Suplementary Material for Score Priors Guided Deep Variational Inference for Unsupervised Real-World Single Image Denoising

Jun Cheng, Tao Liu, Shan Tan*
Huazhong University of Science and Technology, Wuhan, China
{jcheng24, hust_liutao, shantan}@hust.edu.cn

S1. Derivations

S1.1. Proof of Theorem 1

Proof. Suppose a noisy observation $\widetilde{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{n} \in \mathbb{R}^N$ where $\boldsymbol{n} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ has the zero mean and covariance Σ, and $\boldsymbol{x} \sim p(\boldsymbol{x})$. Then the density $p(\widetilde{\boldsymbol{x}})$ is $p(\widetilde{\boldsymbol{x}}) = \int \mathcal{N}(\boldsymbol{x}, \Sigma) p(\boldsymbol{x}) d\boldsymbol{x}$. The gradient of $p(\widetilde{\boldsymbol{x}})$ is

$$\begin{aligned}
\nabla_{\widetilde{\boldsymbol{x}}} p(\widetilde{\boldsymbol{x}}) &= \nabla_{\widetilde{\boldsymbol{x}}} \int \frac{1}{(2\pi)^{0.5N} |\Sigma|^{0.5}} \exp\left\{-\frac{1}{2}(\widetilde{\boldsymbol{x}} - \boldsymbol{x})^T \Sigma^{-1} (\widetilde{\boldsymbol{x}} - \boldsymbol{x})\right\} p(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int \frac{1}{(2\pi)^{0.5N} |\Sigma|^{0.5}} \exp\left\{-\frac{1}{2}(\widetilde{\boldsymbol{x}} - \boldsymbol{x})^T \Sigma^{-1} (\widetilde{\boldsymbol{x}} - \boldsymbol{x})\right\} \nabla_{\widetilde{\boldsymbol{x}}} \left(-\frac{1}{2}(\widetilde{\boldsymbol{x}} - \boldsymbol{x})^T \Sigma^{-1} (\widetilde{\boldsymbol{x}} - \boldsymbol{x})\right) p(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int \Sigma^{-1} (\boldsymbol{x} - \widetilde{\boldsymbol{x}}) p(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} = \Sigma^{-1} \int (\boldsymbol{x} - \widetilde{\boldsymbol{x}}) p(\widetilde{\boldsymbol{x}}, \boldsymbol{x}) d\boldsymbol{x} = \Sigma^{-1} \int (\boldsymbol{x} - \widetilde{\boldsymbol{x}}) p(\boldsymbol{x}|\widetilde{\boldsymbol{x}}) p(\widetilde{\boldsymbol{x}}) d\boldsymbol{x}
\end{aligned} \tag{S1}$$

Multiplying both sides of (S1) by $\frac{1}{p(\widetilde{\boldsymbol{x}})}$ gives

$$\frac{\nabla_{\widetilde{\boldsymbol{x}}} p(\widetilde{\boldsymbol{x}})}{p(\widetilde{\boldsymbol{x}})} = \nabla_{\widetilde{\boldsymbol{x}}} \log p(\widetilde{\boldsymbol{x}}) = \Sigma^{-1} \int (\boldsymbol{x} - \widetilde{\boldsymbol{x}}) p(\boldsymbol{x}|\widetilde{\boldsymbol{x}}) d\boldsymbol{x} = \Sigma^{-1} \left(\mathrm{E}_{p(\boldsymbol{x}|\widetilde{\boldsymbol{x}})}(\boldsymbol{x}) - \widetilde{\boldsymbol{x}}\right) \tag{S2}$$

where $\mathrm{E}_{p(\boldsymbol{x}|\widetilde{\boldsymbol{x}})}(\boldsymbol{x})$ is the conditional mean for noisy input $\widetilde{\boldsymbol{x}}$ corrupted by structured Gaussian noise. The proof is completed. □

S1.2. Score priors for the optimization of $\mathrm{KL}(q(\boldsymbol{x})||p(\boldsymbol{x}))$

Assume variational posterior $q(\boldsymbol{x}) = \prod_{i=1}^{N} \mathcal{N}(\mu_i, \sigma_i^2)$, then we have

$$\mathrm{KL}(q(\boldsymbol{x})||p(\boldsymbol{x})) \approx \mathrm{E}_{q(\boldsymbol{x})} \log q(\boldsymbol{x}) - \frac{1}{M} \sum_{m=1}^{M} \log p(\boldsymbol{x}_m) = -\frac{1}{2} \sum_{i=1}^{N} \log \sigma_i^2 + \mathrm{const} - \frac{1}{M} \sum_{m=1}^{M} \log p(\boldsymbol{x}_m) \tag{S3}$$

where $\boldsymbol{x}_m = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, I)$, and $\odot$ denotes element-wise multiplication.

