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 fangleyuan@gmail.com

A. Analysis on Convergence

Firstly, we can find:

$$log\, p(\boldsymbol{\phi}, \boldsymbol{\theta}|\mathcal{X}, \mathcal{Y}) = log\, p(\mathcal{Z}, \boldsymbol{\phi}, \boldsymbol{\theta}|\mathcal{X}, \mathcal{Y}) - log\, p(\mathcal{Z}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi}). \tag{1}$$

Then, we take the expectation of $\mathcal{Z}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi}^{(t)}, \boldsymbol{\theta}^{(t)}$ for the left and right sides of the equation:

$$\begin{aligned} &\mathbb{E}_{\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi}^{(t)},\boldsymbol{\theta}^{(t)}}[log\, p(\boldsymbol{\phi}, \boldsymbol{\theta}|\mathcal{X}, \mathcal{Y})] = \\ &\quad \mathbb{E}_{\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi}^{(t)},\boldsymbol{\theta}^{(t)}}[log\, p(\mathcal{Z}, \boldsymbol{\phi}, \boldsymbol{\theta}|\mathcal{X}, \mathcal{Y})] - \\ &\mathbb{E}_{\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi}^{(t)},\boldsymbol{\theta}^{(t)}}[log\, p(\mathcal{Z}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi})], \end{aligned} \tag{2}$$

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

$$\begin{aligned} log\, p(\boldsymbol{\phi}, \boldsymbol{\theta}|\mathcal{X}, \mathcal{Y}) = \mathcal{L}(\mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{X}, \mathcal{Y}) - \\ \mathbb{E}_{\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi}^{(t)},\boldsymbol{\theta}^{(t)}}[log\, p(\mathcal{Z}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi})], \end{aligned} \tag{3}$$

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

In M-step, the parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are updated by maximizing the ELBO:

$$\boldsymbol{\phi}, \boldsymbol{\theta} = arg\, max\, \mathcal{L}(\mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{X}, \mathcal{Y}). \tag{4}$$

Thus, we have:

$$\mathcal{L}(\mathcal{Z}, \boldsymbol{\theta}^{(t+1)}, \boldsymbol{\phi}^{(t+1)}; \mathcal{X}, \mathcal{Y}) \geq \mathcal{L}(\mathcal{Z}, \boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{X}, \mathcal{Y}). \tag{5}$$

It's easy to find:

$$\mathcal{L}(\mathcal{Z}, \boldsymbol{\theta}^{(t+1)}, \boldsymbol{\phi}^{(t+1)}; \mathcal{X}, \mathcal{Y}) \geq \mathcal{L}(\mathcal{Z}, \boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}^{(t)}; \mathcal{X}, \mathcal{Y}). \tag{6}$$

For $\mathbb{E}_{\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi}^{(t)},\boldsymbol{\theta}^{(t)}}[log\, p(\mathcal{Z}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi})]$, because:

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

It's equivalent to:

$$\mathbb{E}_{\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi}^{(t)},\boldsymbol{\theta}^{(t)}}[log\, \frac{p(\mathcal{Z}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi}^{(t+1)}, \boldsymbol{\theta}^{(t+1)})}{p(\mathcal{Z}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi}^{(t)}, \boldsymbol{\theta}^{(t)})}] \leq 0. \tag{8}$$

Thus, we have:

$$p(\mathcal{Z}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi}^{(t+1)}, \boldsymbol{\theta}^{(t+1)}) \geq p(\mathcal{Z}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi}^{(t)}, \boldsymbol{\theta}^{(t)}). \tag{9}$$

With combining the Tab. 2, we have:

$$log\, p(\boldsymbol{\phi}^{(t+1)}, \boldsymbol{\theta}^{(t+1)}|\mathcal{X}, \mathcal{Y}) \geq log\, p(\boldsymbol{\phi}^{(t)}, \boldsymbol{\theta}^{(t)}|\mathcal{X}, \mathcal{Y}). \tag{10}$$

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

$$p(\mathcal{Z}^{(t+1)}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi}^{(t+1)}, \boldsymbol{\theta}^{(t+1)}) \geq p(\mathcal{Z}^{(t)}|\mathcal{X}, \mathcal{Y}, \boldsymbol{\phi}^{(t)}, \boldsymbol{\theta}^{(t)}). \tag{11}$$

