

Recorrupted-to-Recorrupted: Unsupervised Deep Learning for Image Denoising (Supplemental Materials)

Tongyao Pang¹, Huan Zheng¹, Yuhui Quan², and Hui Ji¹

¹Department of Mathematics, National University of Singapore, 119076, Singapore

²School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
matpt@nus.edu.sg, huan_zheng@u.nus.edu, csyhquan@scut.edu.cn, and matjh@nus.edu.sg

1. Proof of Theorem 1

Proof. Denote $\hat{\mathbf{n}} = \mathbf{n} + \mathbf{A}\mathbf{z}$ and $\tilde{\mathbf{n}} = \mathbf{n} - \mathbf{B}\mathbf{z}$. That is

$$\begin{pmatrix} \hat{\mathbf{n}} \\ \tilde{\mathbf{n}} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{I} & -\mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{n} \\ \mathbf{z} \end{pmatrix}. \quad (1)$$

Since $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \Sigma_x)$, $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and they are independent, we have

$$\begin{pmatrix} \hat{\mathbf{n}} \\ \tilde{\mathbf{n}} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma'), \quad (2)$$

where

$$\begin{aligned} \Sigma' &= \begin{pmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{I} & -\mathbf{B} \end{pmatrix} \begin{pmatrix} \Sigma_x & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{A}^\top & -\mathbf{B}^\top \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_x + \mathbf{A}\mathbf{A}^\top & \Sigma_x - \mathbf{A}\mathbf{B}^\top \\ \Sigma_x - \mathbf{B}\mathbf{A}^\top & \Sigma_x + \mathbf{B}\mathbf{B}^\top \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_x + \mathbf{A}\mathbf{A}^\top & \mathbf{0} \\ \mathbf{0} & \Sigma_x + \mathbf{B}\mathbf{B}^\top \end{pmatrix}. \end{aligned} \quad (3)$$

Thus, $\hat{\mathbf{n}}$ and $\tilde{\mathbf{n}}$ are also independent Gaussian random variables. It yields that

$$\mathbb{E}_{\mathbf{x}, \hat{\mathbf{n}}, \tilde{\mathbf{n}}} \left\{ \tilde{\mathbf{n}}^\top \mathcal{F}_\theta(\mathbf{x} + \hat{\mathbf{n}}) \right\} = 0.$$

Then our loss function can be rewritten as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}; \mathbf{A}, \mathbf{B}) &= \mathbb{E}_{\mathbf{y}, \mathbf{z}} \|\mathcal{F}_\theta(\mathbf{y} + \mathbf{A}\mathbf{z}) - (\mathbf{y} - \mathbf{B}\mathbf{z})\|_2^2 = \mathbb{E}_{\mathbf{x}, \hat{\mathbf{n}}, \tilde{\mathbf{n}}} \|\mathcal{F}_\theta(\mathbf{x} + \hat{\mathbf{n}}) - (\mathbf{x} + \tilde{\mathbf{n}})\|_2^2 \\ &= \mathbb{E}_{\mathbf{x}, \hat{\mathbf{n}}, \tilde{\mathbf{n}}} \left\{ \|\mathcal{F}_\theta(\mathbf{x} + \hat{\mathbf{n}}) - \mathbf{x}\|_2^2 - 2\tilde{\mathbf{n}}^\top \mathcal{F}_\theta(\mathbf{x} + \hat{\mathbf{n}}) + 2\tilde{\mathbf{n}}^\top \mathbf{x} + \tilde{\mathbf{n}}^\top \tilde{\mathbf{n}} \right\} \\ &= \mathbb{E}_{\mathbf{x}, \hat{\mathbf{n}}, \tilde{\mathbf{n}}} \left\{ \|\mathcal{F}_\theta(\mathbf{x} + \hat{\mathbf{n}}) - \mathbf{x}\|_2^2 + \tilde{\mathbf{n}}^\top \tilde{\mathbf{n}} \right\} \\ &= \mathbb{E}_{\mathbf{x}, \hat{\mathbf{n}}, \tilde{\mathbf{n}}} \|\mathcal{F}_\theta(\mathbf{x} + \hat{\mathbf{n}}) - \mathbf{x}\|_2^2 + \mathbb{E}_{\mathbf{x}} \text{trace}(\Sigma_x + \mathbf{B}\mathbf{B}^\top) \\ &= \tilde{\mathcal{L}}(\boldsymbol{\theta}; \mathbf{A}) + \text{const.} \end{aligned} \quad (4)$$

