Supplementary Materials of "DDS2M: Self-Supervised Denoising Diffusion Spatio-Spectral Model for Hyperspectral Image Restoration"

Yuchun Miao[1] Lefei Zhang[1,3*] Liangpei Zhang[2] Dacheng Tao[4]

[1]National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University

[2]State Key Lab. of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University

[3]Hubei Luojia Laboratory, [4]Sydney AI Centre, School of Computer Science, The University of Sydney

{miaoyuchun, zhanglefei, zlp62}@whu.edu.cn, dacheng.tao@gmail.com

1. Diffusion Process for DDS2M

Given a degraded HSI $\mathbf{y}$, the diffusion model defined in DDS2M is Markov chain $\mathbf{x}_T \to \mathbf{x}_{T-1} \to \ldots \to \mathbf{x}_1 \to \mathbf{x}_0$ conditioned on $\mathbf{y}$ [4], where $\mathbf{x_0}$ is the underlying high-quality HSI (final diffusion output). In order to perform inference, the following variational distribution is considered:

$$q\left(\mathbf{x}_{1:T}|\mathbf{x}_0,\mathbf{y}\right) = q\left(\mathbf{x}_T|\mathbf{x}_0,\mathbf{y}\right)\prod_{t=0}^{T-1} q\left(\mathbf{x}_t|\mathbf{x}_{t+1},\mathbf{x}_0,\mathbf{y}\right), \tag{1}$$

where

$$q\left(\overline{\mathbf{x}}_T^{(i)}|\mathbf{x}_0,\mathbf{y}\right) = \begin{cases} \mathcal{N}\left(\overline{\mathbf{y}}^{(i)}, \sigma_T^2 - \frac{\sigma_{\mathbf{y}}^2}{s_i^2}\right) & \text{if } s_i > 0 \\ \mathcal{N}\left(\overline{\mathbf{x}}_0^{(i)}, \sigma_T^2\right) & \text{if } s_i = 0 \end{cases}$$

$$q\left(\overline{\mathbf{x}}_t^{(i)}|\mathbf{x}_{t+1},\mathbf{x}_0,\mathbf{y}\right) = \begin{cases} \mathcal{N}\left(\overline{\mathbf{x}}_0^{(i)} + \sqrt{1-\eta^2}\sigma_t \frac{\overline{\mathbf{x}}_{t+1}^{(i)} - \overline{\mathbf{x}}_0^{(i)}}{\sigma_{t+1}}, \eta^2\sigma_t^2\right) & \text{if } s_i = 0 \\ \mathcal{N}\left(\overline{\mathbf{x}}_0^{(i)} + \sqrt{1-\eta^2}\sigma_t \frac{\overline{\mathbf{y}}^{(i)} - \overline{\mathbf{x}}_0^{(i)}}{\sigma_{\mathbf{y}}/s_i}, \eta^2\sigma_t^2\right) & \text{if } \sigma_t < \frac{\sigma_{\mathbf{y}}}{s_i} \\ \mathcal{N}\left(\left(1-\eta_b\right)\overline{\mathbf{x}}_0^{(i)} + \eta_b\overline{\mathbf{y}}^{(i)}, \sigma_t^2 - \frac{\sigma_{\mathbf{y}}^2}{s_i^2}\eta_b^2\right) & \text{if } \sigma_t \geq \frac{\sigma_{\mathbf{y}}}{s_i} \end{cases} \tag{2}$$

where $\bar{\mathbf{x}}_t{}^{(i)}$ is the i-th index of vector $\bar{\mathbf{x}}_t = \mathbf{V}^\mathsf{T}\mathbf{x}_t$, $\bar{\mathbf{y}}^{(i)}$ is the i-th index of $\bar{\mathbf{y}} = \mathbf{\Sigma}^\dagger\mathbf{U}^\mathsf{T}\mathbf{y}$, σ_t depending on the hyperparameter $\beta_{1:T}$ denotes the variance of diffusion noise in $\mathbf{x}_t$, and η, η_b are the hyperparameters, which control the level of noise injected at each timestep.