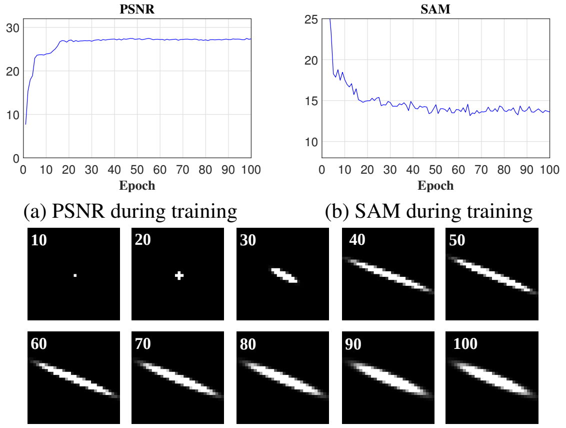
Thus, we have:

$$p(\mathcal{Z}^{(t+1)}, \boldsymbol{\phi}^{(t+1)}, \boldsymbol{\theta}^{(t+1)}|\mathcal{X}, \mathcal{Y}) \geq p(\mathcal{Z}^{(t)}, \boldsymbol{\phi}^{(t)}, \boldsymbol{\theta}^{(t)}|\mathcal{X}, \mathcal{Y}). \tag{12}$$

Since $p(\mathcal{Z}, \boldsymbol{\phi}, \boldsymbol{\theta}|\mathcal{X}, \mathcal{Y})$ 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



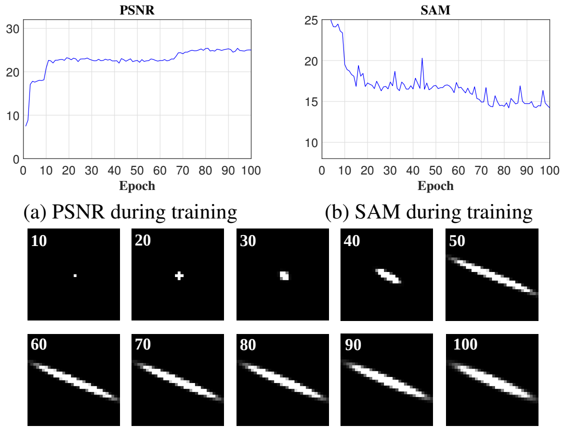
(a) PSNR during training (b) SAM during training

(c) Estimated PSFs during training

Figure 1. Metrics and estimated blur kernels during training with the coordination optimization.

recovered HR-HSI grow in step with the degradation estimation, which confirms the existence of the postive feedback loop between fusion and degradation estimation.

As a comparsion, 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.



(a) PSNR during training (b) SAM during training

(c) Estimated PSFs during training

Figure 2. Metrics and estimated blur kernels during training without the coordination optimization.

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 dirction. 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.

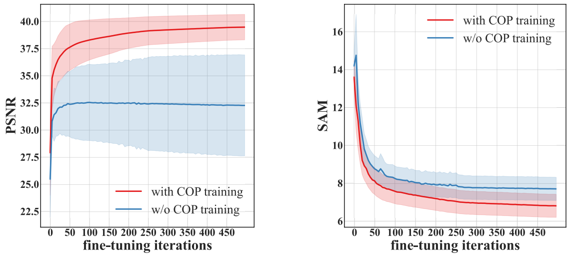


Figure 3. Performance comparsion between fine-tuning with coordination optimization (with COP) and without coordination optimization (w/o COP).

