Supplementary Materials of
Neighbor2Neighbor: Self-Supervised Denoising from Single Noisy Images

A. Proof of Theorem 1

Theorem 1. *Let* $\mathbf{y}$ *and* $\mathbf{z}$ *be two independent noisy images conditioned on* $\mathbf{x}$*, and assume that there exists an* $\varepsilon \neq \mathbf{0}$ *such that* $\mathbb{E}_{\mathbf{y}|\mathbf{x}}(\mathbf{y}) = \mathbf{x}$ *and* $\mathbb{E}_{\mathbf{z}|\mathbf{x}}(\mathbf{z}) = \mathbf{x} + \varepsilon$*. Let the variance of* $\mathbf{z}$ *be* $\boldsymbol{\sigma}_{\mathbf{z}}^2$*. Then it holds that*

$$\begin{aligned}\mathbb{E}_{\mathbf{x},\mathbf{y}} \|f_\theta(\mathbf{y}) - \mathbf{x}\|_2^2 = \mathbb{E}_{\mathbf{x},\mathbf{y},\mathbf{z}} \|f_\theta(\mathbf{y}) - \mathbf{z}\|_2^2 - \boldsymbol{\sigma}_{\mathbf{z}}^2 \\ + 2\boldsymbol{\varepsilon}\mathbb{E}_{\mathbf{x},\mathbf{y}}(f_\theta(\mathbf{y}) - \mathbf{x}).\end{aligned} \tag{1}$$

Proof. First, similar to the derivation in Section 2 of the supplementary materials of [1], we have

$$\begin{aligned}\mathbb{E}_{\mathbf{y}|\mathbf{x}} \|f_\theta(\mathbf{y}) - \mathbf{x}\|_2^2 &= \mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}} \|f_\theta(\mathbf{y}) - \mathbf{z} + \mathbf{z} - \mathbf{x}\|_2^2 \\ &= \mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}} \|f_\theta(\mathbf{y}) - \mathbf{z}\|_2^2 + \mathbb{E}_{\mathbf{z}|\mathbf{x}} \|\mathbf{z} - \mathbf{x}\|_2^2 \\ &+ 2\mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}}(f_\theta(\mathbf{y}) - \mathbf{z})^\top(\mathbf{z} - \mathbf{x}) \\ &= \mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}} \|f_\theta(\mathbf{y}) - \mathbf{z}\|_2^2 + \boldsymbol{\sigma}_{\mathbf{z}}^2 \\ &+ 2\mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}}(f_\theta(\mathbf{y}) - \mathbf{x} + \mathbf{x} - \mathbf{z})^\top(\mathbf{z} - \mathbf{x}) \\ &= \mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}} \|f_\theta(\mathbf{y}) - \mathbf{z}\|_2^2 + \boldsymbol{\sigma}_{\mathbf{z}}^2 \\ &+ 2\mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}}(f_\theta(\mathbf{y}) - \mathbf{x})^\top(\mathbf{z} - \mathbf{x}) \\ &+ 2\mathbb{E}_{\mathbf{z}|\mathbf{x}}(\mathbf{x} - \mathbf{z})^\top(\mathbf{z} - \mathbf{x}) \\ &= \mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}} \|f_\theta(\mathbf{y}) - \mathbf{z}\|_2^2 - \boldsymbol{\sigma}_{\mathbf{z}}^2 \\ &+ 2\mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}}(f_\theta(\mathbf{y}) - \mathbf{x})^\top(\mathbf{z} - \mathbf{x}).\end{aligned} \tag{2}$$

Due to the independence between $\mathbf{y}$ and $\mathbf{z}$ given $\mathbf{x}$, it holds that

$$\begin{aligned}\mathbb{E}_{\mathbf{y}|\mathbf{x}} \|f_\theta(\mathbf{y}) - \mathbf{x}\|_2^2 &= \mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}} \|f_\theta(\mathbf{y}) - \mathbf{z}\|_2^2 - \boldsymbol{\sigma}_{\mathbf{z}}^2 \\ &+ 2\mathbb{E}_{\mathbf{y}|\mathbf{x}}(f_\theta(\mathbf{y}) - \mathbf{x})^\top \mathbb{E}_{\mathbf{z}|\mathbf{x}}(\mathbf{z} - \mathbf{x}) \\ &= \mathbb{E}_{\mathbf{y},\mathbf{z}|\mathbf{x}} \|f_\theta(\mathbf{y}) - \mathbf{z}\|_2^2 - \boldsymbol{\sigma}_{\mathbf{z}}^2 \\ &+ 2\boldsymbol{\varepsilon}\mathbb{E}_{\mathbf{y}|\mathbf{x}}(f_\theta(\mathbf{y}) - \mathbf{x}).\end{aligned} \tag{3}$$

