
Supplementary Material of "Variational Denoising Network: Toward Blind Noise Modeling and Removal"

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 In this supplementary material, we provide more calculation details on the deduc-
2 tion of the variational lower bound, and demonstrate more experimental results in
3 blind image denoising.

4 1 Calculation Details on the Variational Lower Bound

5 1.1 Model Formation

6 Let's denote $\mathbf{y} \in \mathbb{R}^d$ as the observed noisy image and $\mathbf{z} \in \mathbb{R}^d$ the latent clean image. Different
7 from most of the traditional methods, we assumed the noise is distributed as non-i.i.d. Gaussian
8 distribution, i.e.,

$$y_i \sim \mathcal{N}(y_i | z_i, \sigma_i^2), \quad i = 1, 2, \dots, d, \quad (1)$$

9 where $\mathcal{N}(\cdot | \mu, \sigma^2)$ represents the Gaussian distribution with mean μ variance σ^2 .

10 The simulated clean image \mathbf{x} evidently provides a strong prior to the latent variable \mathbf{z} . Accordingly
11 we impose the following conjugate Gaussian prior on \mathbf{z} :

$$z_i \sim \mathcal{N}(z_i | x_i, \varepsilon_0^2), \quad i = 1, 2, \dots, d, \quad (2)$$

12 where ε_0 is a hyper-parameter and can be easily set as a small value.

13 Besides, for $\boldsymbol{\sigma}^2 = \{\sigma_1^2, \sigma_2^2, \dots, \sigma_d^2\}$, we also introduce a rational conjugate prior as follows:

$$\sigma_i^2 \sim \text{IG}\left(\sigma_i^2 \mid \frac{p^2}{2} - 1, \frac{p^2 \xi_i}{2}\right), \quad i = 1, 2, \dots, d, \quad (3)$$

14 where $\text{IG}(\cdot | \alpha, \beta)$ is the inverse gamma distribution with parameter α and β , $\xi = \mathcal{G}((\hat{\mathbf{y}} - \hat{\mathbf{x}})^2; p)$
15 represents the filtering output of the variance map $(\hat{\mathbf{y}} - \hat{\mathbf{x}})^2$ by a Gaussian filter with $p \times p$ window,
16 $\hat{\mathbf{y}}, \hat{\mathbf{x}} \in \mathbb{R}^{h \times w}$ are the matrix (image) forms of $\mathbf{y}, \mathbf{x} \in \mathbb{R}^d$, respectively. Note that the mode of above
17 IG distribution is ξ_i , which is a rational approximate evaluation of σ_i^2 under $p \times p$ window.

18 Combining Eqs (1)-(3), a full Bayesian model for the problem can be obtained. The goal then turns
19 to construct a variational strategy to infer the posterior of latent variables \mathbf{z} and $\boldsymbol{\sigma}^2$ from noisy image
20 \mathbf{y} , i.e., $p(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})$.

21 1.2 Variational Lower Bound

22 Instead of calculating the posterior $p(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})$ directly, we introduced another distribution $q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})$
23 to approximate it. Based on such approximate distribution, we can decompose the marginal likelihood

24 of \mathbf{y} as follows:

$$\begin{aligned}
\log p(\mathbf{y}; \mathbf{z}, \boldsymbol{\sigma}^2) &= \int q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y}) \log p(\mathbf{y} | \mathbf{z}, \boldsymbol{\sigma}^2) d\mathbf{z} d\boldsymbol{\sigma}^2 \\
&= \int q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y}) \log \left[\frac{p(\mathbf{y} | \mathbf{z}, \boldsymbol{\sigma}^2) p(\mathbf{z}) p(\boldsymbol{\sigma}^2)}{p(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})} \right] d\mathbf{z} d\boldsymbol{\sigma}^2 \\
&= \int q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y}) \log \left[\frac{p(\mathbf{y} | \mathbf{z}, \boldsymbol{\sigma}^2) p(\mathbf{z}) p(\boldsymbol{\sigma}^2)}{q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})} + \frac{q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})}{p(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})} \right] d\mathbf{z} d\boldsymbol{\sigma}^2 \\
&= \int q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y}) \log \left[\frac{p(\mathbf{y} | \mathbf{z}, \boldsymbol{\sigma}^2) p(\mathbf{z}) p(\boldsymbol{\sigma}^2)}{q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})} \right] d\mathbf{z} d\boldsymbol{\sigma}^2 \\
&\quad + \int q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y}) \log \left[\frac{q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})}{p(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})} \right] d\mathbf{z} d\boldsymbol{\sigma}^2 \\
&= E_{q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})} [\log p(\mathbf{y} | \mathbf{z}, \boldsymbol{\sigma}^2) p(\mathbf{z}) p(\boldsymbol{\sigma}^2) - \log q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})] \\
&\quad + D_{KL}(q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y}) || p(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})). \quad (4)
\end{aligned}$$

