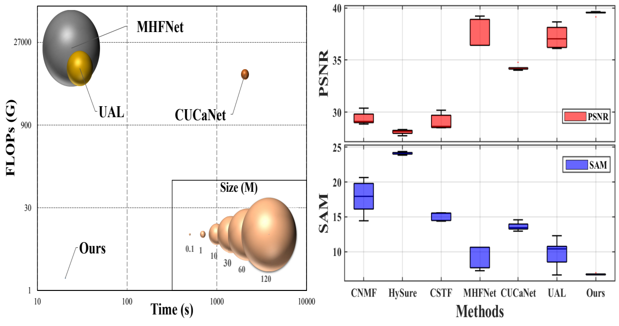
Toward Stable, Interpretable, and Lightweight Hyperspectral Super-resolution

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

Abstract

For real applications, existing HSI-SR methods are not only limited to unstable performance under unknown scenarios but also suffer from high computation consumption. In this paper, we develop a new coordination optimization framework for stable, interpretable, and lightweight HSI-SR. Specifically, we create a positive cycle between fusion and degradation estimation under a new probabilistic framework. The estimated degradation is applied to fusion as guidance for a degradation-aware HSI-SR. Under the framework, we establish an explicit degradation estimation method to tackle the indeterminacy and unstable performance caused by the black-box simulation in previous methods. Considering the interpretability in fusion, we integrate spectral mixing prior into the fusion process, which can be easily realized by a tiny autoencoder, leading to a dramatic release of the computation burden. Based on the spectral mixing prior, we then develop a partial fine-tune strategy to reduce the computation cost further. Comprehensive experiments demonstrate the superiority of our method against the state-of-the-arts under synthetic and real datasets. For instance, we achieve a 2.3 *dB promotion on PSNR with* $120\times$ *model size reduction and* $4300\times$ *FLOPs reduction under the CAVE dataset. Code is available in* `https://github.com/WenjinGuo/DAEM`.



(a) (b)

Figure 1. Comparison among recent state-of-the-art (SOTA) methods and our method. We report the computational efficiency (Size, FLOPs, and Time) in (a) and the distribution of two measurement metrics (PSNR, SAM) under various degradations in (b). Obviously, our method is remarkably superior to others in lightweight, fidelity, and stability.

1. Introduction

Different from traditional optical images with a few channels, hyperspectral images (HSIs) with tens to hundreds of bands hold discriminative information about materials, leading to a great advantage in a wide range of applications, e.g., the monitoring and management of ecosystems, biodiversity, and disasters [1–8]. However, the physical limitation in imaging causes a trade-off between spatial and spectral resolution. HSIs are suffered from low spatial resolution in real applications. Therefore, hyperspectral super-resolution (HSI-SR), which aims to promote the spatial resolution of HSIs, has become a significant task, and always performs as a necessary pre-processing of HSI applications, i.e. detection and classification [9–16].

Due to the common optical platforms which are equipped with both HSI sensor and multispectral sensor (such as satellites, airborne platforms), HSIs and multispectral images (MSIs) imaged in the same scene are easy to access. Fusion-based HSI-SR aims to estimate the desired high resolution HSI (HR-HSI) from the corresponding low resolution HSI (LR-HSI) and high resolution MSI (HR-MSI). Naturally, HSI-SR can be modeled as a maximum a posterior (MAP) estimation:

$$p(\mathcal{Z}|\mathcal{X},\mathcal{Y}) \propto p(\mathcal{X},\mathcal{Y}|\mathcal{Z},\boldsymbol{\theta})p(\mathcal{Z}|\boldsymbol{\phi}), \tag{1}$$

where $\mathcal{Z} \in \mathbb{R}^{H\times W\times B}$ is the target HR-HSI with H, W and B as its height, width and number of bands, respectively. $\mathcal{X} \in \mathbb{R}^{h\times w\times B}$ is the observed LR-HSI with h and w as

[*] Equal contribution.
[†] Cooresponding author.

its height and width ($h < H, w < W$). $\mathcal{Y} \in \mathbb{R}^{H \times W \times b}$ is the HR-MSI with b bands ($b < B$). $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are parameters in the likelihood and the prior, respectively. As a pre-processing task, HSI-SR faces two challenges, high fidelity recovery and efficient processing (i.e., low computational burden and fast processing). Referred to Eq. (1), we analyse these two ingredients from three perspectives: the likelihood term, the prior term, and coordination between them.