Since $\mathbb{E}_{\mathbf{x},\mathbf{y}} = \mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}|\mathbf{x}}$, we further have

$$\begin{aligned}\mathbb{E}_{\mathbf{x},\mathbf{y}} \|f_\theta(\mathbf{y}) - \mathbf{x}\|_2^2 = \mathbb{E}_{\mathbf{x},\mathbf{y},\mathbf{z}} \|f_\theta(\mathbf{y}) - \mathbf{z}\|_2^2 - \boldsymbol{\sigma}_{\mathbf{z}}^2 \\ + 2\boldsymbol{\varepsilon}\mathbb{E}_{\mathbf{x},\mathbf{y}}(f_\theta(\mathbf{y}) - \mathbf{x}).\end{aligned} \tag{4}$$

□

B. Details of Fix-Location Sampling Strategy

Here, we give an illustrative example to describe the details of fix-location sampling strategy. The fix-location sampler randomly generates a pair of sub-sampled images from k^2 to-be-chosen sub-sampled images. In each sub-sampled image, all pixels are from the same location of all the $k \times k$ cells. In Figure A.1, $k = 2$, and locations chosen for four sub-sampled images (red, blue, yellow, and green pixels) are totally the same in each 2×2 cell on the left. Consequently, four sub-sampled images are generated, filled in red, blue, yellow, and green in the middle. Then, the sub-sampled paired images $(g_1(\mathbf{y}), g_2(\mathbf{y}))$ are randomly selected which is shown as the blue patch and the green patch on the right.

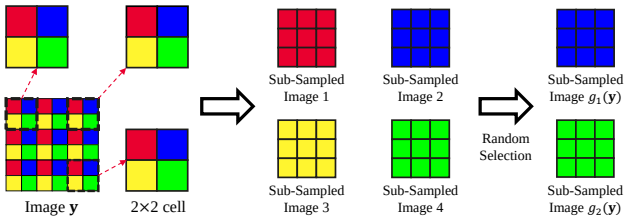

Figure A.1: Example of image pair generation with a fix-location sub-sampler $G = (g_1, g_2)$. Best viewed in color.

C. More Visualization Comparisons

C.1. More Results on Synthetic Experiments

In Figure A.2 and Figure A.3, we provide more qualitative comparisons of the denoised images for Gaussian noise with noise level $\sigma = 25$ and Poisson noise with noise level $\lambda = 30$.

C.2. Additional Visualization Results

Different from Section C.1, here we compare the visual quality of the denoised images by different denoisers on noisy images with varied noise level. Figure A.4 and Figure A.5 show the visual comparison of our methods against other competing methods in the setting of Gaussian $\sigma \in [5, 50]$ and Poisson $\lambda \in [5, 50]$ respectively. In detail, all network-based denoising methods are trained on the synthetic noise of Gaussian $\sigma \in [5, 50]$ or Poisson $\lambda \in [5, 50]$. While testing, we use the noisy images of noise level σ or λ with the fixed noise level 5, 10, 25 or 50 respectively.

C.3. More Results on Real-world Experiments

In Figure A.6, we show more denoising results on the SIDD Benchmark dataset.

References

[1] Magauiya Zhussip, Shakarim Soltanayev, and Se Young Chun. Extending stein's unbiased risk estimator to train deep denoisers with correlated pairs of noisy images. In *Advances in Neural Information Processing Systems*, volume 32, pages 1465–1475, 2019. 1

Figure A.2: Visual comparison of our method against other competing methods in the setting of Gaussian $\sigma = 25$. The quantitative PSNR(dB)/SSIM results are listed underneath the images. Best viewed in color.