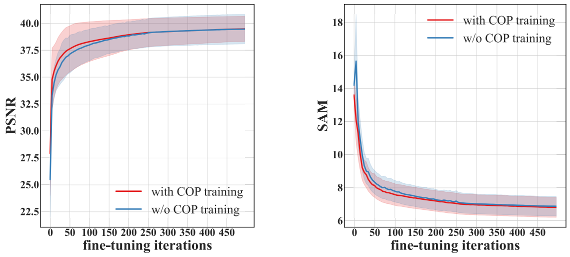
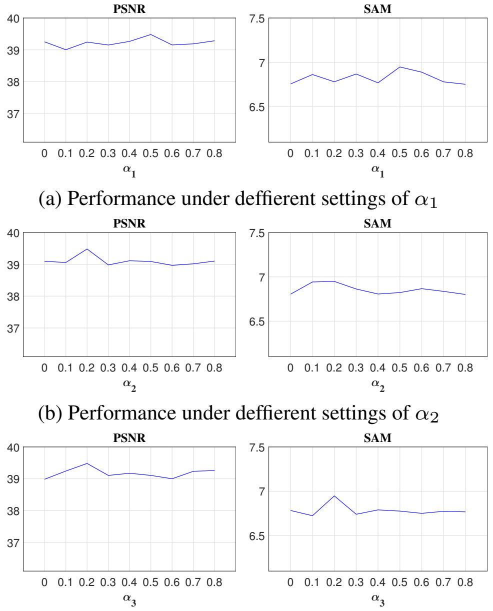


Figure 4. Performance comparsion between fine-tuning from the model trained with coordination optimization (with COP training) and fine-tuning from the model without coordination optimization (w/o COP training).

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.



(a) Performance under deffierent settings of α_1

(b) Performance under deffierent settings of α_2

(c) Performance under deffierent settings of α_3

Figure 5. Analysis on parameters related to degradation estimation.

The hyper-parameter β 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 β. Finally, we select the best setting $[0.5, 0.2, 0.2, 0.5]$ as the default parameter weights of our method.

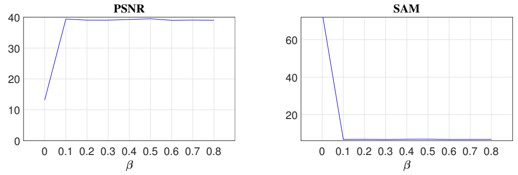


Figure 6. Analysis on parameter related to fusion.

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.

Table 1. Detailed network architecture of the fusion module.

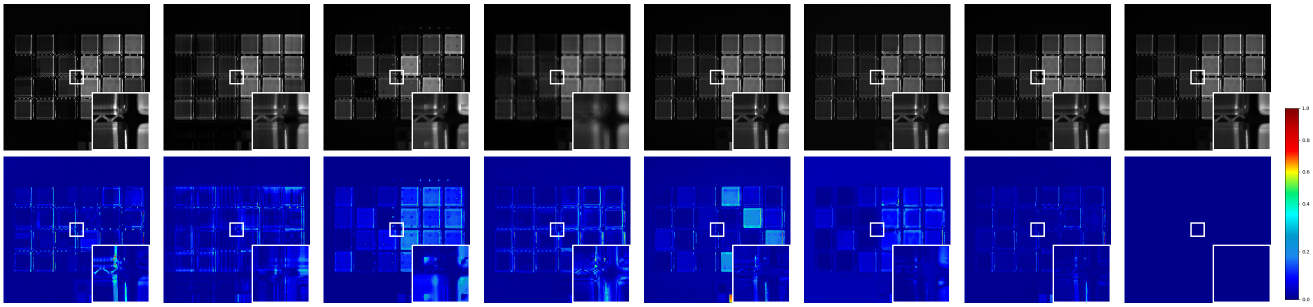
dataset		CAVE/ Harvard	Chikusei	WV 2
module	operation	kernel size		
Encoder	CONV+ LReLU	$4 \times 3 \times 1 \times 1$		
	Concat	HR-MSI		
	CONV+ LReLU	$7 \times 8 \times 1 \times 1$		
	Concat	HR-MSI		
	CONV+ LReLU	$11 \times 16 \times 1 \times 1$		
	Concat	HR-MSI		
	CONV+ LReLU	$19 \times 24 \times 1 \times 1$		
	Concat	HR-MSI		
	CONV+ LReLU	$24 \times 16 \times 1 \times 1$		
Decoder$_{\boldsymbol{\mu}}$	FC+ LReLU	16×24	16×64	16×18
	FC+ ReLU	24×31	64×128	18×8
Decoder$_{\boldsymbol{\sigma}}$	FC+ LReLU	16×24	16×64	16×18
	FC+ ReLU	24×31	64×128	18×8

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. Refered 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.