Likelihood Term

The likelihood term reflects the degradation process, and the widely accepted degradation model can be formulated as:

$$\begin{aligned} \mathcal{X} &= (\mathcal{Z} * C) \downarrow_s + \mathcal{N}_{\mathcal{X}}, \\ \mathcal{Y} &= \mathcal{Z} \times_3 R + \mathcal{N}_Y, \end{aligned} \tag{2}$$

where $C \in \mathbb{R}^{s \times s}$, $R \in \mathbb{R}^{B \times b}$ represent the PSF (point spread function) and SRF (spectral response function), respectively. $\downarrow_s$ represents spatial downsampling with s times. $\times_3$ is matrix multiple on the third dimension. Early methods demand precise degradation parameters to optimize the likelihood [17–19]. However, PSF and SRF are various and unknown in real scenarios, hence the high demand for blind HSI-SR. In [20], two convolution blocks are built to simulate the degradation process, which learn the degradation in training sets and apply it to testing samples. Unsupervised blind HSI-SR methods estimate degradation in each image-pair individually [21–23]. Yao *et al*. [22] propose a spatial-spectral consistency to further constrain the degradation estimation. Despite better applicability, recent blind HSI-SR methods are limited by the DL-based estimation. On one hand, the implicit modeling of degradation imports a black box in optimization and draws uncertainty, especially in volatile degradation. On the other hand, degradation estimation is an inverse problem with multiple candidates, the common over-fitting in neural networks will lead to inaccurate and unstable performance.

Prior Term

If only considering the likelihood term, HSI-SR is an ill-posed problem with infinite solutions. The prior term shrinks the solution space as a regularization. Earlier works make assumptions on prior through manual induction of data characteristics, such as low rank [24, 25], and sparsity [26–30]. Recent supervised deep learning (DL)-based methods replace this process through data characterization by neural networks. These inductive methods will drop a lot on performance when testing samples differ training samples, hence a critical need for general prior assumption. As an inherent feature in HSIs, spectral mixing prior simulates the common spectral mixing phenomenon in imaging. Unmixing-based methods make a breakthrough in fidelity [17, 18, 31, 32]. Moreover, to promote the generalization, coupled autoencoder is proposed to simulate the spectral mixing with deep learning toolkit [21, 22]. Despite superior fidelity, recent unsupervised DL-based methods overfocus on network architecture and underestimate the effort of prior knowledge in HSIs, resulted in two drawbacks. On one hand, the over-designed network structures weaken the role of the prior assumption, which harms the interpretability. On the other hand, the complicated network structures pose a heavy burden on power and memory.

Coordination between Likelihood and Prior

The most recent blind HSI-SR methods generally contain two modules, the fusion module and the degradation estimator [22,23,33]. From the viewpoint of MAP problem, the former aims to recover the HR-HSI through the observations under the regularization of the prior. The later minimizes the likelihood term by estimating degradation parameters. The unknown HR-HSI and degradation determine the observations. Naturally, the estimated degradation can be fed to the fusion module as a guidance. However, in recent methods, the degradation estimator and fusion module only interacts in backward stage. The learned degradation parameters are not involved in the updating of fusion module or prior parameters, which brings two weaknesses. Firstly, the nearly independent optimization of likelihood and prior causes more updating steps and slows the convergence. Secondly, optimization with less interaction inducts conflict optimization directions and results in local optimum with unreal recovery.

In this paper, we develop a novel coordination optimization framework for stable, interpretable, and lightweight HSI-SR. Through integrating the Wald protocol in the MAP problem, we construct a postive feedback loop between prior and likelihood. Based on the framework, we explore an explicit degradation estimation to remedy the unstable performance of black-box estimation. As for the prior, we establish a lightweight autoencoder to simulate the spectral mixing prior in HSIs for a interpretable fusion. Our contributions are summarized as follows:

1. We explore a coordination optimization framework for HSI-SR with fast convergence and stable performance. Under the framework, the degradation estimation and fusion process promote each other concordantly. To the best of our knowledge, it is the first work to explore the cooperative relationship between prior and likelihood in HSI-SR.

2. An explicit estimation method is established in HSI-SR firstly. Through modeling PSF and SRF with anisotropic Gaussian kernel and Gaussian mixture, tiny parameters realize stable and precise estimation.

3. We simulate the general spectral mixing prior with only one interpretable autoencoder. Compared with recent complicated models, the network makes a breakthrough in lightweight. Based on the fusion netwrok, we explore a partial optimization strategy in test stage, which only updates the decoder that handles the individual spectral feature of

images, reducing large computaion and time consumption.

2. Related Work

2.1. Spectral Mixing Prior in HSI-SR

Recent unmixing-based methods implement the spectral mixing prior through coupled autoencoder [21, 22, 33]. The observed LR-HSI and HR-MSI are reconstructed by two autoencoders to extract abundances (latent features) and endmembers (weights of decoders). Then HR-HSI is generated by the decoder of LR-HSI with the latent feature of HR-MSI as input. There are two limitations. Firstly, the separate unmixing of LR-HSI and HR-MSI leads to a particular sensitivity to the divergence of autoencoders' features. Even several modifications are applied, such as cross-attention [22] and alternative optimization [21]. Secondly, high-dimensional hyperspectral data dramatically increases the computational burden on coupled autoencoder. By contrast, we simulate the unmixing prior with only one autoencoder, leading to more accurate results while reducing the computational complexity significantly.