It has been proved that the variational distribution defined in Eqn. (1) and (2) has the following marginal distribution equivalent to that in [3, 8]:

$$q\left(\mathbf{x}_t|\mathbf{x}_0\right) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, \left(1-\bar{\alpha}_t\right)\mathbf{I}\right) \tag{3}$$

*Corresponding Author

And the diffusion process (i.e., forward process) can be derived from Bayes' rule:

$$q\left(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{x}_0,\mathbf{y}\right) = \frac{q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)q\left(\mathbf{x}_t|\mathbf{x}_0,\mathbf{y}\right)}{q\left(\mathbf{x}_{t-1}|\mathbf{x}_0,\mathbf{y}\right)} \tag{4}$$

2. Loss Function Derivations

Below is the deviation of our variational inference-based function:

$$\begin{aligned}
&\mathbb{E}_{q(\mathbf{x}_0),q(\mathbf{y}|\mathbf{x}_0)}\left[\log p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_0|\mathbf{y}\right)\right] \\
&\geq \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[\log\frac{p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{0:T}|\mathbf{y}\right)}{q\left(\mathbf{x}_{1:T}|\mathbf{x}_0,\mathbf{y}\right)}\right] \\
&= \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[\log p\left(\mathbf{x}_T|\mathbf{y}\right) + \sum_{t\geq 1}\log\frac{p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right)}{q\left(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{y}\right)}\right] \\
&= \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[\log p\left(\mathbf{x}_T|\mathbf{y}\right) + \sum_{t>1}\log\frac{p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right)}{q\left(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{y}\right)}\right. \\
&\qquad\left. + \log\frac{p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_0|\mathbf{x}_1,\mathbf{y}\right)}{q\left(\mathbf{x}_1|\mathbf{x}_0,\mathbf{y}\right)}\right] \\
&= \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[\log p\left(\mathbf{x}_T\right) + \sum_{t>1}\log\frac{p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right)}{q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)}\right. \\
&\qquad\left. \cdot\frac{q\left(\mathbf{x}_{t-1}|\mathbf{x}_0,\mathbf{y}\right)}{q\left(\mathbf{x}_t|\mathbf{x}_0,\mathbf{y}\right)} + \log\frac{p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_0|\mathbf{x}_1,\mathbf{y}\right)}{q\left(\mathbf{x}_1|\mathbf{x}_0,\mathbf{y}\right)}\right] \\
&= \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[\log\frac{p\left(\mathbf{x}_T|\mathbf{y}\right)}{q\left(\mathbf{x}_T|\mathbf{x}_0,\mathbf{y}\right)}\right. \\
&\qquad\left. + \sum_{t>1}\log\frac{p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right)}{q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)} + \log p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_0|\mathbf{x}_1,\mathbf{y}\right)\right] \\
&= \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[-D_{\mathrm{KL}}\left(q\left(\mathbf{x}_T|\mathbf{x}_0,\mathbf{y}\right)\|p\left(\mathbf{x}_T|\mathbf{y}\right)\right)\right. \\
&\qquad - \sum_{t>1}D_{\mathrm{KL}}\left(q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)\|p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right)\right) \\
&\qquad \left. + \log p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_0|\mathbf{x}_1,\mathbf{y}\right)\right]
\end{aligned} \tag{5}$$

By maximizing the variational lower bound of $\mathbb{E}_{q(\mathbf{x}_0),q(\mathbf{y}|\mathbf{x}_0)}\left[\log p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_0|\mathbf{y}\right)\right]$, we have

$$\underset{\{\boldsymbol{\theta},\boldsymbol{\zeta}\}}{\arg\max}\ \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[\log\frac{p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{0:T}|\mathbf{y}\right)}{q\left(\mathbf{x}_{1:T}|\mathbf{x}_0,\mathbf{y}\right)}\right] \tag{6}$$

$$\begin{aligned}
&= \underset{\{\boldsymbol{\theta},\boldsymbol{\zeta}\}}{\arg\max}\ \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[-D_{\mathrm{KL}}\left(q\left(\mathbf{x}_T|\mathbf{x}_0,\mathbf{y}\right)\|p\left(\mathbf{x}_T|\mathbf{y}\right)\right)\right.\\
&\qquad -\sum_{t>1}D_{\mathrm{KL}}\left(q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)\|p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right)\right)\\
&\qquad \left.+\log p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_0|\mathbf{x}_1,\mathbf{y}\right)\right]\\
&= \underset{\{\boldsymbol{\theta},\boldsymbol{\zeta}\}}{\arg\max}\ \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[\log p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_0|\mathbf{x}_1,\mathbf{y}\right)\right.\\
&\qquad \left.-\sum_{t>1}D_{\mathrm{KL}}\left(q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)\|p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right)\right)\right]\\
&= \underset{\{\boldsymbol{\theta},\boldsymbol{\zeta}\}}{\arg\min}\ \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[-\log p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_0|\mathbf{x}_1,\mathbf{y}\right)\right.\\
&\qquad \left.+\sum_{t>1}D_{\mathrm{KL}}\left(q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)\|p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right)\right)\right]
\end{aligned} \tag{7}$$

