# Supplementary Material

## A. Sample Reconstructions

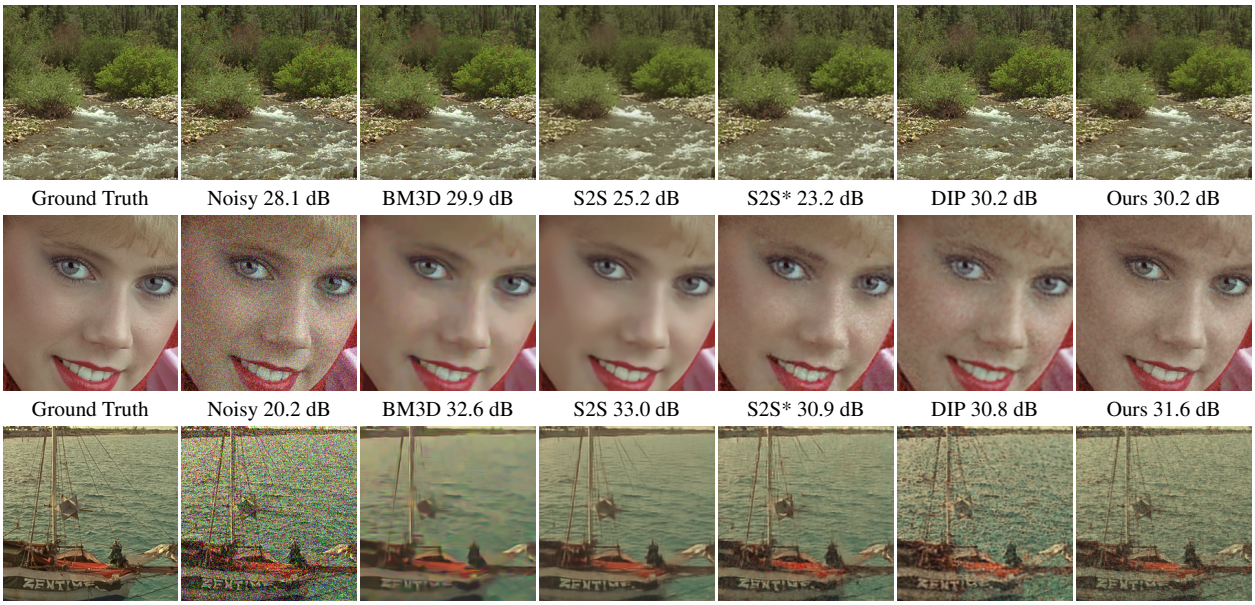


Figure 4. Gaussian denoising on Kodak24 images. Upper row: $\sigma = 10$, middle row: $\sigma = 25$, lower row; $\sigma = 50$. Note how Self2Self fails on the low noise level (top row, $\sigma = 10$), and produces an image noisier than the input noisy image.