25 The secode term is a KL divergence of the approximation $q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})$ to the true posterior $p(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})$,
26 which is non-negative, and thus the first term constitutes a *variational lower bound* on the marginal
27 likelihood of $p(\mathbf{y} | \mathbf{z}, \boldsymbol{\sigma}^2)$, i.e.,

$$\begin{aligned}
\log p(\mathbf{y}; \mathbf{z}, \boldsymbol{\sigma}^2) &\geq \mathcal{L}(\mathbf{z}, \boldsymbol{\sigma}^2; \mathbf{y}) \\
&= E_{q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})} [\log p(\mathbf{y} | \mathbf{z}, \boldsymbol{\sigma}^2) p(\mathbf{z}) p(\boldsymbol{\sigma}^2) - \log q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})]. \quad (5)
\end{aligned}$$

28 Similar to the traditional mean-field variation methods, we assumed the independence between
29 variable \mathbf{z} and $\boldsymbol{\sigma}^2$, i.e.,

$$q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y}) = q(\mathbf{z} | \mathbf{y}) q(\boldsymbol{\sigma}^2 | \mathbf{y}). \quad (6)$$

30 Based on the conjugate priors in Eq. 2 and 3, it is natural to formulate variational posterior forms of
31 \mathbf{z} and $\boldsymbol{\sigma}^2$ as follows:

$$q(\mathbf{z} | \mathbf{y}) = \prod_i^d \mathcal{N}(z_i | \mu_i(\mathbf{y}; W_D), m_i^2(\mathbf{y}; W_D)), \quad q(\boldsymbol{\sigma}^2 | \mathbf{y}) = \prod_i^d \text{IG}(\sigma_i^2 | \alpha_i(\mathbf{y}; W_S), \beta_i(\mathbf{y}; W_S)), \quad (7)$$

32 where $\mu_i(\mathbf{y}; W_D)$ and $m_i^2(\mathbf{y}; W_D)$ are designed as the prediction functions for getting posterior
33 parameters of latent variable \mathbf{z} directly from \mathbf{y} . The function is represented as a network, called
34 denoising network or *D-Net*, with parameters W_D . Similarly, $\alpha_i(\mathbf{y}; W_S)$ and $\beta_i(\mathbf{y}; W_S)$ denote
35 the prediction functions for evaluating posterior parameters of $\boldsymbol{\sigma}^2$ from \mathbf{y} , where W_S represents the
36 parameters of a network, called Sigma network or *S-Net*, for predicting them. Our aim is then to
37 optimize these two network parameters W_D and W_S so as to get the explicit functions for predicting
38 clean image variable \mathbf{z} as well as noise knowledge $\boldsymbol{\sigma}^2$ from any test noisy image \mathbf{y} . A rational
39 objective function with respect to W_D and W_S is thus necessary for using gradient decent strategies
40 to train both networks.

41 For notation convenience, we simply write $\mu_i(\mathbf{y}; W_D), m_i^2(\mathbf{y}; W_D), \alpha_i(\mathbf{y}; W_S), \beta_i(\mathbf{y}; W_S)$ as μ_i ,
42 m_i^2, α_i, β_i in the following calculations.

43 Combining Eqs (5), (6) and Eq (7), the lower bound can be rewritten as:

$$\mathcal{L}(\mathbf{z}, \boldsymbol{\sigma}^2; \mathbf{y}) = E_{q(\mathbf{z}, \boldsymbol{\sigma}^2 | \mathbf{y})} [\log p(\mathbf{y} | \mathbf{z}, \boldsymbol{\sigma}^2)] - D_{KL}(q(\mathbf{z} | \mathbf{y}) || p(\mathbf{z})) - D_{KL}(q(\boldsymbol{\sigma}^2 | \mathbf{y}) || p(\boldsymbol{\sigma}^2)), \quad (8)$$

