1. Diffusion Process for DDS2M

Given a degraded HSI $y$, 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 $y$ [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:

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

(1)

where

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

(2)

where $\mathbf{x}_t^{(i)}$ is the $i$-th index of vector $\mathbf{x}_t = \mathbf{V}^T \mathbf{x}_t$, $\mathbf{y}_t^{(i)}$ is the $i$-th index of $\mathbf{y} = \Sigma^T \mathbf{U}^T \mathbf{y}$, $\sigma_t$ depending on the hyperparameter $\beta_{1:T}$ denotes the variance of diffusion noise in $\mathbf{x}_t$, and $\eta$, $\eta_b$ 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]:

$$
q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t} \mathbf{x}_0, (1 - \alpha_t) \mathbf{I})
$$

(3)

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

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

(4)

2. Loss Function Derivations

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

$$
E_{q(\mathbf{x}_0), q(\mathbf{y}|\mathbf{x}_0)} [\log p_{\theta, \zeta}(\mathbf{x}_0|\mathbf{y})] \\
\geq E_{q(\mathbf{x}_0), q(\mathbf{y}|\mathbf{x}_0)} \left[ \log \sum_{t>1} \frac{p_{\theta, \zeta}(\mathbf{x}_{t-1}|\mathbf{x}_0, \mathbf{y})}{q(\mathbf{x}_{t-1}|\mathbf{x}_0, \mathbf{y})} \right] \\
=q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0, y) \left[ \log \frac{p_{\theta, \zeta}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0, \mathbf{y})}{q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0, \mathbf{y})} \right]
$$

(5)
By maximizing the variational lower bound of \( \mathbb{E}_{q(x_0), q(y|x_0)} [\log p_{\theta, \zeta}(x_0, y)] \), we have

\[
\arg \max_{\{\theta, \zeta\}} \mathbb{E}_{q(x_{0:T}), q(y|x_0)} \left[ \log \frac{p_{\theta, \zeta}(x_{0:T}|y)}{q(x_{1:T}|x_0, y)} \right] \quad (6)
\]

\[
= \arg \max_{\{\theta, \zeta\}} \mathbb{E}_{q(x_{0:T}), q(y|x_0)} \left[ -D_{KL} \left( q(x_{T}|x_0, y) \parallel p(x_T|y) \right) \right.
\]

\[
- \sum_{t>1} D_{KL} \left( q(x_{t-1}|x_t, x_0, y) \parallel p_{\theta, \zeta}(x_{t-1}|x_t, y) \right)
\]

\[
+ \log p_{\theta, \zeta}(x_0|x_1, y)]
\]

\[
= \arg \max_{\{\theta, \zeta\}} \mathbb{E}_{q(x_{0:T}), q(y|x_0)} \left[ \log p_{\theta, \zeta}(x_0|x_1, y) \right.
\]

\[
- \sum_{t>1} D_{KL} \left( q(x_{t-1}|x_t, x_0, y) \parallel p_{\theta, \zeta}(x_{t-1}|x_t, y) \right)
\]

\[
+ \log p_{\theta, \zeta}(x_0|x_1, y)]
\]

\[
= \arg \min_{\{\theta, \zeta\}} \mathbb{E}_{q(x_{0:T}), q(y|x_0)} \left[ \log p_{\theta, \zeta}(x_0|x_1, y) \right.
\]

\[
+ \sum_{t>1} D_{KL} \left( q(x_{t-1}|x_t, x_0, y) \parallel p_{\theta, \zeta}(x_{t-1}|x_t, y) \right)
\]

\[
(7)
\]

For \( t > 1 \):

\[
\arg \min_{\{\theta, \zeta\}} \mathbb{E}_{q(x_{0:T}), q(y|x_0)} \left[ \sum_{t>1} D_{KL} \left( q(x_{t-1}|x_t, x_0, y) \parallel p_{\theta, \zeta}(x_{t-1}|x_t, y) \right) \right]
\]

\[
= \arg \min_{\{\theta, \zeta\}} \mathbb{E}_{q(x_{0:T}), q(y|x_0)} \left[ \sum_{t>1} D_{KL} \left( q(x_{t-1}|x_t, x_0, y) \parallel q(x_{t-1}|x_1, x_{\theta, \zeta}, y) \right) \right]
\]

\[
= \arg \min_{\{\theta, \zeta\}} \mathbb{E}_{q(x_{0:T}), q(y|x_0)} \|x_0 - x_{\theta, \zeta}\|_F^2
\]

\[
(8)
\]

For \( t = 1 \):

\[
\arg \min_{\{\theta, \zeta\}} \mathbb{E}_{q(x_{0:T}), q(y|x_0)} \left[ \log p_{\theta, \zeta}(x_0|x_1, y) \right]
\]

\[
= \arg \min_{\{\theta, \zeta\}} \mathbb{E}_{q(x_{0:T}), q(y|x_0)} \|x_0 - x_{\theta, \zeta}\|_F^2
\]

\[
(9)
\]

Therefore, the objective in Eqn. (5) can be reduced into a denoising objective, i.e., estimating the underlying high quality HSI \( x_0 \) from the noisy version \( x_t \). Inspired by the self-supervised loss functions in [9], our variational inference-based loss function can be designed as follows:

\[
\arg \min_{\{\theta, \zeta\}} \left\| x_t - \text{vec} \left( \sqrt{\alpha_t} \sum_{r=1}^R S_{\theta}(z_r) \circ C_{\zeta}(w_r) \right) \right\|_F^2.
\]

\[
(10)
\]

3. HSI Decomposition Utilized in DDS2M

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

\[
\chi = \sum_{r=1}^R S_r \circ c_r,
\]

\[
(11)
\]

where \( S_r \in \mathbb{R}^{J \times J} \) and \( c_r \in \mathbb{R}^K \) represent the \( r \)-th end-member’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

\[
\chi^{(i,j,k)} = \sum_{r=1}^R S_r^{(i,j)} c_r^k.
\]

\[
(12)
\]

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.](image)

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

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.