2.2. Degradation Estimation

Image enhancement is a typical ill-posed problem with infinite solutions, hence the urgent necessity for precise degradation estimation to constrain the solution space.

Estimation of Blur Kernel

In single image super-resolution (SISR), estimation of blur kernel become a hot topic. Wang *et al.* [34] propose an contrastive learning scheme for degradation representation under relative distances between different blur kernels. Luo *et al.* [35] introduce probabilistic model and build a single network to mapping the distribution of blur kernels. In [36], an explicit degradation estimation scheme is proposed with modeling blur kernel with anisotropic Gaussian. In HSI-SR, Qu *et al.* [21] fisrt estimate PSF unsupervisedly with several convolution layers. Yao *et al.* [22] regularize the estimation with spatial-spectral consistency. In [33], a single convolution layer is applied to simulate the PSF and realize closer result to groundtruth.

Estimation of SRF

SRF is a specific degradation in HSI-SR. Previous methods model SRF with 1×1 convolution layers or fully-connect layers [21–23, 33, 37].

2.3. Wald Protocol

The target HR-HSI with same spatial resolution to HR-MSI is not accessible in real world. To measure the performance of HSI-SR methods, Wald protocol indicates that the performance of recovering LR-HSI from degraded HR-MSI and LR-HSI is consistent to the recovery of unknown HR-HSI, which links the degradation and fusion [38].

3. The Proposed Method

3.1. Problem Formulation

According to Eq. (1), HSI-SR is a MAP problem determined by parameters $\boldsymbol{\theta} = \{C, R\}$ and $\boldsymbol{\phi}$. In practice, the degradation parameter $\boldsymbol{\theta}$ and prior knowledge $\boldsymbol{\phi}$ are not completely known. Thus, we re-model blind HSI-SR with an additional inference of parameters:

$$\begin{aligned} max_{\mathcal{Z},\boldsymbol{\theta},\boldsymbol{\phi}}\, log\, p(\mathcal{X},\mathcal{Y}|\mathcal{Z},\boldsymbol{\theta}) + log\, p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi}) \\ + log\, p(\boldsymbol{\phi}|\mathcal{X},\mathcal{Y},\boldsymbol{\theta}) + log\, p(\boldsymbol{\theta}|\mathcal{X},\mathcal{Y}) \end{aligned}, \tag{3}$$

where $p(\boldsymbol{\theta}|\mathcal{X},\mathcal{Y})$ represents inferring degradation parameters $\boldsymbol{\theta}$ without the guidance of HR-HSI, leading to unsupervised degradation estimation. The estimated degradation $\boldsymbol{\theta}$ is involved in the inference of fusion module, i.e., $p(\boldsymbol{\phi}|\mathcal{X},\mathcal{Y},\boldsymbol{\theta})$, which limits the freedom of original MAP model for a coordination optimization.

We will discuss each item in Eq. (3) at the remainder of this section. The solution of Eq. (3) will be introduced in Sec. 4.

3.2. Coordination between Likelihood and Prior

In Wald protocol, the fusion module should recover LR-HSI from the degraded LR-HSI and HR-MSI. Based on Wald protocol, we can find:

$$\mathcal{X} = \mathcal{F}((\mathcal{X} * C)\downarrow_s, (\mathcal{Y} * C)\downarrow_s; \boldsymbol{\phi}) + \mathcal{N}_{\mathcal{X}}, \tag{4}$$

where $\mathcal{F}(\cdot)$ represents the fusion module. $\mathcal{N}_{\mathcal{X}}$ is a Gaussian noise with zero-mean, thus we have:

$$p(\boldsymbol{\phi}|\mathcal{X},\mathcal{Y},\boldsymbol{\theta}) = \prod_{i=1}^{h}\prod_{j=1}^{w}\mathcal{N}(\hat{\mathcal{X}}_{i,j}(\boldsymbol{\phi})|\boldsymbol{x}_{i,j}, \epsilon_1^2 I), \tag{5}$$

where $\hat{\mathcal{X}} = \mathcal{F}((\mathcal{X} * C)\downarrow_s, (\mathcal{Y} * C)\downarrow_s; \boldsymbol{\phi})$. Evidently, the estimated degradation parameters guide the optimization of the fusion module. Then, the guided fusion result $\mathcal{Z} = \mathcal{F}(\mathcal{X},\mathcal{Y};\boldsymbol{\phi})$ will facilitates the update of degradation parameters, i.e., maximum of $p(\mathcal{X},\mathcal{Y}|\mathcal{Z},\boldsymbol{\theta})$. In this way, the coordination optimization framework forms a virtuous circle between degradation estimation and fusion, leading to a robust HSI-SR. Meanwhile, the framework is orthogonal to the degradation estimation module and fusion module. Thus, it can be easily combined with well-designed modules for a further promotion on performance. Next, we will discuss these two modules.