Note that $\log p(\boldsymbol{x}_m)$ is not directly computable, but its derivatives can be obtained with the help of score functions from MMSE Non- $i.i.d$ Gaussian denoisers, that is

$$\nabla_{\boldsymbol{\mu}} \mathrm{KL}(q(\boldsymbol{x})||p(\boldsymbol{x})) \approx \nabla_{\boldsymbol{\mu}} \left(-\frac{1}{M} \sum_{m=1}^{M} \log p(\boldsymbol{x}_m)\right) = -\frac{1}{M} \sum_{m=1}^{M} \nabla_{\boldsymbol{x}_m} \log p(\boldsymbol{x}_m) \odot \nabla_{\boldsymbol{\mu}} \boldsymbol{x}_m = -\frac{1}{M} \sum_{m=1}^{M} \nabla_{\boldsymbol{x}_m} \log p(\boldsymbol{x}_m) \tag{S4}$$

$$\begin{aligned}
\nabla_{\boldsymbol{\sigma}^2} \mathrm{KL}(q(\boldsymbol{x})||p(\boldsymbol{x})) \approx \nabla_{\boldsymbol{\sigma}^2} \left(-\frac{1}{2} \sum_{i=1}^{N} \log \sigma_i^2 - \frac{1}{M} \sum_{m=1}^{M} \log p(\boldsymbol{x}_m)\right) &= -\frac{1}{2\boldsymbol{\sigma}^2} - \frac{1}{M} \sum_{m=1}^{M} \nabla_{\boldsymbol{x}_m} \log p(\boldsymbol{x}_m) \odot \nabla_{\boldsymbol{\sigma}^2} \boldsymbol{x}_m \\
&= -\frac{1}{2\boldsymbol{\sigma}^2} - \frac{1}{M} \sum_{m=1}^{M} \nabla_{\boldsymbol{x}_m} \log p(\boldsymbol{x}_m) \odot \boldsymbol{\epsilon}
\end{aligned} \tag{S5}$$

We can then replace the score in (S4) and (S5) with

$$\nabla_{\boldsymbol{x}_m} \log p(\boldsymbol{x}_m) = (\mathrm{E}(X|\boldsymbol{x}_m) - \boldsymbol{x}_m) \odot \frac{1}{\boldsymbol{\sigma}^2} \approx (\mathcal{G}(\boldsymbol{x}_m) - \boldsymbol{x}_m) \odot \frac{1}{\boldsymbol{\sigma}^2} \in \mathbb{R}^N \tag{S6}$$

where $\mathrm{E}(X|\boldsymbol{x}_m)$ is the conditional expectation given noisy input $\boldsymbol{x}_m$ and $\mathcal{G}(\boldsymbol{x}_m)$ is the output of a MMSE Non-$i.i.d$ Gaussian denoiser $\mathcal{G}$, which serves as the approximation of $\mathrm{E}(X|\boldsymbol{x}_m)$.

S1.3. Derivations of the complete ELBO objective and its components

The KL divergence between the trivial variational posterior $q(X, \Phi, Z, \Omega)$ and the true posterior $p(X, \Phi, Z, \Omega|\boldsymbol{y})$ is

$$\begin{aligned}
\mathrm{KL}(q(X, \Phi, Z, \Omega)||p(X, \Phi, Z, \Omega|\boldsymbol{y})) &= \int_{X,\Phi,Z,\Omega} \log \frac{q(X, \Phi, Z, \Omega)}{p(X, \Phi, Z, \Omega|\boldsymbol{y})} q(X, \Phi, Z, \Omega) dX d\Phi dZ d\Omega \\
&= \int_{X,\Phi,Z,\Omega} \log \frac{q(X, \Phi, Z, \Omega)p(y)}{p(\boldsymbol{y}|X, \Phi, Z, \Omega)p(X, \Phi, Z, \Omega)} q(X, \Phi, Z, \Omega) dX d\Phi dZ d\Omega \\
&= \log p(\boldsymbol{y}) - \underbrace{(\mathrm{E}_{q(X,\Phi,Z,\Omega)}(\log p(\boldsymbol{y}|X, \Phi, Z, \Omega)) - \mathrm{KL}(q(X, \Phi, Z, \Omega)||p(X, \Phi, Z, \Omega)))}_{\text{ELBO}}
\end{aligned} \tag{S7}$$

As $\mathrm{KL}(q(X, \Phi, Z, \Omega)||p(X, \Phi, Z, \Omega|\boldsymbol{y})) \geq 0$ and $\log p(\boldsymbol{y})$ is not computable, we instead maximize the Evidence Low Bound (ELBO) in Eq. (S7). $\mathrm{KL}(q(X, \Phi, Z, \Omega)||p(X, \Phi, Z, \Omega)$ in the ELBO can be further decomposed into

