HIR-Diff: Unsupervised Hyperspectral Image Restoration Via Improved Diffusion Models

Supplementary Material

1. Band Index Selection Using RRQR

1.1. Notation

The QR factorization of matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ is defined as

$$\mathbf{X} = \mathbf{Q}\mathbf{R} \equiv \mathbf{Q} \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix}, \tag{25}$$

where $\mathbf{Q} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix and $\mathbf{R} \in \mathbb{R}^{m \times n}$ is an upper triangular matrix. Then $\mathcal{R}_k(\mathbf{X})$ is defined as

$$\mathcal{R}_k(\mathbf{X}) = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix}. \tag{26}$$

For $\mathbf{R}_{11}$, $1/\omega_i(\mathbf{R}_{11})$ denotes the 2-norm of the ith row of $\mathbf{R}_{11}^{-1}$. For $\mathbf{R}_{22}$, $1/\gamma_i(\mathbf{R}_{22})$ denotes the 2-norm of the jth column of $\mathbf{R}_{22}$. Then $\rho(\mathbf{R}, k)$ is defined as

$$\rho(\mathbf{R}, k) = \max_{\substack{1 \le i \le k \\ 1 \le j \le n-k}} \sqrt{\left(\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\right)_{i,j}^2 + \left(\gamma_j(\mathbf{R}_{22})/\omega_i(\mathbf{R}_{11})\right)^2}. \tag{27}$$

$\mathbf{\Pi}_{i,j}$ denotes the permutation that interchanges the ith and jth columns of a matrix.

1.2. RRQR

In this section, we provide a more detailed description of the RRQR algorithm process [12] employed in our work. As introduced in Sect. 3.3, the RRQR is used to determine the band index $(i_1, i_2, ..., i_K)$ so that $|\det(\mathbf{V}_s)|$ is prevented from being zero and each band in the reduced image $\mathcal{A}$ is able to encode different image information. Given a matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ with $m \ge n$, the QR factorization of $\mathbf{M}$ with its columns permuted can be formulated as

$$\mathbf{M}\mathbf{\Pi} = \mathbf{Q}\mathbf{R} \equiv \mathbf{Q} \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix}, \tag{28}$$

where $\mathbf{Q} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix, $\mathbf{R} \in \mathbb{R}^{m \times n}$ is an upper triangular matrix and $\mathbf{\Pi}$ is a permutation matrix.

The RRQR aims to choose $\mathbf{\Pi}$ such that $\sigma_{\min}(\mathbf{R_{11}})$ is sufficiently large and $\sigma_{\max}(\mathbf{R_{22}})$ is sufficiently small, where $\sigma(\cdot)$ denotes the singular values. The RRQR factorization algorithm works by interchanging any pair of columns that sufficiently increases $|\det(\mathbf{R_{11}})|$, resulting in a large $|\det(\mathbf{R_{11}})|$. The details of the RRQR process are illustrated in Algorithm 2, where k and $f \ge 1$ are hyperparameters. It can be proven that $|\det(\mathbf{R_{11}})|$ increases strictly with every interchange and the algorithm is completed within a finite number of permutations. Readers could refer to [12] for more detailed proof. We replace $\mathbf{M}$ in Eq.(28) with $\mathbf{V}^{\mathrm{T}}$ and we have

Algorithm 2: RRQR Algorithm

Input: $\mathbf{M}, k, f$
Output: Permutation matrix $\mathbf{\Pi}$

Compute $\mathbf{R} := \mathcal{R}_k(M)$ and $\mathbf{\Pi} = \mathbf{I}$
while True **do**
  step 1: Compute $\rho(\mathbf{R}, k)$
  step 2: **if** $\rho(\mathbf{R}, k) \le f$ **then Break**;
  step 3: Find i and j such that $\rho(\mathbf{R}, k) > f$
  step 4: Compute $\mathbf{R} := \mathcal{R}_k(\mathbf{R}\mathbf{\Pi}_{i,j+k})$ and $\mathbf{\Pi} := \mathbf{\Pi}\mathbf{\Pi}_{i,j+k}$
end
return $\mathbf{\Pi}$

$$\mathbf{V}^{\mathrm{T}}\mathbf{\Pi_v} = \left[\, \mathbf{Q}_v\mathbf{R}_{v1} \;\; \mathbf{Q}_v\mathbf{R}_{v2} \,\right] \tag{29}$$

We define $\mathbf{V}_s^{\mathrm{T}}$ as $\mathbf{Q}_v\mathbf{R}_{v1}$ and then we could readily obtain $|\det(\mathbf{V}_s)| = |\det(\mathbf{R}_{v1})|$. Therefore, by solving the RRQR factorization problem of $\mathbf{V}^{\mathrm{T}}$ utilizing the algorithm proposed in [12], the permutation matrix $\mathbf{\Pi}_v$ is obtained and $|\det(\mathbf{V}_s)|$ is maximized. The indices corresponding to the first K columns of the permuted $\mathbf{V}^{\mathrm{T}}$ are defined as the band selection index.