3.3. Explicit Degradation Estimation

Here, we aim to capture the inherent pattern of degradations explicitly, expecting to a stable and precise estimation. Similar to the widely recognized degradation model

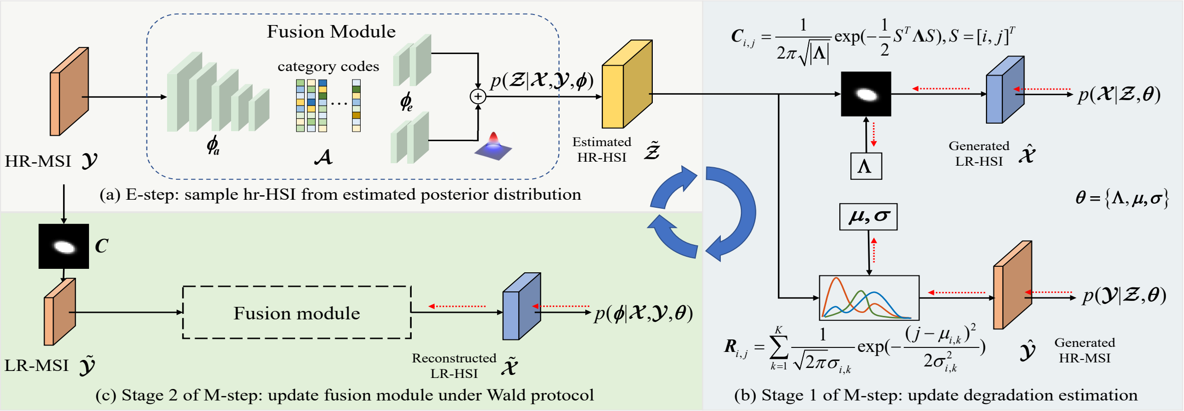

Fusion Module
category codes
$p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\phi)$
ϕ_e
ϕ_a
$\mathcal{A}$
HR-MSI $\mathcal{Y}$
Estimated HR-HSI $\tilde{\mathcal{Z}}$
(a) E-step: sample hr-HSI from estimated posterior distribution
$C_{i,j}=\frac{1}{2\pi\sqrt{|\Lambda|}}\exp(-\frac{1}{2}S^T\Lambda S), S=[i,j]^T$
$p(\mathcal{X}|\mathcal{Z},\theta)$
Λ
Generated LR-HSI $\hat{\mathcal{X}}$
$\theta=\{\Lambda,\mu,\sigma\}$
μ,σ
$p(\mathcal{Y}|\mathcal{Z},\theta)$
$R_{i,j}=\sum_{k=1}^{K}\frac{1}{\sqrt{2\pi}\sigma_{i,k}}\exp(-\frac{(j-\mu_{i,k})^2}{2\sigma_{i,k}^2})$
Generated HR-MSI $\hat{\mathcal{Y}}$
(b) Stage 1 of M-step: update degradation estimation
C
Fusion module
$p(\phi|\mathcal{X},\mathcal{Y},\theta)$
LR-MSI $\tilde{\mathcal{Y}}$
Reconstructed LR-HSI $\tilde{\mathcal{X}}$
(c) Stage 2 of M-step: update fusion module under Wald protocol

Figure 2. An overview of the proposed coordination optimization HSI-SR framework. Degradation estimation and fusion facilitate a positive feedback loop in the framework. As shown in (c), the estimated PSF is involved in the optimization of the fusion module as a guidance. Explicit modeling of PSF and SRF are built under anisotropic Gaussian kernel and mixture Gaussian for a stable degradation estimation. A lightweight autoencoder is constructed to recover the HR-HSI under the spectral mixing prior for an interpretable fusion. To solve the HSI-SR structure flexibly and efficiently, a Monte Carlo EM algorithm is applied with alternative optimization.

[36, 39], we simulate the PSF C with anisotropic Gaussian kernel:

$$\begin{aligned} C_{i,j} &= \frac{1}{2\pi}\sqrt{|\Lambda|}exp(-\frac{1}{2}S^T\Lambda S),\\ S &= [i,j]^T, i,j \in -s/2,\cdots,s/2, \end{aligned} \tag{6}$$

where $\Lambda \in \mathbb{R}^2$ is the covariance of the Gaussian blur kernel, and Λ is a symmetric positive definite matrix. Thus, only 3 parameters can handle the blur kernel[1].