$$\begin{aligned}
\mathrm{KL}(q(X, \Phi, Z, \Omega)||p(X, \Phi, Z, \Omega)) &= \int_{X,\Phi,Z,\Omega} \log \frac{q(X, \Phi, Z, \Omega)}{p(X, \Phi, Z, \Omega)} q(X, \Phi, Z, \Omega) dX d\Phi dZ d\Omega \\
&= \int_{X,\Phi,Z,\Omega} \log \frac{q(X|Z)q(\Phi|Z)q(Z)q(\Omega)}{p(X|Z)p(\Phi|Z)p(Z|\Omega)p(\Omega)} q(X|Z)q(\Phi|Z)q(Z)q(\Omega) dX d\Phi dZ d\Omega \\
&= \mathrm{E}_{q(X|Z)}\mathrm{E}_{q(\Phi|Z)}\mathrm{E}_{q(Z)}\mathrm{E}_{q(\Omega)} \left(\log \frac{q(X|Z)}{p(X|Z)} + \log \frac{q(\Phi|Z)}{p(\Phi|Z)} + \log \frac{q(Z)}{p(Z|\Omega)} + \log \frac{p(\Omega)}{p(\Omega)} \right) \\
&= \mathrm{E}_{q(Z)}\mathrm{E}_{q(\Omega)} \left(\mathrm{KL}(q(X|Z||p(X|Z)) + \mathrm{KL}(q(\Phi|Z||p(\Phi|Z)) + \log \frac{q(Z)}{p(Z|\Omega)} + \log \frac{q(\Omega)}{p(\Omega)} \right) \\
&= \underbrace{\mathrm{E}_{q(Z)}\mathrm{KL}(q(X|Z||p(X|Z))}_{\mathcal{L}_2} + \underbrace{\mathrm{E}_{q(Z)}\mathrm{KL}(q(\Phi|Z||p(\Phi|Z))}_{\mathcal{L}_3} \\
&+ \underbrace{\mathrm{E}_{q(\Omega)}\mathrm{KL}(q(Z||p(Z|\Omega))}_{\mathcal{L}_4} + \underbrace{\mathrm{KL}(q(\Omega||p(\Omega))}_{\mathcal{L}_5}
\end{aligned} \tag{S8}$$

with each component derived as:

$$\begin{aligned}
\mathrm{E}_{q(Z)}\left(\mathrm{KL}(q(\Phi|Z)||p(\Phi|Z)\right) &= \mathrm{E}_{q(Z)q(\Phi|Z)} \log \left(\prod_{i=1}^{N} \prod_{k=1}^{K} \left(\frac{\phi_{ik}^{\hat{\alpha}_{ik}-1} \exp^{-\hat{\beta}_{ik}\phi_{ik}} \hat{\beta}_{ik}^{\hat{\alpha}_{ik}}}{\Gamma(\hat{\alpha}_{ik})} \right)^{z_{ik}} \left(\frac{\Gamma(\alpha)}{\phi_{ik}^{\alpha-1} \exp^{-\beta\phi_{ik}} \beta^{\alpha}} \right)^{z_{ik}} \right) \\
&= \int_{\Phi,Z} \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \log \left(\frac{\phi_{ik}^{\hat{\alpha}_{ik}-1} \exp^{-\hat{\beta}_{ik}\phi_{ik}} \hat{\beta}_{ik}^{\hat{\alpha}_{ik}}}{\Gamma(\hat{\alpha}_{ik})} \frac{\Gamma(\alpha)}{\phi_{ik}^{\alpha-1} \exp^{-\beta\phi_{ik}} \beta^{\alpha}} \right) q(\Phi|Z)q(Z) d\Phi dZ \\
&= \int_{Z} \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left(\log \frac{\Gamma(\alpha)}{\Gamma(\hat{\alpha}_{ik})} + (\hat{\alpha}_{ik} - \alpha)(\psi(\hat{\alpha}_{ik}) - \log \hat{\beta}_{ik}) + (\beta - \hat{\beta}_{ik}) \frac{\hat{\alpha}_{ik}}{\hat{\beta}_{ik}} + \hat{\alpha}_{ik} \log \hat{\alpha}_{ik} - \alpha \log \beta \right) q(Z) dZ \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} \pi_{ik} \left(\log \frac{\Gamma(\alpha)}{\Gamma(\hat{\alpha}_{ik})} + (\hat{\alpha}_{ik} - \alpha)\psi(\hat{\alpha}_{ik}) + \alpha \log \frac{\hat{\beta}_{ik}}{\beta} + (\beta - \hat{\beta}_{ik}) \frac{\hat{\alpha}_{ik}}{\hat{\beta}_{ik}} \right)
\end{aligned} \tag{S9}$$

where $\Gamma(\cdot)$ and $\psi(\cdot)$ denotes gamma function and digamma function, respectively.

$$
\begin{aligned}
\mathrm{E}_{q(Z)}\left(\mathrm{KL}(q(X|Z)||p(X|Z)\right) &= \int_{X,Z} \log \frac{\prod_{i=1}^N \prod_{k=1}^K \left(\mathcal{N}(\mu_{ik}, \sigma_{ik}^2)\right)^{z_{ik}}}{\prod_{i=1}^N \prod_{k=1}^K p(x_{ik})^{z_{ik}}} q(X|Z)q(Z)dXdZ \\
&= \int_Z \left(\sum_{k=1}^K \sum_{i=1}^N -\frac{1}{2} z_{ik} (\log \sigma_{ik}^2 + \log 2\pi) - \sum_{k=1}^K \sum_{i=1}^N z_{ik} \mathrm{E}_{q(x_{ik})} \log p(x_{ik}) \right) q(Z)dZ \\
&= \sum_{k=1}^K \sum_{i=1}^N -\frac{1}{2} \pi_{ik} \log \sigma_{ik}^2 - \sum_{k=1}^K \sum_{i=1}^N \pi_{ik} \mathrm{E}_{q(x_{ik})} \log p(x_{ik}) + C
\end{aligned} \tag{S10}
$$

where $C = \sum_{k=1}^K \sum_{i=1}^N -\frac{1}{2} z_{ik} \log 2\pi$.

