Toward Stable, Interpretable, and Lightweight Hyperspectral Super-resolution
Supplementary Material

Wen-jin Guo 1, Weiying Xie 1, Kai Jiang 1, Yunsong Li 1, Jie Lei 1, Leyuan Fang 2
1 State Key Laboratory of Integrated Services Networks, Xidian University
2 College of Electrical and Information Engineering, Hunan University
guowenjin@stu.xidian.edu.cn wyxie@xidian.edu.cn xdjiangkai@foxmail.com
jielei, ysli@mail.xidian.edu.cn fangleyan@gmail.com

A. Analysis on Convergence

Firstly, we can find:

\[ \log p(\phi, \theta | \mathcal{X}, \mathcal{Y}) = \log p(\mathcal{Z}, \phi, \theta | \mathcal{X}, \mathcal{Y}) - \log p(\mathcal{Z} | \mathcal{X}, \mathcal{Y}, \phi). \] (1)

Then, we take the expectation of \( Z|X \), \( \phi(t) \), \( \theta(t) \) for the left and right sides of the equation:

\[
E_{Z|X,\phi(t),\theta(t)}[\log p(\phi, \theta | \mathcal{X}, \mathcal{Y})] = \\
E_{Z|X,\phi(t),\theta(t)}[\log p(\mathcal{Z}, \phi, \theta | \mathcal{X}, \mathcal{Y})] - \\
E_{Z|X,\phi(t),\theta(t)}[\log p(\mathcal{Z} | \mathcal{X}, \mathcal{Y}, \phi)], \] (2)

where \( \phi(t) \) and \( \theta(t) \) are parameters in \( t \)-th iteration. After simplifications, we can find:

\[
\log p(\phi, \theta | \mathcal{X}, \mathcal{Y}) = \mathcal{L}(\mathcal{Z}, \theta, \phi; \mathcal{X}, \mathcal{Y}) - \\
E_{Z|X,\phi(t),\theta(t)}[\log p(\mathcal{Z} | \mathcal{X}, \mathcal{Y}, \phi)], \] (3)

where \( \mathcal{L}(\mathcal{Z}, \theta, \phi; \mathcal{X}, \mathcal{Y}) \) is the ELBO in \( t \)-th iteration.

In M-step, the parameters \( \phi \) and \( \theta \) are updated by maximizing the ELBO:

\[ \phi, \theta = \arg \max \mathcal{L}(\mathcal{Z}, \theta, \phi; \mathcal{X}, \mathcal{Y}). \] (4)

Thus, we have:

\[ \mathcal{L}(\mathcal{Z}, \theta(t+1), \phi(t+1); \mathcal{X}, \mathcal{Y}) \geq \mathcal{L}(\mathcal{Z}, \theta, \phi; \mathcal{X}, \mathcal{Y}). \] (5)

It’s easy to find:

\[ \mathcal{L}(\mathcal{Z}, \theta(t+1), \phi(t+1); \mathcal{X}, \mathcal{Y}) \geq \mathcal{L}(\mathcal{Z}, \theta(t), \phi(t); \mathcal{X}, \mathcal{Y}). \] (6)

For \( E_{Z|X,Y,\phi(t),\theta(t)}[\log p(\mathcal{Z} | \mathcal{X}, \mathcal{Y}, \phi)] \), because:

\[ \mathbb{D}_{KL}(\mathcal{Z}|\mathcal{X},\mathcal{Y},\phi(t),\theta(t)||\mathcal{Z}|\mathcal{X},\mathcal{Y},\phi(t+1),\theta(t+1)) \geq 0. \] (7)

It’s equivalent to:

\[ E_{Z|X,Y,\phi(t),\theta(t)}[\log p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\phi(t+1),\theta(t+1)) / p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\phi(t),\theta(t))] \leq 0. \] (8)

Thus, we have:

\[ p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\phi(t+1),\theta(t+1)) \geq p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\phi(t),\theta(t)). \] (9)

With combining the Tab. 2, we have:

\[ \log p(\phi(t+1), \theta(t+1) | \mathcal{X}, \mathcal{Y}) \geq \log p(\phi(t), \theta(t) | \mathcal{X}, \mathcal{Y}). \] (10)

For the posterior of \( Z \), \( p(Z|X,Y,\phi,\theta) = p(Z|\phi)p(X|Y|Z,\theta) \), since the prior of \( Z \) is nearly static (the prior knowledge is contained in the network structure), and the likelihood is maximized in M-step, we can find:

\[ p(Z(t+1)|X,Y,\phi(t+1),\theta(t+1)) \geq p(Z(t)|X,Y,\phi(t),\theta(t)). \] (11)

Thus, we have:

\[ p(Z(t+1), \phi(t+1), \theta(t+1) | X, Y) \geq p(Z(t), \phi(t), \theta(t) | X, Y). \] (12)

Since \( p(Z, \phi, \theta | X, Y) \) is upper bounded by 1, thus the EM-algorithm in our method is convergent.

