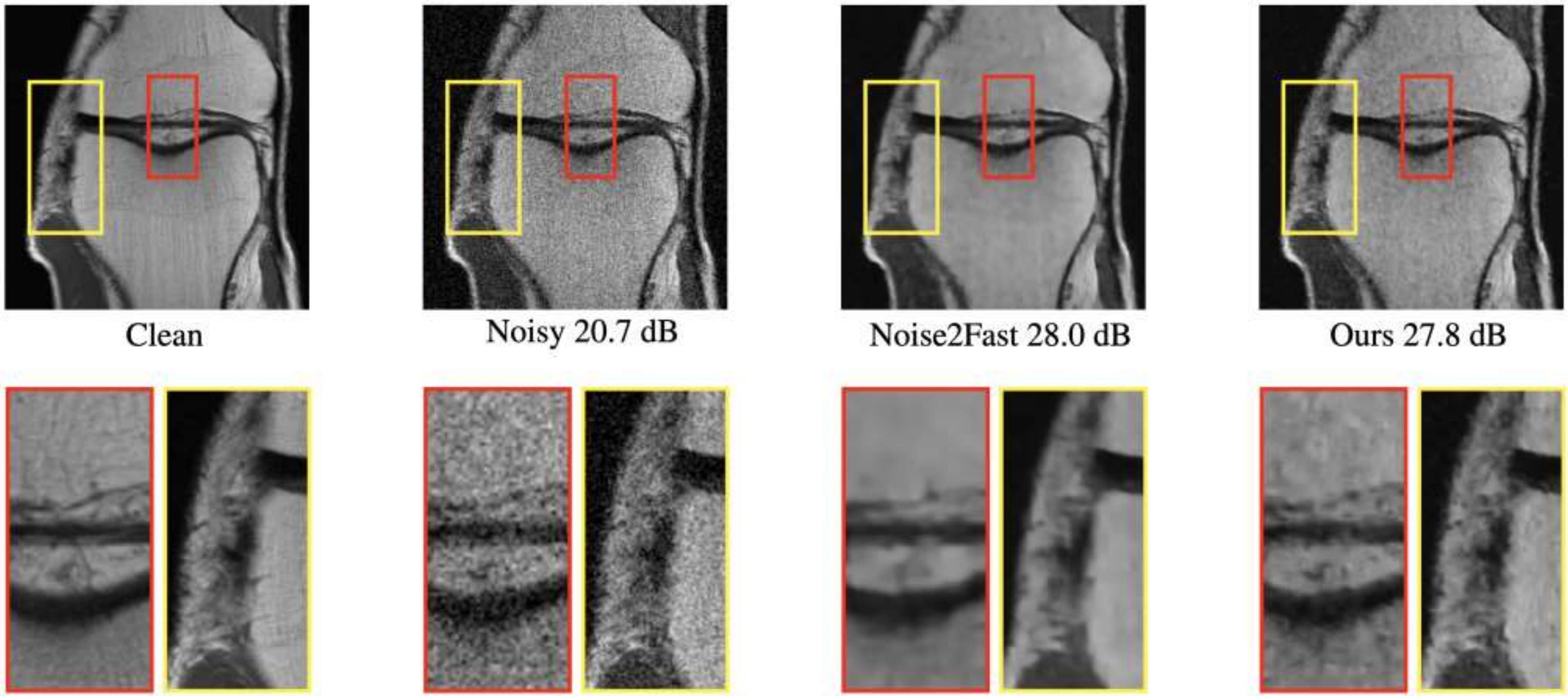
Image	Photon BPAE	Photon Mice	Confocal BPAE	Average
ZS-N2N	30.73	31.42	35.85	**32.67**
DIP	29.22	30.01	35.51	31.58
S2S	30.90	**31.51**	31.01	31.14
S2S*	29.49	29.99	29.54	29.67
BM3D	27.19	29.48	33.23	29.97
N2F [6]	**30.93**	31.07	**36.01**	**32.67**

Real-world microscope noise (FMDD)