$$
\begin{aligned}
\mathrm{E}_{q(\Omega)}\left(\mathrm{KL}(q(Z)||p(Z|\Omega)\right) &= \int_{Z,\pi} \log \frac{\prod_{i=1}^N \prod_{k=1}^K \pi_{ik}^{z_{ik}}}{\prod_{i=1}^N \prod_{k=1}^K \omega_{ik}^{z_{ik}}} q(Z)dZq(\Omega)d\Omega \\
&= \sum_{i=1}^N \sum_{k=1}^K \pi_{ik} \left(\log \pi_{ik} - \psi(\hat{d}_{ik}) + \psi(\sum_{k=1}^K \hat{d}_{ik}) \right)
\end{aligned} \tag{S11}
$$

$$
\begin{aligned}
\mathrm{KL}(q(\Omega)||p(\Omega)) &= \int_\Omega \log \frac{\prod_{i=1}^N Z_{\text{Dir}}(\hat{\boldsymbol{d}}_i) \prod_{k=1}^K \omega_{ik}^{\hat{d}_{ik}-1}}{\prod_{i=1}^N Z_{\text{Dir}}(\boldsymbol{d}) \prod_{k=1}^K \omega_{ik}^{d_k - 1}} q(\Omega)d\Omega \\
&= \sum_{i=1}^N \log \frac{Z_{\text{Dir}}(\hat{\boldsymbol{d}}_i)}{Z_{\text{Dir}}(\boldsymbol{d})} + \sum_{i=1}^N \sum_{k=1}^K (\hat{d}_{ik} - d_k)(\psi(\hat{d}_{ik}) - \psi(\sum_{k=1}^K \hat{d}_{ik}))
\end{aligned} \tag{S12}
$$

where $Z_{\text{Dir}}(\cdot)$ is the normalizing parameter of Dirichlet distribution.

Finally, $\mathcal{L}_1 = -\mathrm{E}_{q(X,\Phi,Z,\Omega)}(\log p(\boldsymbol{y}|X,\Phi,Z,\Omega))$ is derived as

$$
\begin{aligned}
-\mathrm{E}_{q(X,\Phi,Z,\Omega)}(\log p(\boldsymbol{y}|X,\Phi,Z,\Omega)) &= -\int_{X,\Phi,Z} \log \prod_{i=1}^N \prod_{k=1}^K \mathcal{N}(x_{ik}, \phi_{ik}^{-1})^{z_{ik}} q(X|Z)q(\Phi|Z)q(Z)dXd\Phi dZ \\
&= -\int_{X,\Phi,Z} \sum_{i=1}^N \sum_{k=1}^K z_{ik} \left(-\frac{(x_{ik} - y_i)^2 \phi_{ik}}{2} - \frac{1}{2} \log 2\pi + \frac{\log \phi_{ik}}{2} \right) q(X|Z)q(\Phi|Z)q(Z)dXd\Phi dZ \\
&= -\int_{\Phi,Z} \sum_{i=1}^N \sum_{k=1}^K z_{ik} \left(-\frac{\left((\mu_{ik} - y_i)^2 + \sigma_{ik}^2\right)\phi_{ik}}{2} - \frac{1}{2} \log 2\pi + \frac{\log \phi_{ik}}{2} \right) q(\Phi|Z)q(Z)d\Phi dZ \\
&= -\int_Z \sum_{i=1}^N \sum_{k=1}^K z_{ik} \left(-\frac{\left((\mu_{ik} - y_i)^2 + \sigma_{ik}^2\right)\hat{\alpha}_{ik}}{2\hat{\beta}_{ik}} - \frac{1}{2} \log 2\pi + \frac{\psi(\hat{\alpha}_{ik}) - \log \hat{\beta}_{ik}}{2} \right) q(Z)dZ \\
&= -\sum_{i=1}^N \sum_{k=1}^K \pi_{ik} \left(-\frac{\left((\mu_{ik} - y_i)^2 + \sigma_{ik}^2\right)\hat{\alpha}_{ik}}{2\hat{\beta}_{ik}} + \frac{\psi(\hat{\alpha}_{ik}) - \log \hat{\beta}_{ik}}{2} \right) - C
\end{aligned} \tag{S13}
$$

Note that $-C$ in $\mathcal{L}_1$ and C in $\mathcal{L}_2$ cancel out, so we simply ignore them in the main paper and the subsequent optimization.