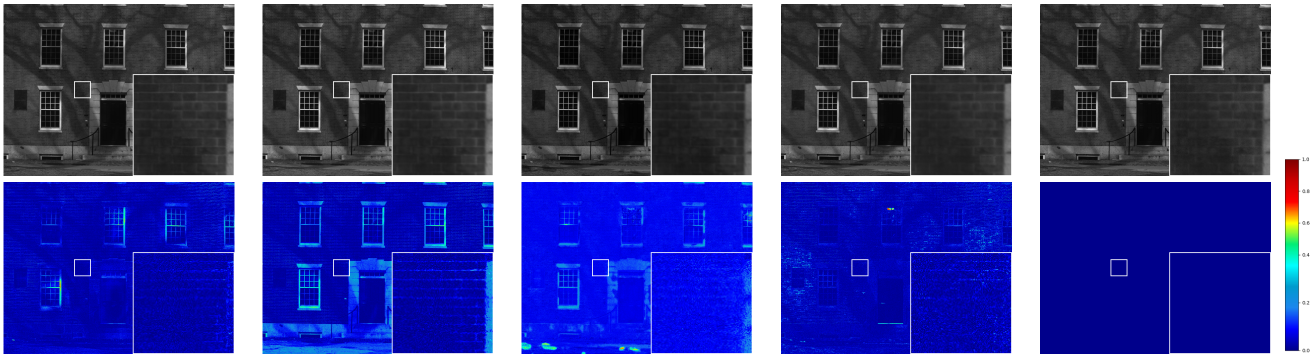
Table 2. Average performance of test methods on three synthetic datasets under inconsistent degradation and various degradation.

Datasets	CAVE				Harvard				Chikusei			
Methods	PSNR	SAM	ERGAS	UIQI	PSNR	SAM	ERGAS	UIQI	PSNR	SAM	ERGAS	UIQI
CNMF	27.4	17.9	1.33	0.815	25.6	13.3	1.67	0.941	26.7	19.7	1.56	0.849
Hysure	28.0	24.0	1.25	0.791	35.3	6.97	0.638	0.927	21.9	10.8	2.71	0.631
CSTF	28.2	15.8	1.12	0.816	34.8	6.91	0.632	0.941	24.6	9.74	1.76	0.849
MHFNet	32.2	10.0	1.17	0.761	40.8	3.52	0.317	0.985	32.1	5.51	0.705	0.958
CUCaNet	34.1	9.10	0.964	0.922	37.8	6.14	1.42	0.917	28.1	5.77	0.781	0.891
UAL	34.9	8.97	0.797	0.845	40.7	8.90	0.931	0.938	26.2	6.23	1.42	0.789
Ours	**39.5**	**6.83**	**0.355**	**0.944**	**42.2**	**3.23**	**0.280**	**0.995**	**32.2**	**3.80**	**0.665**	**0.979**



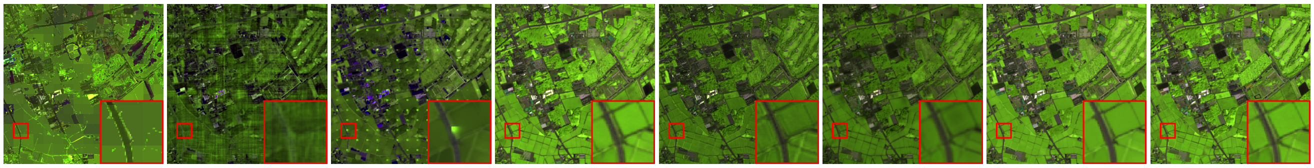
(a) CNMF (b) CSTF (c) HySure (d) MHFNet (e) CUCaNet (f) UAL (g) Ours (h) Ground truth

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.



(a) MHFNet (b) CUCaNet (c) UAL (d) Ours (e) Ground truth

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.



(a) CNMF (b) CSTF (c) HySure (d) MHFNet (e) CUCaNet (f) UAL (g) Ours (h) HR-HSI

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.