44 Next we calculated the three terms in Eq (8) one by one as follows:

$$\begin{aligned}
E_{q(\mathbf{z}, \sigma^2 | \mathbf{y})} [\log p(\mathbf{y} | \mathbf{z}, \sigma^2)] &= \int q(\mathbf{z}, \sigma^2 | \mathbf{y}) \log p(\mathbf{y} | \mathbf{z}, \sigma^2) d\mathbf{z} d\sigma^2 \\
&= \sum_i^n \int q(z_i, \sigma_i^2 | \mathbf{y}) \log p(y_i | z_i, \sigma_i^2) dz_i d\sigma_i^2 \\
&= \sum_i^n \int q(z_i | \mathbf{y}) q(\sigma_i^2 | \mathbf{y}) \left\{ -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma_i^2 - \frac{(y_i - z_i)^2}{2\sigma_i^2} \right\} dz_i d\sigma_i^2 \\
&= \sum_i \left\{ -\frac{1}{2} \log 2\pi - \frac{1}{2} \int q(\sigma_i^2 | \mathbf{y}) \log \sigma_i^2 d\sigma_i^2 \int q(z_i | \mathbf{y}) dz_i \right. \\
&\quad \left. - \frac{1}{2} \int q(z_i | \mathbf{y}) (y_i - z_i)^2 dz_i \int q(\sigma_i^2 | \mathbf{y}) \frac{1}{\sigma_i^2} d\sigma_i^2 \right\} \\
&= \sum_i^n \left\{ -\frac{1}{2} \log 2\pi - \frac{1}{2} E[\log \sigma_i^2] - \frac{1}{2} E[(y_i - z_i)^2] E\left[\frac{1}{\sigma_i^2}\right] \right\} \\
&= \sum_i^n \left\{ -\frac{1}{2} \log 2\pi - \frac{1}{2} (\log \beta_i - \psi(\alpha_i)) - \frac{\alpha_i}{2\beta_i} [(y_i - \mu_i)^2 + m_i^2] \right\},
\end{aligned} \tag{9}$$

45

$$\begin{aligned}
D_{KL}(q(\mathbf{z} | \mathbf{y}) || p(\mathbf{z})) &= \sum_i^n D_{KL}(\mathcal{N}(z_i | \mu_i, m_i^2) || p(z_i | x_i, \varepsilon_0^2)) \\
&= \sum_i^n \left\{ \frac{(\mu_i - x_i)^2}{2\varepsilon_0^2} + \frac{1}{2} \left[\frac{m_i^2}{\varepsilon_0^2} - \log \frac{m_i^2}{\varepsilon_0^2} - 1 \right] \right\},
\end{aligned} \tag{10}$$

46

$$\begin{aligned}
D_{KL}(q(\sigma^2 | \mathbf{y}) || p(\sigma^2)) &= \sum_i^n D_{KL}\left(\text{IG}(\sigma_i^2 | \alpha_i, \beta_i) || \text{IG}\left(\sigma_i^2 | \frac{p^2}{2} - 1, \frac{p^2 \xi_i}{2}\right)\right) \\
&= \sum_i^n \left\{ \left(\alpha_i - \frac{p^2}{2} + 1 \right) \psi(\alpha_i) + \left[\log \Gamma\left(\frac{p^2}{2} - 1\right) - \log \Gamma(\alpha_i) \right] \right. \\
&\quad \left. + \left(\frac{p^2}{2} - 1 \right) \left(\log \beta_i - \log \frac{p^2 \xi_i}{2} \right) + \alpha_i \left(\frac{p^2 \xi_i}{2\beta_i} - 1 \right) \right\},
\end{aligned} \tag{11}$$

47 Where $\psi(\cdot)$ denotes the digamma function, $E[\cdot]$ represents expectation with some stochastic variables
48 that had been neglected for notation clarity.

49 We can then easily get the expected objective function (i.e., a negative lower bound of the marginal
50 likelihood on entire training set) for optimizing the network parameters of D-Net and S-Net as follows:

51

$$\min_{W_D, W_S} - \sum_{j=1}^n \mathcal{L}(\mathbf{z}_j, \sigma_j^2; \mathbf{y}_j). \tag{12}$$

52 2 More Experimental Results

53 2.1 Experiments on Synthetic Non-I.I.D. Gaussian Noise

54 In this supplementary material, we displayed more denoising results of different methods on the
55 testing dataset in Fig. 1-6.