S2. More details and results of ScoreDVI

S2.1. Network architectures and optimization

The four networks for representing the parameters of variational posteriors, i.e., X-net, Φ-net, Z-net and Ω-net, are presented in detail in Figure S1. In practice, there are some restrictions on these variational posterior parameters. That is,

$$
\sigma_{ik}^2 > 0, \hat{\alpha}_{ik} > 0, \hat{\beta}_{ik} > 0, \pi_{ik} \geq 0, \sum_{k=1}^K \pi_{ik} = 1, \hat{d}_{ik} > 0, i = \{1, \cdots N\}, k = \{1, \cdots K\} \tag{1}
$$

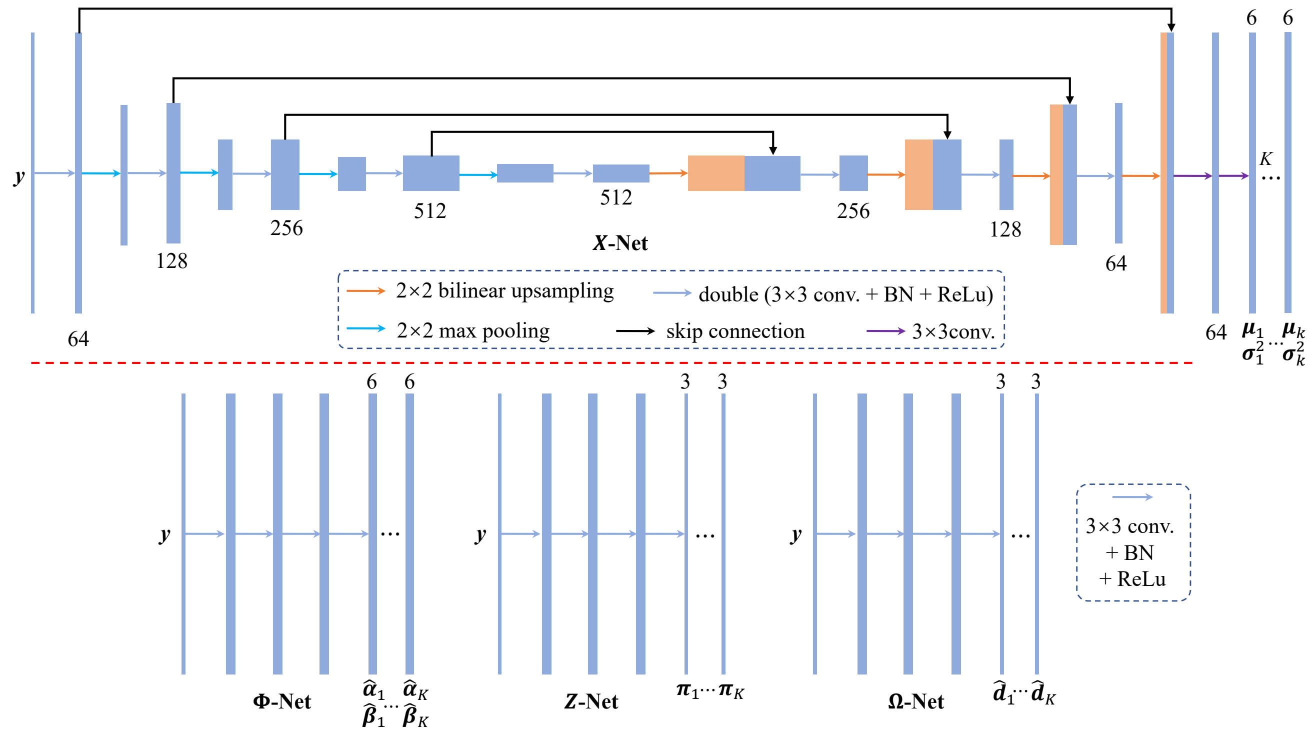

Figure S1: Architectures of four CNNs, i.e., X-net, Φ-net, Z-net and Ω-net. X-net bases on Unet, while the other three nets build upon the DnCNN with 5 convolution layers.

To obtain Θ, we output the logarithms of σ_{ik}^2, $\hat{\alpha}_{ik}$, $\hat{\beta}_{ik}$, and $\hat{d}_{ik}$ using the four networks. We then obtain their actual values by applying exponential operations. As for π_{ik}, we apply the softmax operation channel-wise to the output of Z-net. This ensures that $\pi_{ik} \geq 0$ and that $\sum_{k=1}^{K} \pi_{ik} = 1$.

Note that, we do not explicitly compute $\pi_{ik} \mathrm{E}_{q(x_{ik})} \log p(x_{ik})$ in $\mathcal{L}_2$ but derive its derivative with respect to μ_{ik} and σ_{ik}^2 using score priors as discussed in the main paper. Note that no gradient of π_{ik} is returned from $\pi_{ik} \mathrm{E}_{q(x_{ik})} \log p(x_{ik})$. Hence, we detach π_{ik} from this term and replace it with π_{ik}.detach $\mathrm{E}_{q(x_{ik})} \log p(x_{ik})$ during optimization.

S2.2. Denoising process and uncertainty quantification

The denoising process of our method is illustrated in Figure S2. Importantly, our method provides a natural way to estimate the uncertainty of the denoised image, which is particularly relevant in safety-critical scenarios. Specifically, our method estimates the variance of the posterior image distributions, denoted by $\bar{\boldsymbol{\sigma}}^2$. As shown in Figure S2, $\bar{\boldsymbol{\sigma}}^2(t = 400)$ reflects the degree of uncertainty in the final restored image, especially the edges around the fine details.