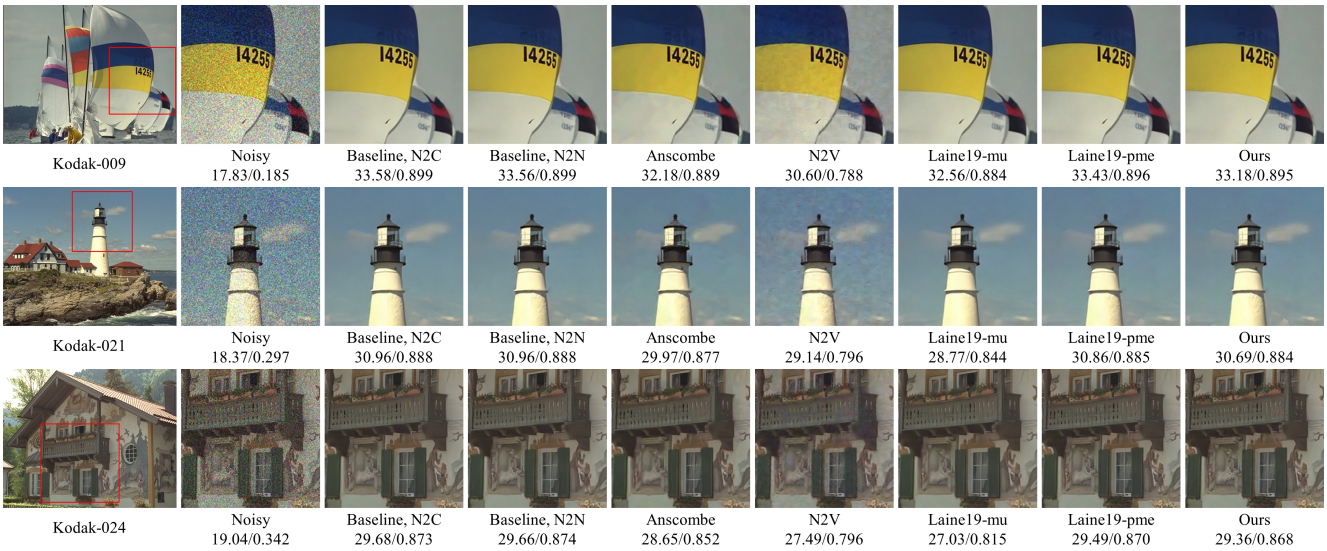

Figure A.3: Visual comparison of our method against other competing methods in the setting of Poisson $\lambda = 30$. The quantitative PSNR(dB)/SSIM results are listed underneath the images. Best viewed in color.

Figure A.4: Visual comparison of our method against other competing methods in the setting of Gaussian $\sigma \in [5, 50]$. The denoised results of four noisy images with fixed Gaussian noise level $\sigma = 5, 10, 25, 50$ are evaluated on the same denoising model which is trained on noisy images with Gaussian $\sigma \in [5, 50]$. The quantitative PSNR(dB)/SSIM results are listed underneath the images. Best viewed in color.

Figure A.5: Visual comparison of our method against other competing methods in the setting of Poisson $\lambda \in [5, 50]$. The denoised results of four noisy images with fixed Poisson noise level $\lambda = 5, 10, 25, 50$ are evaluated on the same denoising model which is trained on noisy images with Poisson $\lambda \in [5, 50]$. The quantitative PSNR(dB)/SSIM results are listed underneath the images. Best viewed in color.

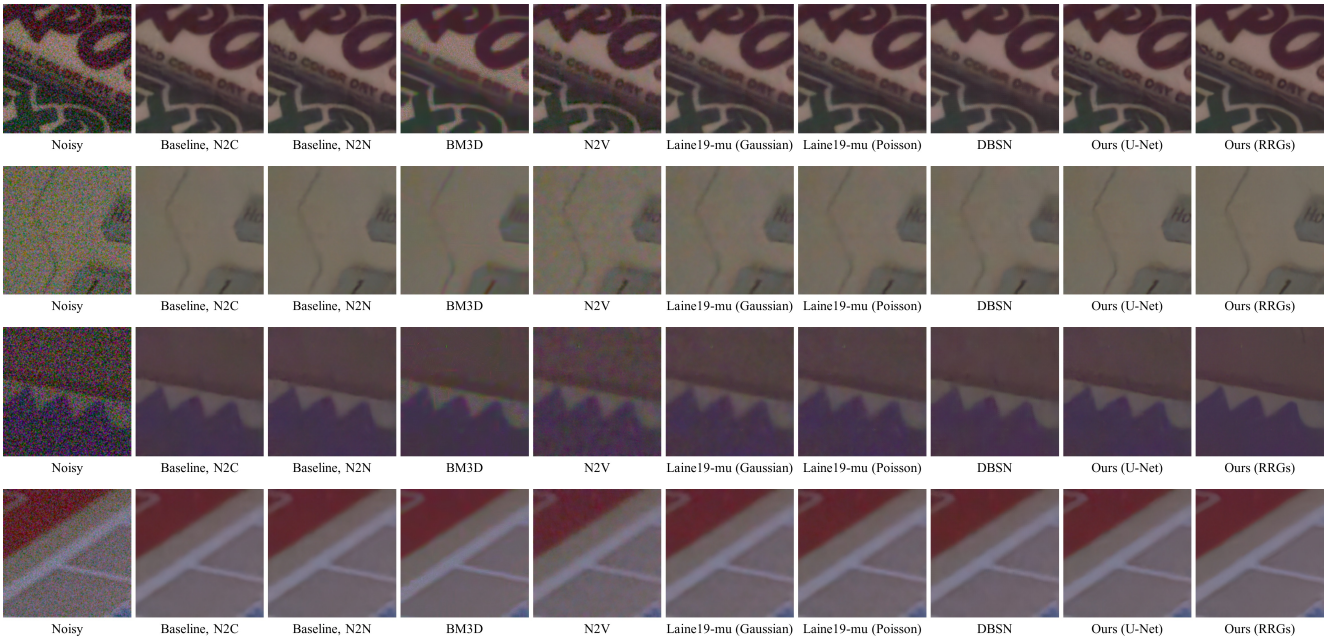

Figure A.6: Visual comparison of our method against other methods on SIDD Benchmark. All images are converted from raw-RGB space to sRGB space by the ISP provided by SIDD[1] for visualization. Best viewed in color.

[1] https://github.com/AbdoKamel/simple-camera-pipeline