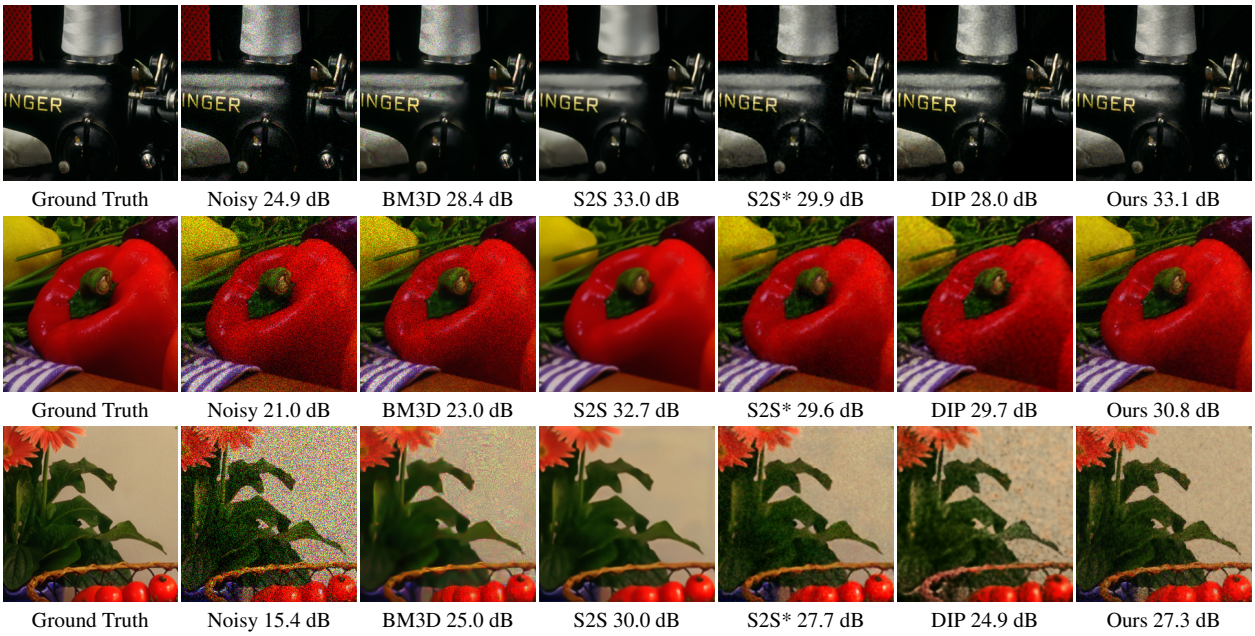


Figure 5. Poisson denoising on McMaster18 images. Upper row: $\lambda = 50$, middle row: $\lambda = 25$, lower row; $\lambda = 10$. Note the inconsistency of BM3D's performance, as it relies on noise level estimation, which varies from image to image.

## B. Ablation Studies

In this section we provide additional experiments and discuss a few variants of our proposed approach to show which elements are essential for good performance. Unless otherwise mentioned, the ablation studies are conducted on the Kodak24 dataset contaminated with Gaussian noise of $\sigma = 25$.

**Loss function** We study 3 variations of our proposed loss function, namely without the symmetric loss, without the consistency loss, and without the residual loss. The results are displayed in table 5. The symmetric loss offers minor improvement to the method's performance, where as the consistency loss has a more significant impact. However the residual loss is necessary, since without it the network just learns the identity mapping.

| Default | w/o symmetric | w/o consistency | w/o residual |
|---|---|---|---|
| 29.07 | 28.65 | 28.01 | 17.93 |

Table 5. Denoising PSNR in dB of ablated versions of the loss function.

**Network size** We saw that compared to deep learning based algorithms, ZS-N2N has few network parameters. In this section, we show that, perhaps surprisingly, even with much fewer parameters, ZS-N2N can still perform well. Moreover, we show that denoising with a UNet fails. This is most likely due to overfitting, as only a single test image is used for training. The results are depicted in table 6. Even with a network as small as 500 parameters, ZS-N2N outperforms DIP that has 2 million parameters.

| Network size | UNet (3.3M) | Default (22k) | 4k | 2k | 1k | 500 |
|---|---|---|---|---|---|---|
| PSNR | 21.01 | 29.07 | 28.66 | 28.28 | 28.07 | 27.71 |

Table 6. Denoising PSNR in dB for reducing the number of parameters of the ZS-N2N network. The network size was reduced by decreasing the number of channels in the hidden layers.

**Data scaling** In the previous experiments, the dataset based methods were trained on only 500 images, and therefore exhibited slightly worse performance than dataset free ones. In this section, we unveil the potential of the supervised Noise2Clean by additionally training on 4000 and 10000 images. As before, a UNet with 3.3M parameters is trained on ImageNet images.

The results are shown in figure 6. Already at 4000 training images, N2C significantly outperforms all other dataset free methods. These findings coincide with the results in the literature, that supervised dataset based methods achieve state-of-the-art results, given enough training data and similarity between the training and test sets.

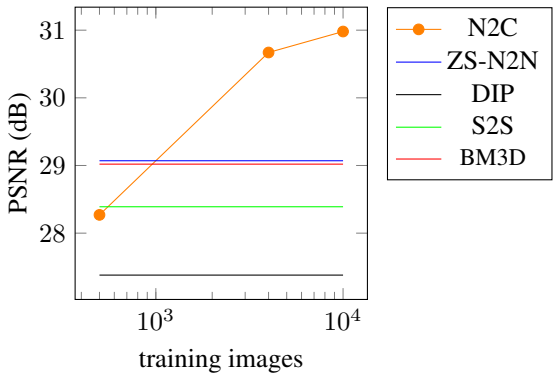


Figure 6. Denoising performance of Noise2Clean as the training set is scaled to larger sizes. Note that the performance of the dataset free methods is constant, since they do not make use of any training data.

**Performance vs optimization iterations** DIP's performance is sensitive to the number of gradient descent iterations. The optimal early stopping point for DIP varies according to noise type and level, which makes it hard to determine in advance. However, unlike DIP and similar to S2S, ZS-N2N's performance only improves with the optimization steps. An example is shown in figure 7. This enables ZS-N2N to be deployed in various use cases with no manual interference or fine tuning.