Gray scale

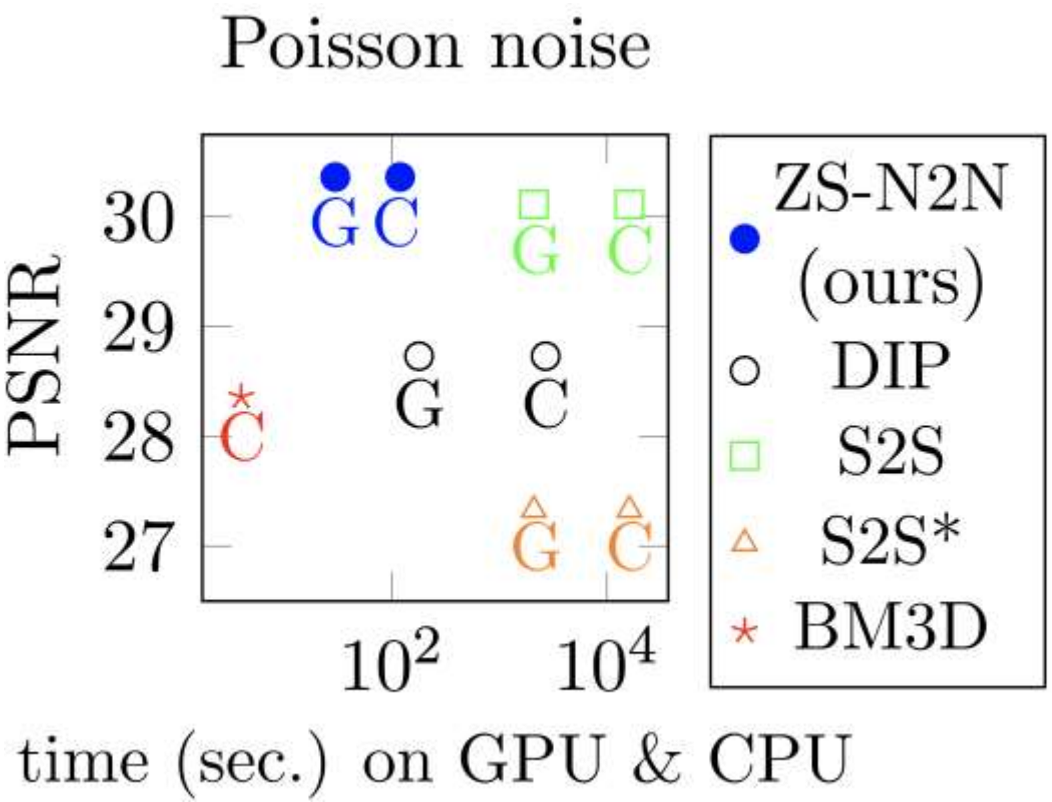
Further Experiments

Newly proposed zero-shot method for grayscale denoising: Noise2Fast [6]



- ➢ Our method produces sharper images

Compute



Method	ZS-N2N Ours	DIP	S2S
Network size	22k	2.2M	1M

Conclusion

- Current zero-shot methods have poor performance or require heavy compute

- Proposed a new zero-shot method that:
  - Performs well
  - Requires moderate compute (Time, Memory, CPU)
  - Good generalization

References

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. “Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering”. In: IEEE Transactions on Image Processing. 2007.
[2] D. Ulyanov, A. Vedaldi, and V. Lempitsky. “Deep Image Prior”. In: IEEE/CVF Conferenceon Computer Vision and Pattern Recognition. 2018.
[3] Y. Quan, M. Chen, T. Pang, and H. Ji. “Self2Self With Dropout: Learning Self-Supervised Denoising From Single Image”. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
[4] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. “Noise2Noise: Learning Image Restoration without Clean Data”. In: International Conference on Machine Learning. 2018.
[5] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu. “Neighbor2Neighbor: Self-Supervised Denoising from Single Noisy Images”. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
[6] J. Lequyer, R. Philip, A. Sharma, W.-H. Hsu, and L. Pelletier. “A Fast Blind Zero-Shot Denoiser”. In: Nature Machine Intelligence. 2022.