For $t>1$:

$$\begin{aligned}
&\underset{\{\boldsymbol{\theta},\boldsymbol{\zeta}\}}{\arg\min}\ \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[\right.\\
&\qquad \left.\sum_{t>1}D_{\mathrm{KL}}\left(q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)\|p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{y}\right)\right)\right]\\
&= \underset{\{\boldsymbol{\theta},\boldsymbol{\zeta}\}}{\arg\min}\ \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[\right.\\
&\qquad \left.\sum_{t>1}D_{\mathrm{KL}}\left(q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0,\mathbf{y}\right)\|q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_{\boldsymbol{\theta},\boldsymbol{\zeta}},\mathbf{y}\right)\right)\right]\\
&= \underset{\{\boldsymbol{\theta},\boldsymbol{\zeta}\}}{\arg\min}\ \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left\|\mathbf{x}_0-\mathbf{x}_{\boldsymbol{\theta},\boldsymbol{\zeta}}\right\|_F^2
\end{aligned} \tag{8}$$

For $t=1$:

$$\begin{aligned}
&\underset{\{\boldsymbol{\theta},\boldsymbol{\zeta}\}}{\arg\min}\ \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left[-\log p_{\boldsymbol{\theta},\boldsymbol{\zeta}}\left(\mathbf{x}_0|\mathbf{x}_1,\mathbf{y}\right)\right]\\
&= \underset{\{\boldsymbol{\theta},\boldsymbol{\zeta}\}}{\arg\min}\ \mathbb{E}_{q(\mathbf{x}_{0:T}),q(\mathbf{y}|\mathbf{x}_0)}\left\|\mathbf{x}_0-\mathbf{x}_{\boldsymbol{\theta},\boldsymbol{\zeta}}\right\|_F^2
\end{aligned} \tag{9}$$

Therefore, the objective in Eqn. (5) can be reduced into a denoising objective, i.e., estimating the underlying high quality HSI $\mathbf{x}_0$ from the noisy version $\mathbf{x}_t$. Inspired by the self-supervised loss functions in [9], our variational inference-based loss function can be designed as follows:

$$\underset{\{\boldsymbol{\theta},\boldsymbol{\zeta}\}}{\arg\min}\ \left\|\mathbf{x}_t-\mathrm{vec}(\sqrt{\bar{\alpha}_t}\sum_{r=1}^{R}\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{z}_r)\circ\mathcal{C}_{\boldsymbol{\zeta}}(\mathbf{w}_r))\right\|_F^2. \tag{10}$$

3. HSI Decomposition Utilized in DDS2M

Under linear mixture model [7], $\mathcal{X}\in\mathbb{R}^{I\times J\times K}$ can be factorized as follows (when the noise is absent):

$$\mathcal{X}=\sum_{r=1}^{R}\mathbf{S}_r\circ\mathbf{c}_r, \tag{11}$$

where $\mathbf{S}_r\in\mathbb{R}^{I\times J}$ and $\mathbf{c}_r\in\mathbb{R}^{K}$ represent the r-th endmember's abundance map and the spectral signature, respectively, and R is the number of endmembers contained in the HSI. This decomposition can also be expressed as

$$\mathcal{X}^{(i,j,k)}=\sum_{r=1}^{R}\mathbf{S}_r^{(i,j)}\mathbf{c}_r^k. \tag{12}$$

Physically, it means that every pixel is a non-negative combination of the spectral signatures of the constituting endmembers in the HSI. An illustration of this decomposition can be found in Figure 1. In addition, this decomposition with a relatively small R can often capture around 98% of the energy of the HSI [1]. Hence, it is a reliable model for HSIs. Indeed, this decomposition has been utilized for a large variety of hyperspectral imaging tasks, e.g., hyperspectral unmixing [11, 2], hyperspectral super-resolution [5], pansharpening [6], and denoising [12], just to name a few.