The proof is done. \square

2. Running time

The inference time of processing the whole BSD68 dataset and SIDD Benchmark is around 115 seconds and 22 seconds respectively, on a NVIDIA TITAN RTX GPU with 24GB Memory. The reason why SIDD Benchmark is larger but takes less

time for inference is that the images in SIDD Benchmark are of the same size and can be processed in batch (a batch size of 32 is used by us), while the images in BSD68 vary in size and are processed one by one. Another cause is that, the AWGN denoiser is trained on the recorrupted images with specific noise level and thus the testing images in BSD68 are recorrupted for multiple times for prediction: $\sum_{j=1}^{50} \mathcal{F}_{\theta^*}(\mathbf{u} + \mathbf{A}\mathbf{z}^j)$, while for SIDD Benchmark, the single forward prediction $\mathcal{F}_{\theta^*}(\mathbf{u})$ is enough since the trained real-world image denoiser is blind to noise level.

3. Visual Comparison of More Examples

In this section, we provide visual comparison of more examples on AWGN denoising and real-world image denoising. See Fig. 1 – 6.

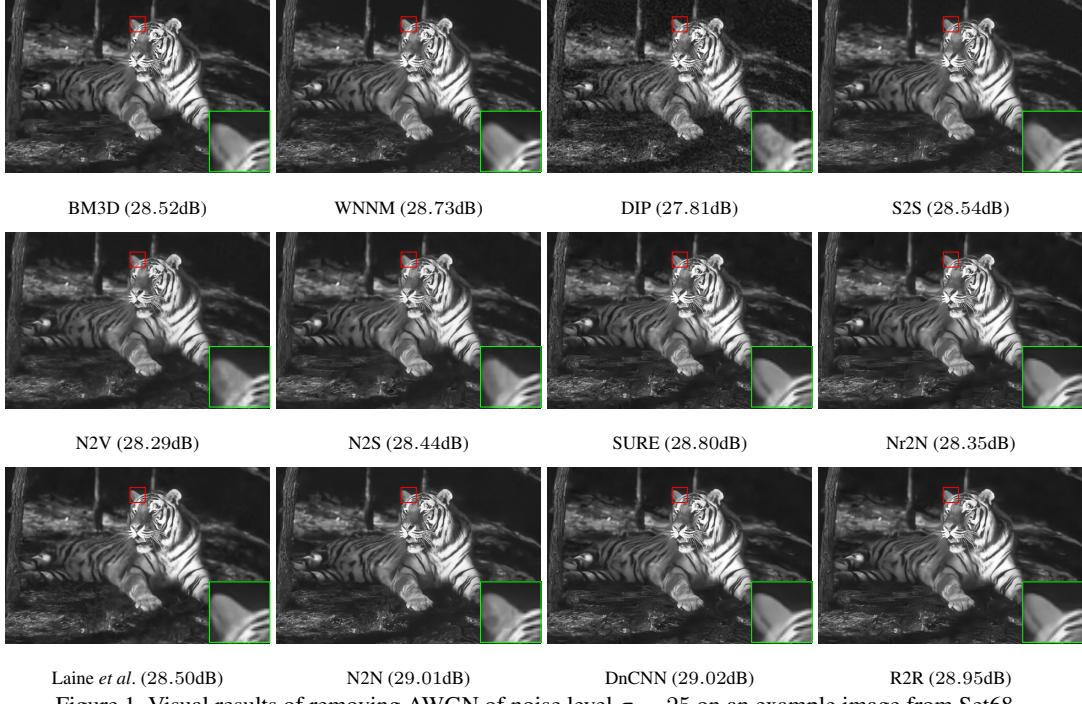


Figure 1. Visual results of removing AWGN of noise level $\sigma = 25$ on an example image from Set68.