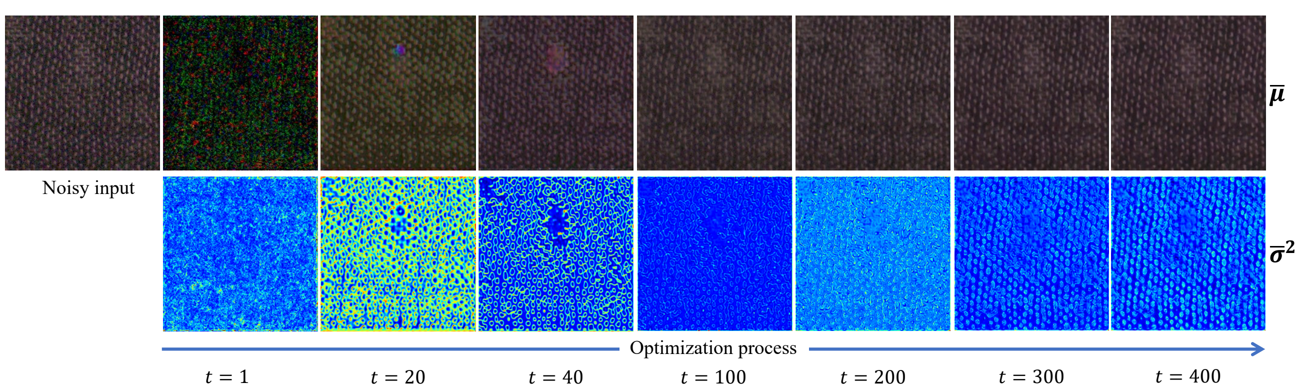

Figure S2: The denoising process of our ScoreDVI for real-world noise removal.

S3. Image-wise fusion strategy

We consider the likelihood function as

$$p(\boldsymbol{y}|X, \Phi, \boldsymbol{\omega}) = \sum_{k=1}^{K} \omega_k \mathcal{N}(\boldsymbol{x}_k, \text{diag}(\boldsymbol{\phi}_k)^{-1}), \sum_{k=1}^{K} \omega_k = 1 \tag{S14}$$

where $\boldsymbol{\omega}$ is the mixing vector that combines K Gaussian components. Eq. (S14) can be equivalently expressed as

$$p(\boldsymbol{y}|X, \Phi, \boldsymbol{z}) = \prod_{k=1}^{K} \mathcal{N}(\boldsymbol{x}_k, \text{diag}(\boldsymbol{\phi}_k^{-1}))^{z_k}, p(\boldsymbol{z}|\boldsymbol{\omega}) = \prod_{k=1}^{K} \omega_k^{z_k}, \sum_{k=1}^{K} z_k = 1 \tag{S15}$$

which allows modeling the image prior and the variational image posterior conditioned on $\boldsymbol{z}$ as

$$p(X|\boldsymbol{z}) = \prod_{k=1}^{K} p(\boldsymbol{x})^{z_k}, q(X|\boldsymbol{z}) = \prod_{k=1}^{K} \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k^2)^{z_k}, q(\boldsymbol{z}) = \prod_{k=1}^{K} \pi_k \tag{S16}$$

Therefore, the mean of variational image posteriors is $\bar{\boldsymbol{\mu}} = \sum_{k=1}^{K} \pi_k \boldsymbol{\mu}_k$, which is the image-wise fusion of different $\boldsymbol{\mu}_k$.

S4. More visual comparisons

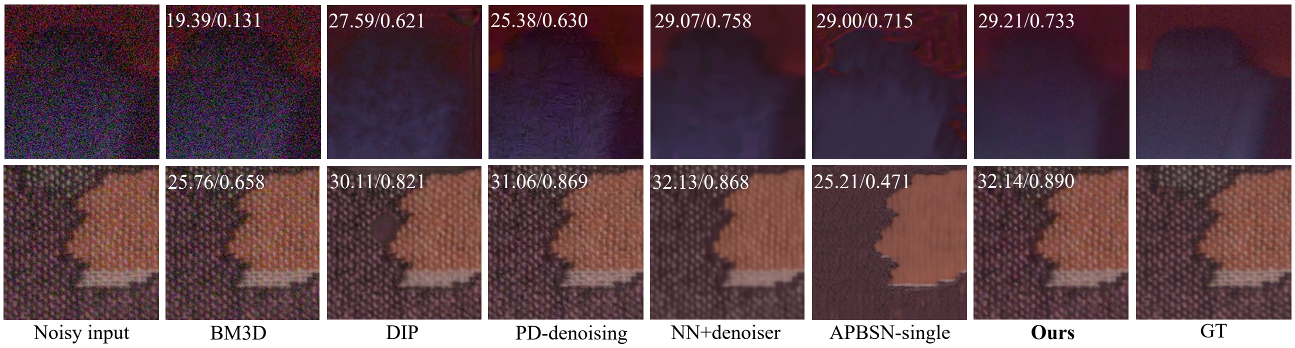

Figure S3: More visual comparisons of our method against other single image-based denoising methods in SIDD validation.