2. Diffusion Model

In HIR-Diff, we employ a pre-trained diffusion model proposed in [10] to generate the reduced image $\mathcal{A}$. The diffusion model is a U-Net proposed in [40] and is trained on an amount of 3-channel remote sensing images without human supervision. Since the network requires the input image to be 3-channel, the rank value K in our work is set as 3 so that the reduced image $\mathcal{A}$ with 3 channels can be denoised with the pre-trained network. Although the rank value is small, we found that it is sufficient to restore the image details and helps to keep noise out of the estimated coefficient matrix E and the restored image, since the matrix $\mathbf{V}$ obtained from the SVD of the observed image as introduced in Sec. 3.3 is cleaner and the low-rank property enables noise reduction of the restored image.

3. Coefficient Matrix Estimation

In our work, we employ SVD and RRQR to estimate the coefficient matrix $\mathbf{E}$ as introduced in Sec. 3.3. The visualization results of the estimated $\mathbf{E}$ for the WDC dataset

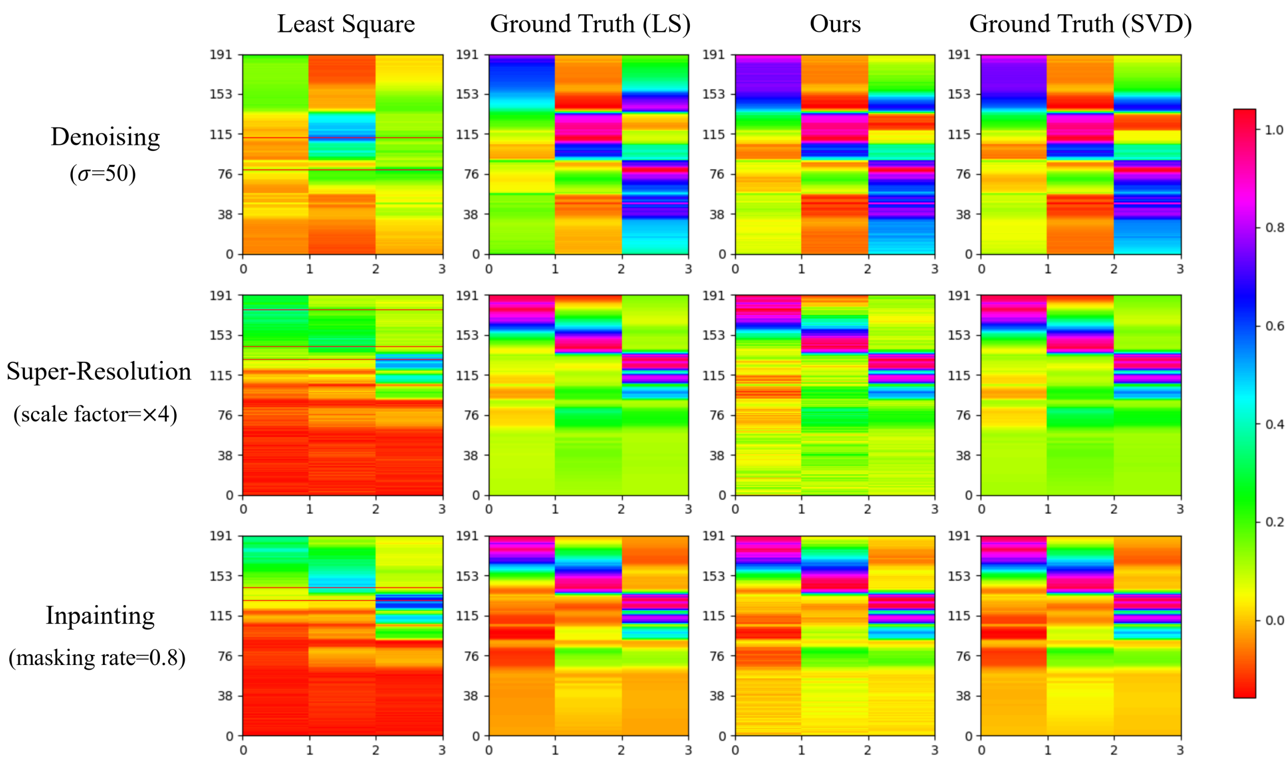

Figure 5. The visualization results of the estimated $\mathbf{E}$. **Least Square** and **Ground Truth (LS)** denote the coefficient matrix $\mathbf{E}$ estimated by employing the least square method with the observed image and the clean image, respectively. **Ours** and **Ground Truth (SVD)** denote the coefficient matrix $\mathbf{E}$ estimated using SVD and RRQR proposed in our work with the observed image and the clean image, respectively.

are demonstrated in Fig. 5. The results of the matrix $\mathbf{E}$ estimated using the least square method proposed in [38] is also provided for comparison. Specifically, the least-squares method directly selects several bands from the observed image $\mathcal{Y}$ as the reduced image $\mathcal{A}$, and then estimates the coefficient matrix E by solving the least-squares problem.

$$\arg\min_{\mathbf{E}} ||\mathcal{Y} - \mathcal{A} \times_3 E||_F^2 \, . \tag{30}$$

Since there is a lot of noise in the reduced image $\mathcal{A}$ as the observed image suffers from various degradation, the estimated $\mathbf{E}$ is unreliable, resulting in undesirable HSI restoration performance. On the contrary, our estimated coefficient matrix $\mathbf{E}$ is robust to noise and exhibits a high degree of similarity to ground truth, verifying the effectiveness of our estimation method.