56 2.2 Experiments on Real-World Noise

57 In Fig. 7, we show more denoising results of different methods on the SIDD validation dataset.

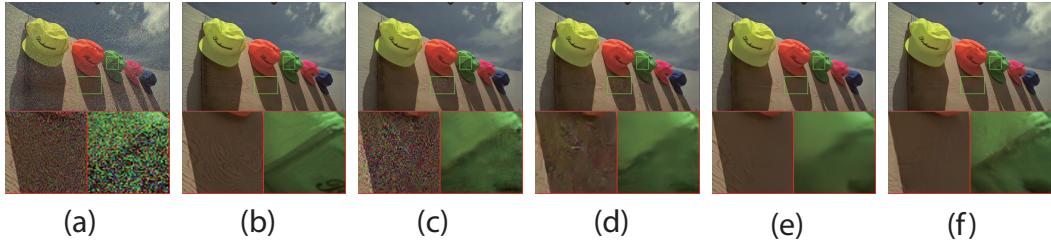


Figure 1: Image denoising results of different methods on the testing data in Case 1. From left to right: (a) Noisy Image, (b) Groundtruth, (c) CBM3D, (d) DnCNN-B, (e) FFDNet, (f) VDN

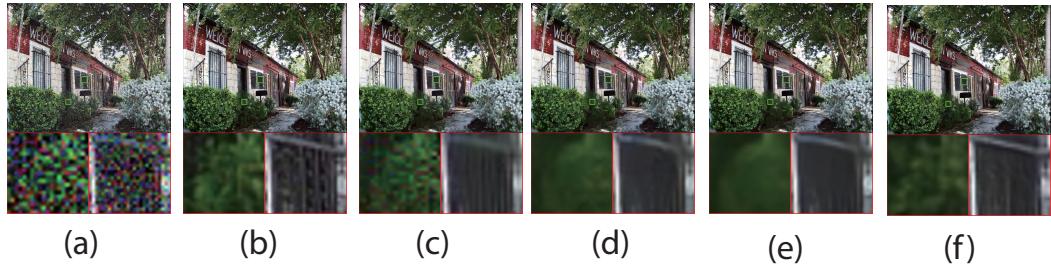


Figure 2: Image denoising results of different methods on the testing data in Case 1. From left to right: (a) Noisy Image, (b) Groundtruth, (c) CBM3D, (d) DnCNN-B, (e) FFDNet, (f) VDN

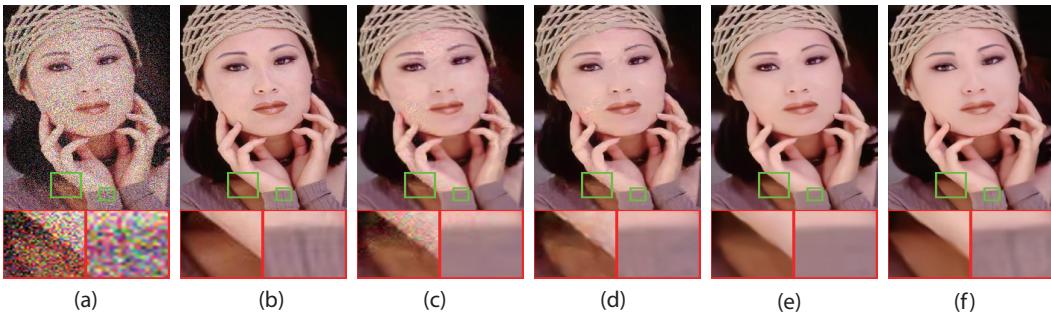


Figure 3: Image denoising results of different methods on the testing data in Case 2. From left to right: (a) Noisy Image, (b) Groundtruth, (c) CBM3D, (d) DnCNN-B, (e) FFDNet, (f) VDN

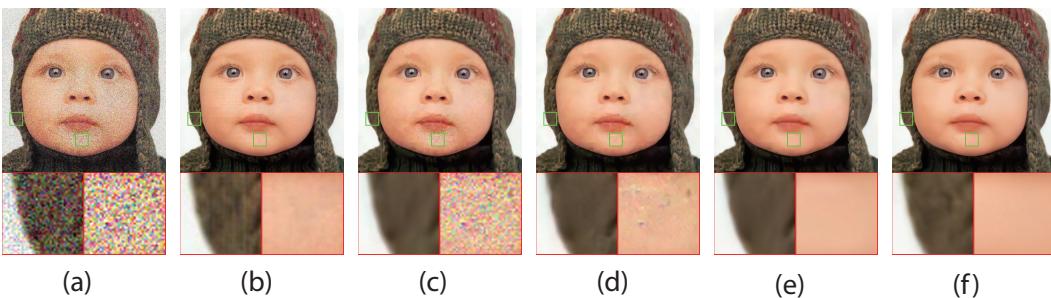


Figure 4: Image denoising results of different methods on the testing data in Case 2. From left to right: (a) Noisy Image, (b) Groundtruth, (c) CBM3D, (d) DnCNN-B, (e) FFDNet, (f) VDN