Similarly, the SRF R lies in simple mode. We can estimate each column of R through mixture Gaussian with K components (only $2K$ parameters can realize estimation, which is evidently less than previous methods with $B \times b$ parameters at least[2]):

$$R_{i,j} = \sum_{k=1}^{K}\frac{1}{\sqrt{2\pi}\sigma_{j,k}}exp(-\frac{(i-\mu_{j,k})^2}{2\sigma_{j,k}^2}), \tag{7}$$

where $\mu_{j,k}$ and $\sigma_{j,k}$ are mean and variance of k-th Gaussian components in jth band of HR-MSI. With the estimated degradation parameters, the likelihood $p(\mathcal{X},\mathcal{Y}|\mathcal{Z},\boldsymbol{\theta})$ can be mathematically expressed as:

$$\begin{aligned} p(\mathcal{X},\mathcal{Y}|\mathcal{Z},\boldsymbol{\theta}) &= p(\mathcal{X}|\mathcal{Z},\boldsymbol{\theta})p(\mathcal{Y}|\mathcal{Z},\boldsymbol{\theta}),\\ p(\mathcal{X}|\mathcal{Z},\boldsymbol{\theta}) &= \prod_{i=1}^{H}\prod_{j=1}^{W}\mathcal{N}(\boldsymbol{x}_{i,j}|((\mathcal{Z}*C)\downarrow_s)_{i,j},\epsilon_2^2 I),\\ p(\mathcal{Y}|\mathcal{Z},\boldsymbol{\theta}) &= \prod_{i=1}^{H}\prod_{j=1}^{W}\mathcal{N}(\boldsymbol{y}_{i,j}|(\mathcal{Z}\times_3 R)_{i,j},\epsilon_3^2 I). \end{aligned} \tag{8}$$

In general, the spectral degradation of LR-HSI and the spatial degradation of HR-HSI stay close to each other, which is the so-called spatial-spectral consistency [22]. Therefore, $p(\boldsymbol{\theta}|\mathcal{X},\mathcal{Y})$ is in direct proportion to the similarity of these two degradations. Naturally, we model the posterior distribution of $\boldsymbol{\theta}$ as:

$$p(\boldsymbol{\theta}|\mathcal{X},\mathcal{Y}) \propto exp(||\mathcal{X}\times_3 R - (\mathcal{Y}*C)\downarrow_s||^2). \tag{9}$$

3.4. Interpretable and Lightweight Fusion

Two ingredients should be considered in fusion. Firstly, the fusion module should reflect the physical generation of HR-HSI as a prior knowledge for a clear interpretability. Secondly, considering the limited computation resource in real applications, the fusion module should be lightweight.

For the first point, we build a two stage model to simulate the spectral mixing phenomenon in HSIs imaging:

$$\begin{aligned} \mathcal{A} &= \mathcal{D}(\mathcal{Y};\phi_a),\\ \mathcal{Z} &= \mathcal{G}(\mathcal{A};\phi_e) = \boldsymbol{\mu}(\mathcal{A}) + \boldsymbol{\sigma}(\mathcal{A})\cdot\boldsymbol{\epsilon}, \boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},I), \end{aligned} \tag{10}$$

where $\mathcal{A}$ is the abundance of different materials in $\mathcal{Z}$. $\mathcal{D}(\cdot)$ is a classifier to detect the material category of each pixel,

[1] Direct simulation of PSF requires $s \times s$ parameters at least. In common, the scale factor s values as 32, means our method achieve 341× parameters reduction.

[2] Take CAVE dataset as example. In this situation, $b = 3$, $B = 31$ and we set $K = 4$, resulted in 11× parameter reduction at least.

essentially. $\mathcal{G}(\cdot)$ aims to map each hyperspectral vector from corresponding category of materials.

For the second point, we implement the fusion module with a tiny autoencoder. The encoder stacks five 1×1 convolution layers, which aims to predict the category of input pixel, while the subsequent decoder recover the HSIs from the predicted class information on each pixel. Because of the simple mode of spectrums in HSIs, we build the decoder with only two fully-connected layers.

4. Optimization

4.1. Training

For an efficient solution, we apply Monte Carlo Expectation Maximum (EM) algorithm on Eq. (3), which alternately updates parameters and desired distribution[3]. The overview of the proposed EM algorithm in training stage is presented in Fig. 2.

E-step

In this stage, we fix the parameters in Eq. (3) and search the optimal distribution of $\mathcal{Z}$. Following the classic EM algorithm, we set it as $p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi}_{old})$ (where $\boldsymbol{\phi}_{old}$ is the current fusion parameter). Thus, the corresponding evident lower bound (ELBO, equivalent to the original MAP problem) can be written as:

$$\mathcal{L}(\mathcal{Z},\boldsymbol{\phi},\boldsymbol{\theta};\mathcal{X},\mathcal{Y}) \propto \int p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi}_{old})\, log\, p(\mathcal{Z},\boldsymbol{\theta},\boldsymbol{\phi}|\mathcal{X},\mathcal{Y}) d\mathcal{Z}. \tag{11}$$

Then we simplify the calculation of ELBO through Monte Carlo approach. Through sampling the latent variable $\mathcal{Z}$ from $p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi}_{old})$, the ELBO is simplified to a accumulation:

$$\mathcal{L}(\mathcal{Z},\boldsymbol{\theta},\boldsymbol{\phi};\mathcal{X},\mathcal{Y}) \approx \sum_{i=1}^{N} logp(\mathcal{Z}_i,\boldsymbol{\theta},\boldsymbol{\phi}|\mathcal{X},\mathcal{Y}) + const, \tag{12}$$

where $\mathcal{Z}_i$ is the i-th sample from $p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi})$ and we set $i = 1$ similar to [40].