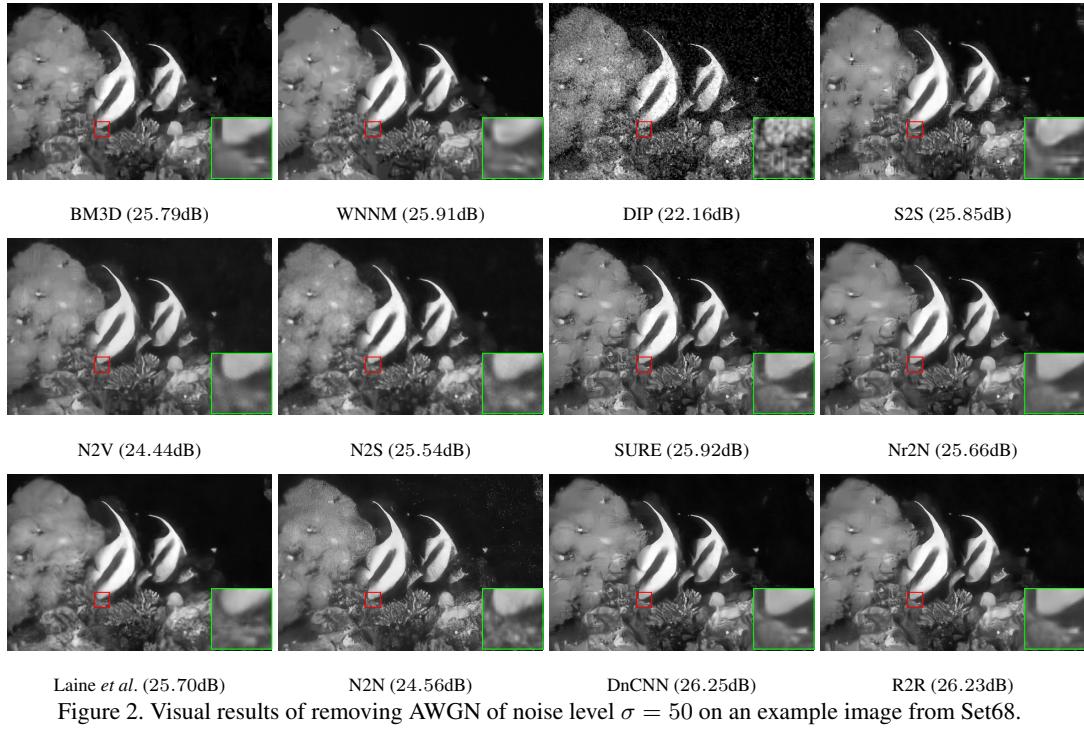


Figure 2. Visual results of removing AWGN of noise level $\sigma = 50$ on an example image from Set68.

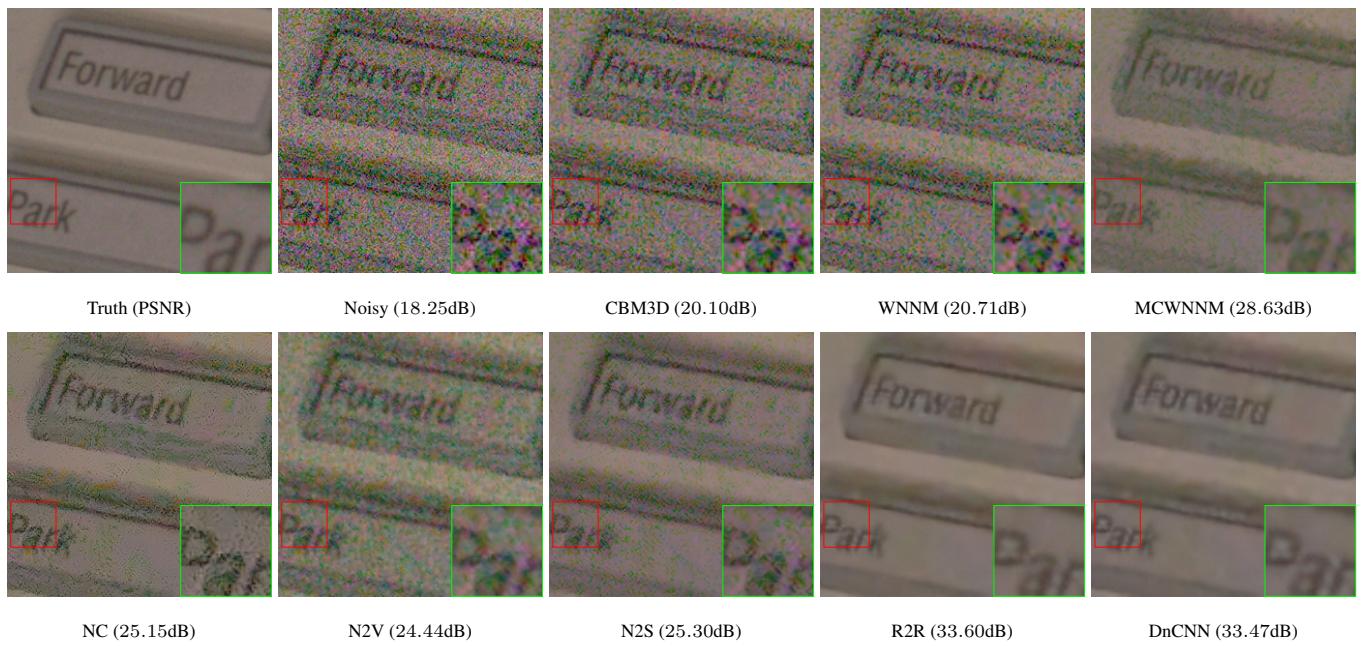


Figure 3. Visual comparison of the results from different methods when denoising an example image from SIDD Validation.