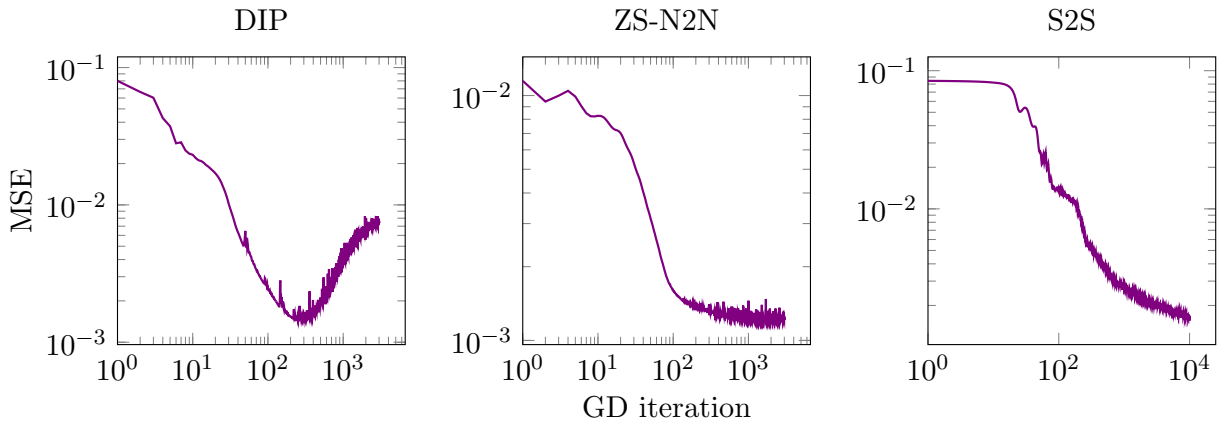


Figure 7. Denoising one image from the Kodak24 dataset. The y-axis is the MSE between the clean image and the output of the network (denoised image). The x-axis is the gradient descent iterations.

## C. Appendix

**Weaknesses and Limitations** ZS-N2N exhibits strong denoising performance and outperforms or is on par with other baselines in low and moderate noise levels. However, in high noise levels such as Gaussian noise with $\sigma = 50$, or Poission noise with $\lambda = 10$, a performance drop is noticed. ZS-N2N's performance in high noise levels is still better than DIP and BM3D, but worse than S2S. Nevertheless, ZS-N2N generalizes better than S2S, since it's performance in the high noise regime is acceptable and above other baselines, while S2S's performance in the low noise regime is poor. Moreover, even with high noise, one could still choose ZS-N2N over S2S if only limited compute is available or short denoising duration is required.

Another weakness that ZS-N2N shares with all dataset free methods, is that they do not make use of training data. Therefore in use cases where abundant data is available that is similar to the test data, dataset based methods will significantly outperform zero-shot methods as seen in the ablation studies.

**Proof of equation 1:**

**Proposition.** *Let $\|\cdot\|_2^2$ denote the squared $\mathcal{L}_2$ norm, and $\boldsymbol{\theta}$ the trainable parameters of a network $f$. Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be two noisy fixed observations of the same clean image $\mathbf{x}$, i.e. $\mathbf{y}_1 = \mathbf{x} + \mathbf{e}_1$ and $\mathbf{y}_2 = \mathbf{x} + \mathbf{e}_2$, where $\mathbf{e}_i$ is noise. Given that the $\mathbf{e}_i$ are independent, and $\mathbb{E}[\mathbf{e}] = 0$, the optimization problem w.r.t the MSE of Noise2Noise is the same as that of Noise2Clean.*

Proof:

$$
\begin{aligned}
\boldsymbol{\theta}_{\mathrm{N2C}} &= \arg\min_{\boldsymbol{\theta}} \mathbb{E}\left[\|f_{\boldsymbol{\theta}}(\mathbf{y}_1) - \mathbf{x}\|_2^2\right] \\
&= \arg\min_{\boldsymbol{\theta}} \mathbb{E}\left[\|f_{\boldsymbol{\theta}}(\mathbf{y}_1)\|_2^2 - 2\mathbf{x}^{\mathsf{T}} f_{\boldsymbol{\theta}}(\mathbf{y}_1)\right]. \\
\boldsymbol{\theta}_{\mathrm{N2N}} &= \arg\min_{\boldsymbol{\theta}} \mathbb{E}\left[\|f_{\boldsymbol{\theta}}(\mathbf{y}_1) - \mathbf{y}_2\|_2^2\right] \\
&= \arg\min_{\boldsymbol{\theta}} \mathbb{E}\left[\|f_{\boldsymbol{\theta}}(\mathbf{y}_1) - \mathbf{x} - \mathbf{e}_2\|_2^2\right] \\
&= \arg\min_{\boldsymbol{\theta}} \mathbb{E}\left[\|f_{\boldsymbol{\theta}}(\mathbf{y}_1)\|_2^2 - 2\mathbf{x}^{\mathsf{T}} f_{\boldsymbol{\theta}}(\mathbf{y}_1) - 2\mathbf{e}_2{}^{\mathsf{T}} f_{\boldsymbol{\theta}}(\mathbf{y}_1)\right] \\
&= \arg\min_{\boldsymbol{\theta}} \mathbb{E}\left[\|f_{\boldsymbol{\theta}}(\mathbf{y}_1)\|_2^2 - 2\mathbf{x}^{\mathsf{T}} f_{\boldsymbol{\theta}}(\mathbf{y}_1)\right] \\
&= \boldsymbol{\theta}_{\mathrm{N2C}},
\end{aligned}
$$

which concludes the proof. Here, the second to last equality follows from the noise being independent and having zero mean.