M-step M-step aims to maximize the approximated ELBO in E-step w.r.t. the parameters $\boldsymbol{\theta},\boldsymbol{\phi}$. We apply alternative optimization strategy to solve this problem.

Firstly, we fix $\boldsymbol{\phi}$ and update $\boldsymbol{\theta}$. According to Eq. (3), the object of optimization on $\boldsymbol{\theta}$ is:

$$\begin{aligned} & max_{\boldsymbol{\theta}}\, log\, p(\mathcal{X},\mathcal{Y}|\mathcal{Z},\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{X},\mathcal{Y}) \\ & \Longrightarrow min_{\boldsymbol{\theta}}\alpha_1||\mathcal{X} - (\mathcal{Z} * C)\downarrow_s||^2 + \alpha_2||\mathcal{Y} - \mathcal{Z}\times_3 R||^2 \\ & \quad + \alpha_3||\mathcal{X}\times_3 R - (\mathcal{Y} * C)\downarrow_s||^2. \end{aligned} \tag{13}$$

Algorithm 1: Fast fine-tune of the proposed method.

```
Input: observed LR-HSI $\mathcal{X}$, HR-MSI $\mathcal{Y}$,
       pre-trained $\boldsymbol{\theta}$, $\boldsymbol{\phi}$
Output: the estimated HR-HSI $\mathcal{Z}$
while not converged do
    E-step: Sample the latent variable/HR-HSI $\mathcal{Z}$
      from $p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi})$ following Eq. (10);
    M-step: Update parameter $\boldsymbol{\phi}_e$ with fixed $\boldsymbol{\theta}$ and
      $\boldsymbol{\phi}_a$ based on Eq. (15).
end
$\mathcal{Z} = \mathcal{G}(\mathcal{D}(\mathcal{Y};\boldsymbol{\phi}_a);\boldsymbol{\phi}_e)$.
```

Secondly, we fix $\boldsymbol{\theta}$ and update $\boldsymbol{\phi}$:

$$\begin{aligned} & max_{\boldsymbol{\phi}}\, log\, p(\boldsymbol{\phi}|\mathcal{X},\mathcal{Y},\boldsymbol{\theta}) \\ & \quad \Longrightarrow min_{\boldsymbol{\theta}}\, \beta||\mathcal{X} - \mathcal{G}(\mathcal{D}((\mathcal{Y} * C^{new})\downarrow_s;\boldsymbol{\phi}_{\boldsymbol{a}});\boldsymbol{\phi}_{\boldsymbol{e}})||^2. \end{aligned} \tag{14}$$

We use $\alpha_1, \alpha_2, \alpha_3, \beta$ to trade-off the efforts of different terms[4]. Note that we omit $p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi})$, beacuse the prior knowledge of HR-HSI is contained in the structure of the fusion module.

At the end of optimization, we can estimate the desired HR-HSI through sampling from $p(\mathcal{Z}|\mathcal{X},\mathcal{Y},\boldsymbol{\phi})$, i.e., $\mathcal{Z} = \mathcal{G}(\mathcal{D}(\mathcal{Y};\boldsymbol{\phi}_a);\boldsymbol{\phi}_e)$.

4.2. Fast Fine-tuning

Different from nature images, hyperspectral images vary in different scene dramaticly. Thus, fine-tuning in the processing images is necessary [21–23]. Considering the circumstance in real scene, the fine-tune stage should be efficient in time and computational consumption. We develop a fast fine-tune strategy to satisfy above requirements. The overview of the proposed EM algorithm in testing stage is presented in Algorithm 1.

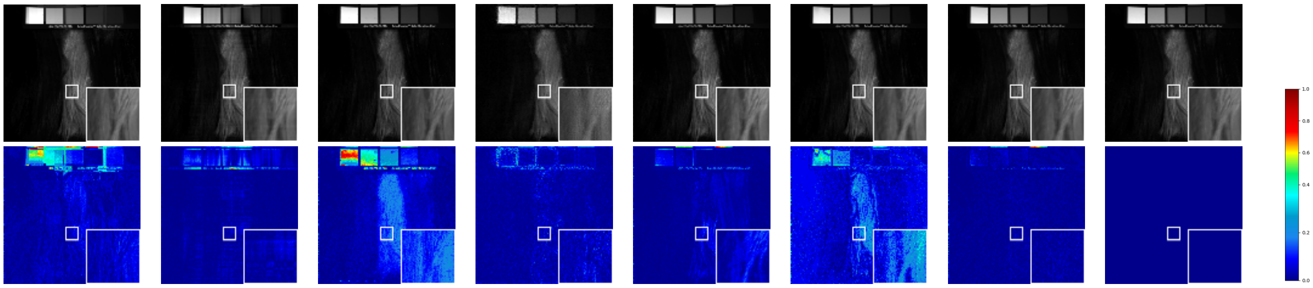
With the learned degradation from the training set, it's unnecessary to update degradation adaption in fine-tuning[5]. We only adapt the fusion module to input images. The two submodules of the fusion module play different roles. As a classifier, the abundance estimator $\mathcal{D}(\cdot)$ handles the relative differences in spectrums and is not necessary to be updated in fine-tune stage. We only optimize the spectral mapping block $\mathcal{G}(\cdot)$:

$$min_{\boldsymbol{\phi}_z}||\mathcal{X} - \mathcal{G}(\tilde{\mathcal{A}};\boldsymbol{\phi}_e)||^2, \tag{15}$$

[3] Analysis on convergence is presented in the Supplemental Materials.