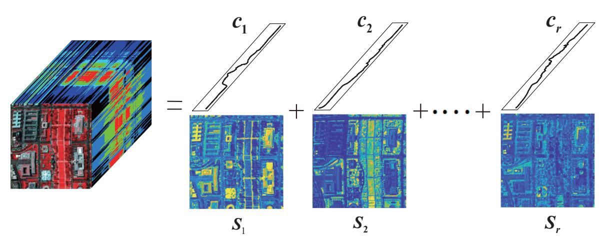

Figure 1. Illustration of the HSI decomposition utilized in DDS2M.

4. Concrete Network Structure of DDS2M

In DDS2M, we propose to introduce the attention mechanism [10] into the U-Net, for abundance map modeling which aims to enhance the self-supervised expression ability of the VS2M. The concrete network structure is illustrated in Figure 2.

5. Visualization of Reverse Diffusion Process

We visualize the sampling process in the Figure 3, and report the history PSNR values during the reverse diffusion process in Figure 4, in which HSI *Balloons* and *Fruits* are selected as examples.

References

[1] J. M. Bioucas-Dias and J. M. P. Nascimento. Hyperspectral subspace identification. *IEEE Trans. Geosci. Remote Sens.*, 46(8):2435–2445, 2008. 2

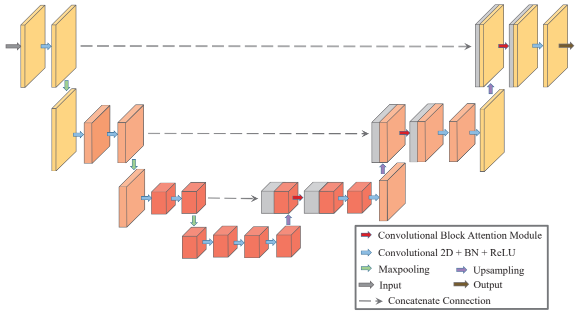

Figure 2. The concrete U-Net structure used in `DDS2M`.

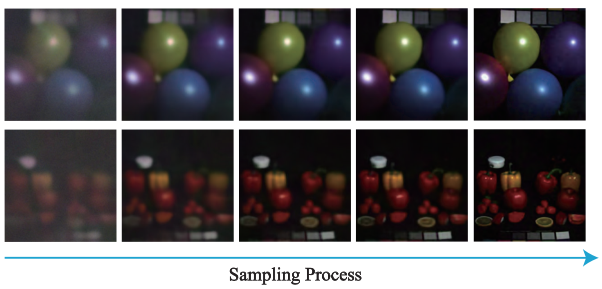

Figure 3. Visualization of the reverse diffusion process in `DDS2M`.

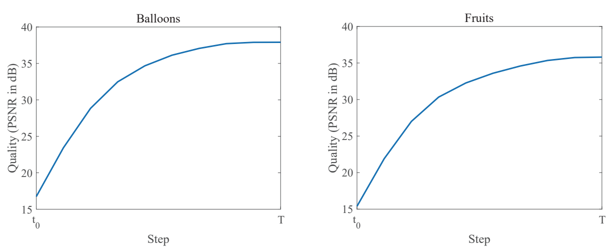

Figure 4. The history PSNR values during the reverse diffusion process.

[2] X. Fu, W.-K. Ma, J. M. Bioucas-Dias, and T.-H. Chan. Semiblind hyperspectral unmixing in the presence of spectral library mismatches. *IEEE Trans. Geosci. Remote Sens.*, 54(9):5171–5184, 2016. 2

[3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33:6840–6851, 2020. 1

[4] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *ICLR Workshops*, 2022. 1

[5] C. Lanaras, E. Baltsavias, and K. Schindler. Hyperspectral super-resolution by coupled spectral unmixing. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 3586–3594, 2015. 2

[6] L. Loncan, J. Chanussot, S. Fabre, and X. Briottet. Hyperspectral pansharpening based on unmixing techniques. In *Workshop Hyperspectral Image Signal Proces.: Evol. Remote Sens.*, pages 1–4, 2015. 2

[7] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Process. Mag.*, 31(1):67–81, 2014. 2

[8] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021. 1

[9] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *CVPR*, pages 9446–9454. IEEE, 2018. 2

[10] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *ECCV*, pages 3–19, 2018. 2

[11] N. Yokoya, T. Yairi, and A. Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.*, 50(2):528–537, 2012. 2

[12] Y. Zhao, J. Yang, C. Yi, and Y. Liu. Joint denoising and unmixing for hyperspectral image. In *Workshop Hyperspectral Image Signal Proces.: Evol. Remote Sens.*, pages 1–4, 2014. 2