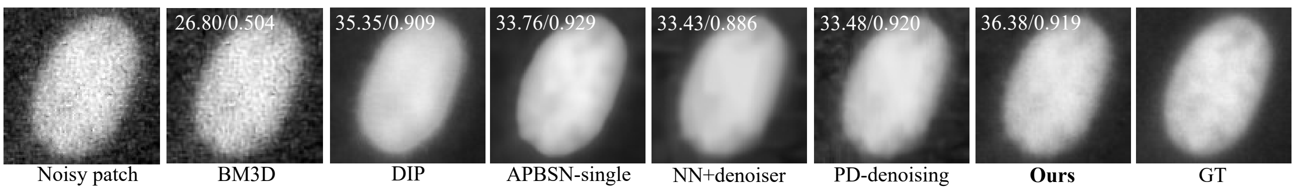

Figure S4: More visual comparisons of our method against other single image-based denoising methods in FMDD dataset. PSNR/SSIM are evaluated on the whole image. The noisy patch is fromThe noisy patch is from WideField_BPAE_B_1.

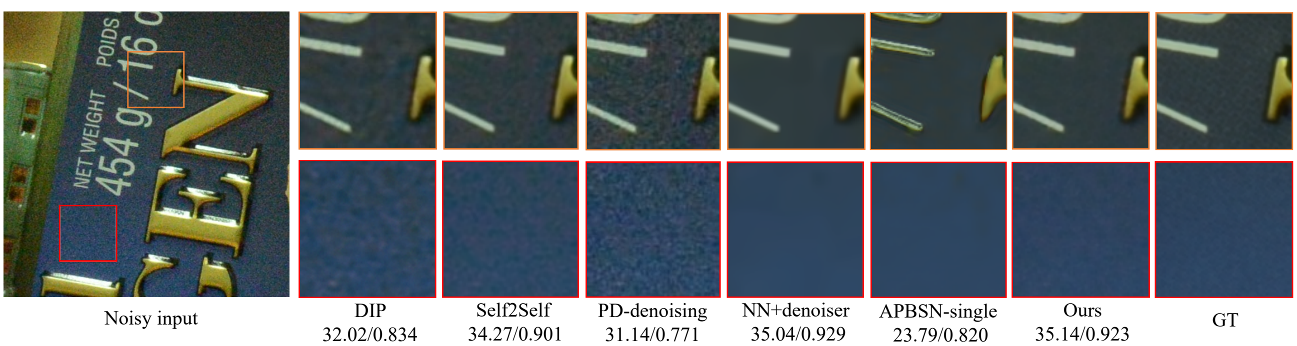

Figure S5: Visual comparisons of our method against other single image-based denoising methods in CC dataset.

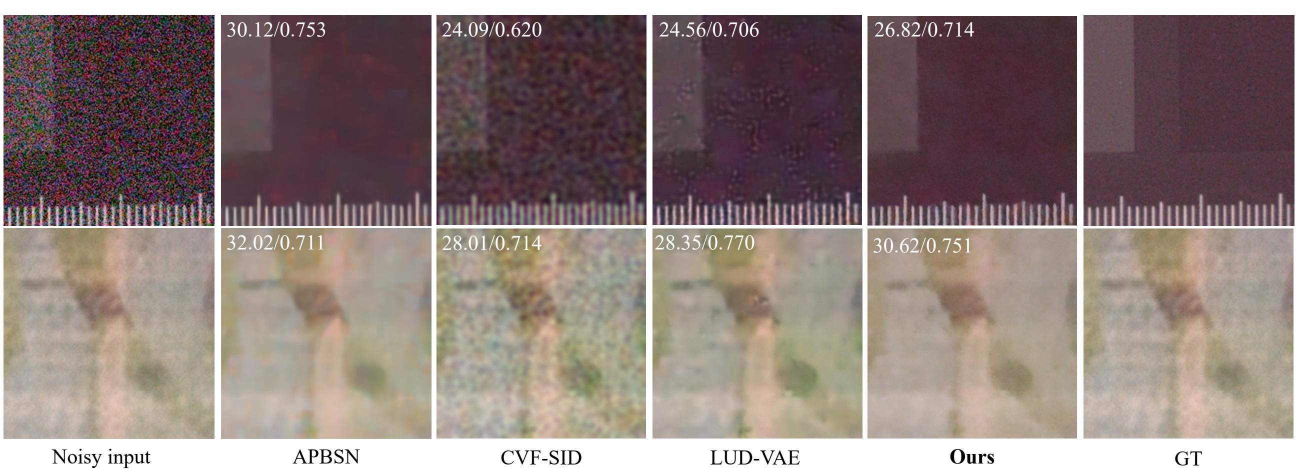

Figure S6: Visual comparisons of our method and other dataset-based denoising methods in SIDD validation dataset.