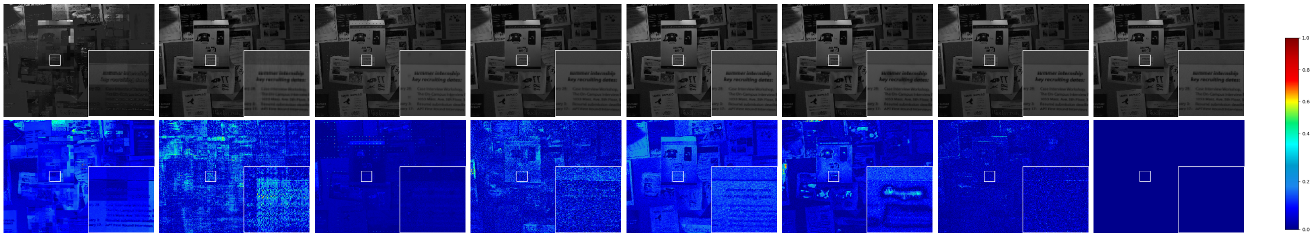
[4] Parameters analysis can be found in the Supplemental Materials.

[5] We assume consistent degradation in training and testing samples. For inconsistent degradation (i.e., the PSF and SRF of imaging sensors are changed), we can load batch of image-pairs to re-train the network with low cost. For dramatic degradation changing where degradation varies in different images, we simulate this circumstance and verify the efficiency of our method, which can be found in the Supplementary Materials.



(a) CNMF [17] (b) CSTF [26] (c) HySure [18] (d) MHFNet [39] (e) CUCaNet [22] (f) UAL [23] (g) Ours (h) Ground truth

Figure 3. Visual SR results and the corresponding error images on scene of $glass\ tiles$ in CAVE dataset under the degradation of $PSF5$, where we display the 22th (620nm) band of the HR-HSI images.



(a) CNMF [17] (b) CSTF [26] (c) HySure [18] (d) MHFNet [39] (e) CUCaNet [22] (f) UAL [23] (g) Ours (h) Ground truth

Figure 4. Visual SR results and the corresponding error images of $imgf27$ in Harvard dataset under degradation of $PSF5$, where we display the 15th (480nm) band of the HR-HSI images.

where $\tilde{\mathcal{A}} = \mathcal{D}((\mathcal{Y} * C) \downarrow_s; \phi_a)$ and is fixed in fine-tuning stage.

5. Experimental Results

5.1. Experiment Setup

Implementation Details

We implement our method under Pytorch with a single 3090 GPU. We apply the ADAM optimizer in the training stage with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\eta = 10^{-8}$. The learning rate in the training and fine-tuning stage is initialized to 10^{-3} and 5×10^{-3}, respectively. We adopt 100 epochs in the training stage and 250 in the testing stage. The detailed network structure varies in the different datasets with different bands of HSIs, and we will list the clear network structure in the Supplementary Materials.

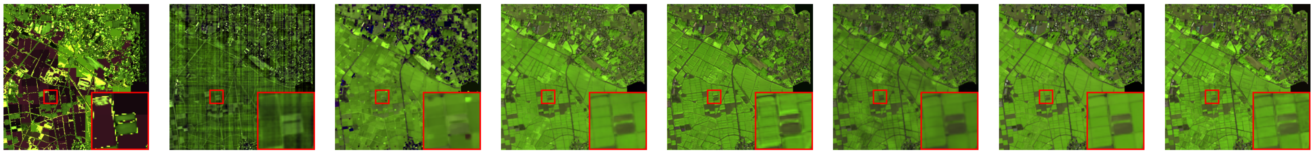
Compared Methods

For comprehensive comparison, we select three numerical methods (including CNMF [17][6], Hysure [18][7], CSTF [26][8]), state-of-the-art supervised learning approach, (MHFNet [39][9]), two unsupervised learning approaches (including CUCaNet [22][10], and UAL [23][11]). All of the numerical methods require the PSF C for a precise recovery. We provide the groundtruth degradation parameters for these methods in synthetic datasets.