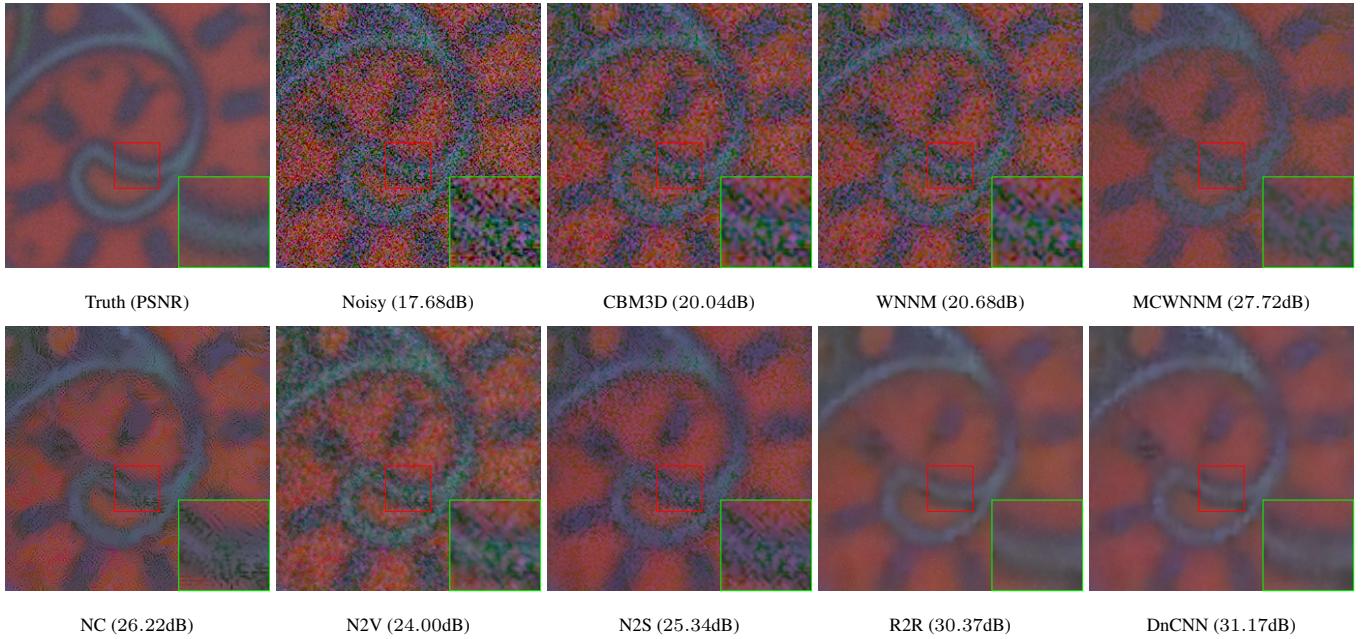


Figure 4. Visual comparison of the results from different methods when denoising an example image from SIDD Validation.