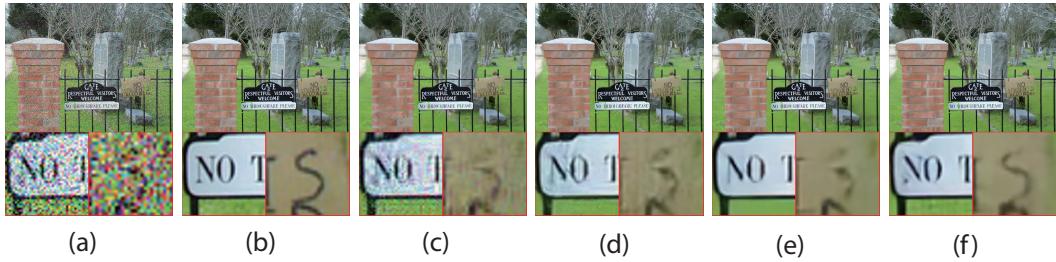


Figure 5: Image denoising results of different methods on the testing data in Case 3. From left to right: (a) Noisy Image, (b) Groundtruth, (c) CBM3D, (d) DnCNN-B, (e) FFDNet, (f) VDN

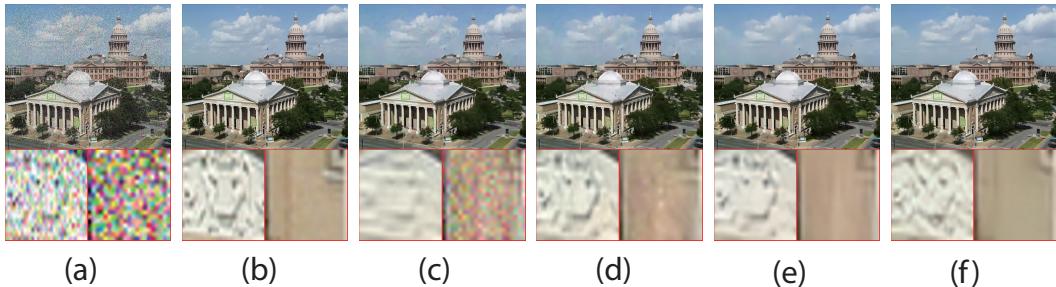


Figure 6: Image denoising results of different methods on the testing data in Case 3. From left to right: (a) Noisy Image, (b) Groundtruth, (c) CBM3D, (d) DnCNN-B, (e) FFDNet, (f) VDN

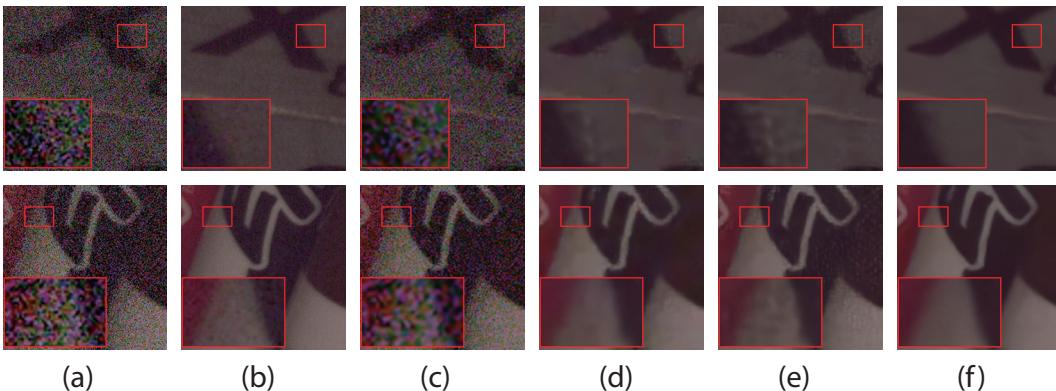


Figure 7: Image denoising results of different methods on the SIDD validation set. From left to right: (a) Noisy image, (b) Simulated "clean" image, (c) WNNM, (d) DnCNN-B, (e) CBDNet, (f) VDN