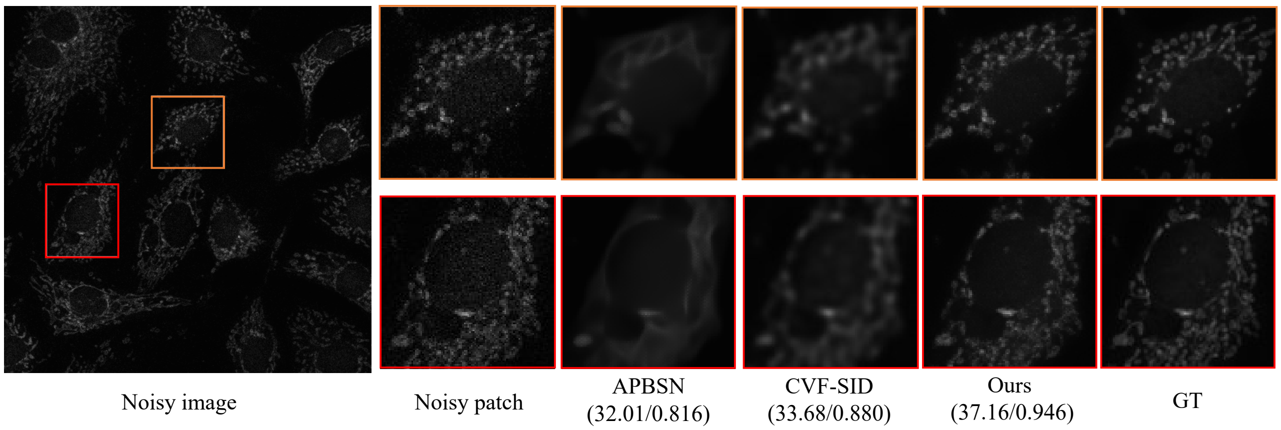

Figure S7: Visual comparisons of our method and other dataset-based denoising methods in FMDD dataset.

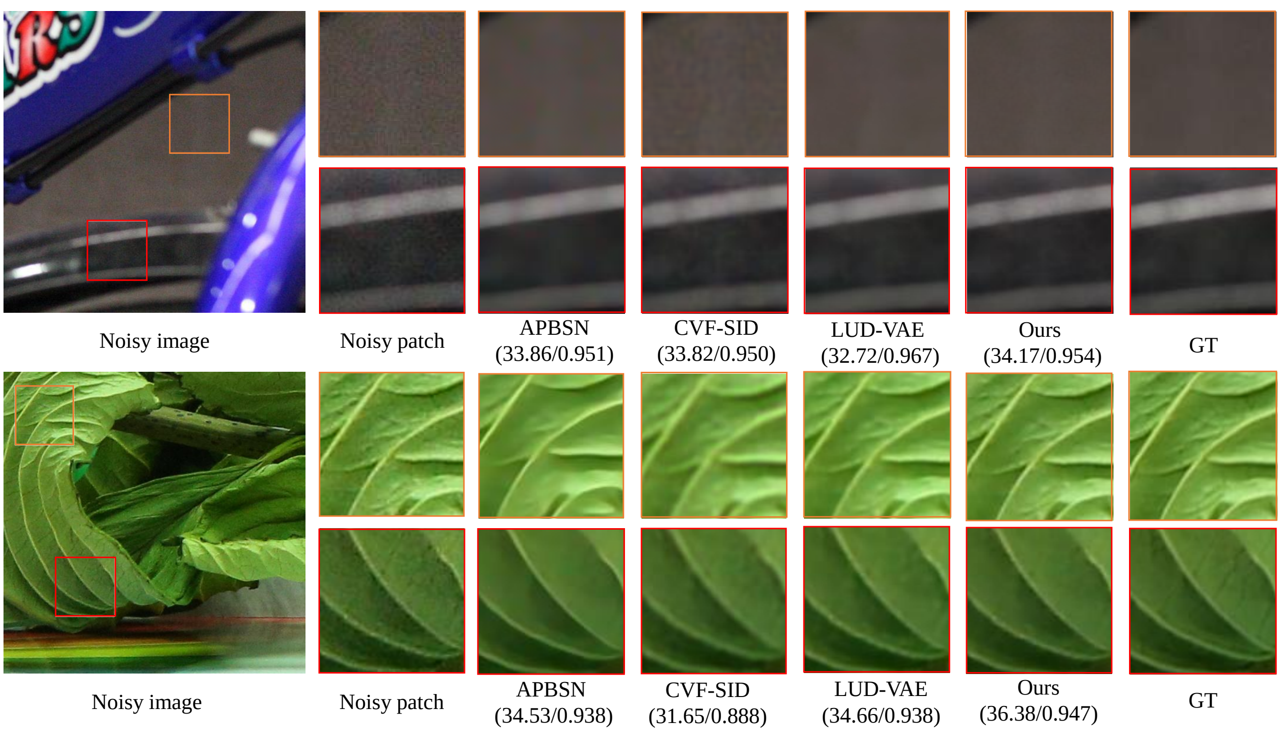

Figure S8: Visual comparisons of our method and other dataset-based denoising methods in PolyU (first row) and CC (second row) dataset.