B. Analysis on the Coordination Optimization

To further explore the effort of the coordination optimization strategy, we record the estimated PSFs and metrics in different iterations during training. As shown in Fig. 1, we initialize the blur kernel as an isotropic Gaussian kernel with variance [0.1, 0.1]. The estimated kernel is gradually adapted toward the groundtruth. Meanwhile, the metrics of
recovered HR-HSI grow in step with the degradation estimation, which confirms the existence of the positive feedback loop between fusion and degradation estimation.

As a comparison, we break the feedback cycle in coordination optimization. In detail, we replace the estimated PSF in the optimization of the fusion module with fixed PSF (isotropic Gaussian kernel with variance $[0.1, 0.1]$). The result is depicted in Fig. 2.

The degradation estimation converges slower compared with the setting under coordination optimization (50 epoch vs 40 epoch). Similarly, the metrics grow more slowly and volatilely. Without the guidance of degradation estimation, the fusion module lost in incorrect optimized direction. Then, the inaccurate recovery confuses the degradation estimation.

We also contrasts the performance of fine-tuning under the fixed degradation and estimated degradation. As shown in Fig. 3, fine-tuning with coordination optimization realizes better and stabler metrics than the setting without coordination optimization, especially in PSNR. The large difference in performance verifies the vital role of the guidance of correct degradation. Moreover, we fine-tune the models trained with coordination optimization and without coordination optimization. In the fine-tuning stage, we guide the fusion model with estimated PSF. As shown in Fig. 4, the two settings have almost the same performance, which further confirms the vital effort of the guidance from estimated degradation.

C. Hyper-parameter Analysis

We set $\alpha_1, \alpha_2, \alpha_3, \beta = 0.5, 0.2, 0.2, 0.5$ as the default setting in our method. To analyze the effort of each parameter, we vary each parameter in the range of $[0, 0.8]$ with a step size of 0.1. During testing one parameter, the other parameters are fixed. $\alpha_1, \alpha_2, \alpha_3$ are parameters corresponding to degradation estimation. As shown in Fig. 5, our method performs very stably in various settings of these parameters. Meanwhile, even the weight of one parameter is set zero, the other parameters will compensate the eliminated effort.
The hyper-parameter $\beta$ determines the step length of gradient descent in the fusion module. With setting $\beta = 0$, the fusion module keeps the initial weights during the whole training stage. The experiment in Fig. 6 confirms above indication though the poor performance when $\beta = 0$. As a contrast, our method perform well and stably in the other settings of $\beta$. Finally, we select the best setting $[0.5, 0.2, 0.2, 0.5]$ as the default parameter weights of our method.

**D. Network Structure**

In our method, the fusion module is established by a variational auto-encoder. The encoder is conducted by 5 convolution layers with concatenating HR-MSI before each layer to maintain the spatial information. The decoder is conducted by two fully-connected layers and the number of nodes in each layer depends on the number of bands in HR-HSI at each dataset. Due to the lightweight structure, the network is computationally efficient.

**E. Experiment under Image-wise Changed Degradation**

In this circumstance, we synthesize LR-HSIs in the training set and testing set through random selecting PSFs from the generated 6 blur kernels. As shown in Tab. 2, the proposed method outperforms other methods with a large margin. Refer to the experiments of consistent degradation in training and testing, we can find that the unsupervised methods have smaller performance decline than supervised methods, which indicates the necessity of fine-tuning in HSI-SR. The visual result is depicted in Figs. 7 to 9. Obviously, our method surpass other methods in edge-preserving.
Table 2. Average performance of test methods on three synthetic datasets under inconsistent degradation and various degradation.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>CAVE</th>
<th>Harvard</th>
<th>Chikusei</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>SAM</td>
<td>ERGAS</td>
</tr>
<tr>
<td>Methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNMF</td>
<td>27.4</td>
<td>17.9</td>
<td>1.33</td>
</tr>
<tr>
<td>Hysure</td>
<td>28.0</td>
<td>24.0</td>
<td>1.25</td>
</tr>
<tr>
<td>CSTF</td>
<td>28.2</td>
<td>15.8</td>
<td>1.12</td>
</tr>
<tr>
<td>MHFNet</td>
<td>32.2</td>
<td>10.0</td>
<td>1.17</td>
</tr>
<tr>
<td>CUCaNet</td>
<td>34.1</td>
<td>9.10</td>
<td>0.964</td>
</tr>
<tr>
<td>UAL</td>
<td>34.9</td>
<td>8.97</td>
<td>0.797</td>
</tr>
<tr>
<td>Ours</td>
<td>39.5</td>
<td>6.83</td>
<td>0.355</td>
</tr>
</tbody>
</table>

Figure 7. Visual SR results and the corresponding error images on scene of glass tiles in CAVE dataset under the image-level changed degradation, where we display the 20th (580nm) band of the HR-HSI images.

Figure 8. Visual SR results and the corresponding error images of imgf33 in Harvard dataset under the image-level changed degradation, where we display the 18th (540nm) band of the HR-HSI images.

Figure 9. Visual SR results and the corresponding error images on Chikusei dataset under the image-level changed degradation, where we display the test images with bands 70-100-36 as R-G-B to show.