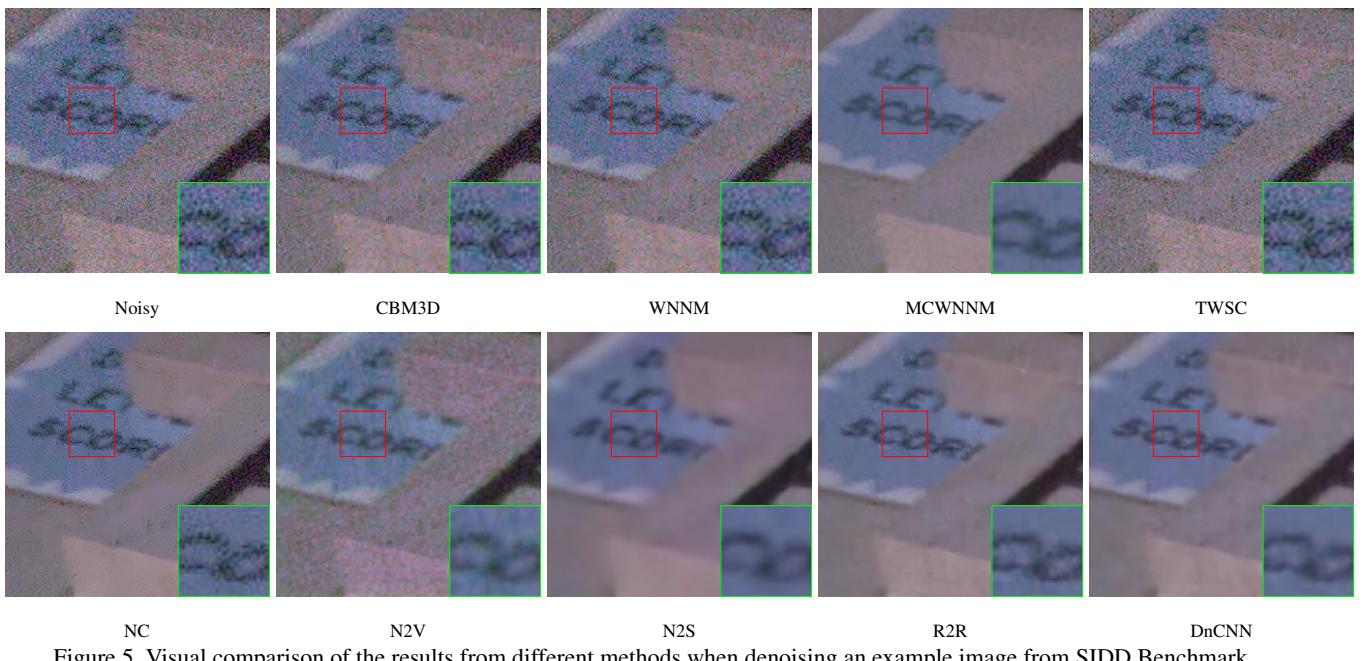


Figure 5. Visual comparison of the results from different methods when denoising an example image from SIDD Benchmark.

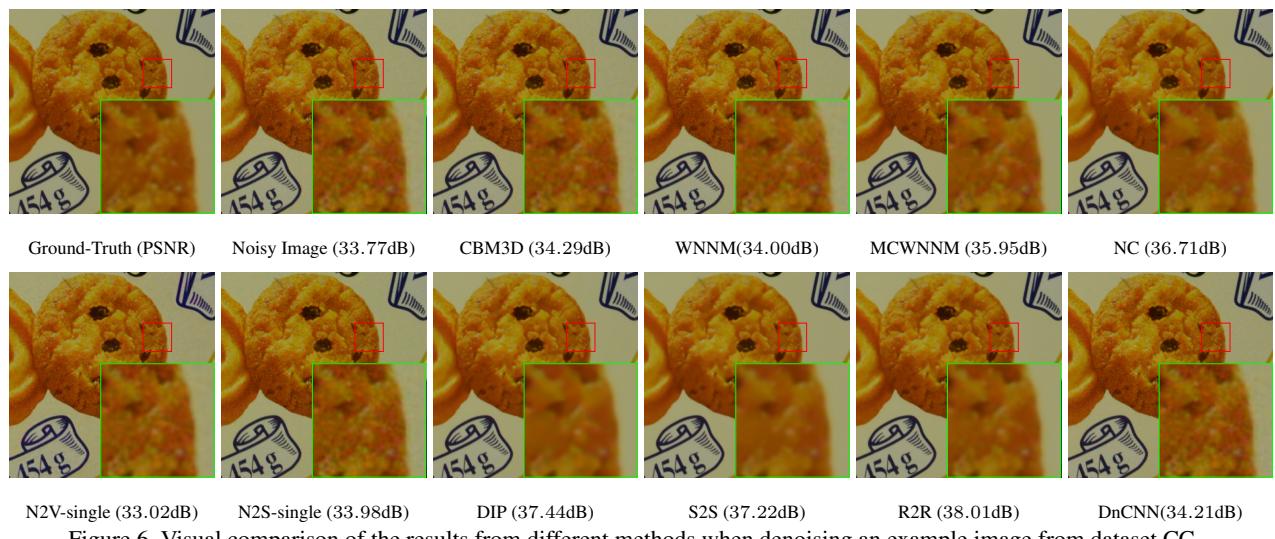


Figure 6. Visual comparison of the results from different methods when denoising an example image from dataset CC.