Benchmarks

We adopt three synthetic HSI benchmarks and one real HSI benchmark for evaluation, including CAVE [41][12], Harvard [42][13], Chikusei [43][14], and Worldview2[15]. The CAVE dataset contains 32 HSIs imaging from 32 different indoor scenes. Each HSI has 512 × 512 pixels and 31 spectral bands measured in the wavelength ranging from 400nm to 700nm. The Harvard dataset consists of 50 HSIs imaged outdoors. Each HSI has 1040 × 1392 pixels and 31 spectral bands measured in the wavelength ranging from 420nm to 720nm. For the experiment, we select the left top of each image with 1024 × 1376 pixels. Chikusei is an airborne hyperspectral dataset with 2517 × 2335 pixels and 128 bands in the spectral range from 363nm to 1018nm. We select the right bottom of Chikusei with 2048 × 2048 pixels and crop it to 16 non-overlapped images with a size of 512 × 512 for the experiment. Worldview2 is a real-world dataset with a

[6] https://naotoyokoya.com/Download.html
[7] https://github.com/alfaiate/HySure
[8] https://github.com/renweidian/CSTF
[9] https://github.com/XieQi2015/MHF-net
[10] https://github.com/danfenghong/ECCV2020_CUCaNet
[11] https://github.com/JiangtaoNie/UAL
[12] http://www.cs.columbia.edu/CAVE/databases/
[13] http://vision.seas.harvard.edu/hyperspec/explore.html
[14] https://naotoyokoya.com/Download.html
[15] https://www.l3harrisgeospatial.com/Data-Imagery/Satellite-Imagery/High-Resolution/WorldView-2



(a) CNMF [17] (b) CSTF [26] (c) HySure [18] (d) MHFNet [39] (e) CUCaNet [22] (f) UAL [23] (g) Ours (h) HR-HSI

Figure 5. Visual SR results and the corresponding error images on Chikusei dataset, where we display the test images with bands 70-100-36 as R-G-B to show.

Table 1. Average performance of test methods on three synthetic datasets under six random generated degradations.

Datasets	CAVE				Harvard				Chikusei			
Methods	PSNR	SAM	ERGAS	UQI	PSNR	SAM	ERGAS	UQI	PSNR	SAM	ERGAS	UQI
CNMF	29.3	17.8	2.16	0.702	25.7	13.3	1.76	0.580	24.5	17.8	1.38	0.721
Hysure	28.0	24.1	1.27	0.789	35.2	7.01	0.652	0.908	21.8	10.3	2.67	0.644
CSTF	29.5	14.8	1.03	0.817	34.7	7.16	0.640	0.923	25.7	11.8	1.56	0.848
MHFNet	37.3	9.57	0.693	0.928	41.1	3.57	0.312	0.988	33.6	7.07	0.638	0.951
CUCaNet	34.3	13.6	0.964	0.963	37.3	6.10	1.42	0.936	27.3	5.82	0.753	0.890
UAL	37.2	9.83	0.785	0.912	40.7	7.11	0.824	0.964	28.9	7.55	1.249	0.888
Ours	**39.5**	**6.74**	**0.340**	**0.944**	**42.4**	**3.20**	**0.276**	**0.995**	**33.7**	**3.77**	**0.591**	**0.980**

LR-HSI of size 419 × 658 × 8 and a HR-MSI of size 1676 × 2632 × 3, while the HR-HSI is unavailable.

We adopt four picture quality indices (PQIs) for quantitative evaluation, including peak signal-to-noise ratio (PSNR), spectral angle mapper (SAM [44]), relative dimensionless global error in synthesis (ERGAS [38]), and universal image quality index (UIQI [45]).

5.2. Experiments on Synthetic Benchmarks

In CAVE and Harvard, we conduct the testing sets with the last 16 and 25 images, respectively, and the remainding images are set as training samples for compared supervised methods. We randomly select 8 cropped images in Chikusei as testing samples, and the others are set as training samples. We apply six randomly generated Gaussian blur kernels as PSF to synthesize LR-HSIs for each dataset. The generated bur kernels are visualized in Fig. 6. Following the settings of [39], we apply SRF of Landset8 to generate HR-MSIs of Chikusei, Nikon 700 for HR-MSIs of CAVE and Harvard. There are 6 versions for each dataset corresponding to 6 blur kernels. As shown in Table 1, the proposed method outperforms other comparsion methods. Fig. 3 depicts the visual performance of different methods in CAVE [41] dataset. As can be observed that our method surpasses other methods in detailed content and global contexts. The same conclusion holds for other datasets, as shown in Fig. 4 and Fig. 5.

5.3. Real Scenario Experiments

Since the HR-HSI in Worldview2 is unavailable, we apply the Wald protocol to build the training set for supervised methods. Both LR-HSI and HR-MSI are downsampled spatially by a factor of 4, and the original LR-HSI is used as ground truth HR-HSI in the training set. As shown in Fig. 7, visual inspection indicates that our method holds a more clear texture than comparison methods.

5.4. Algorithmic Analysis

Explicit Degradation Estimation

Fig. 6 depicts the estimated blur kernels by our method. Obviously, our explicit degradation estimation method accurately captures the degradation characteristics.

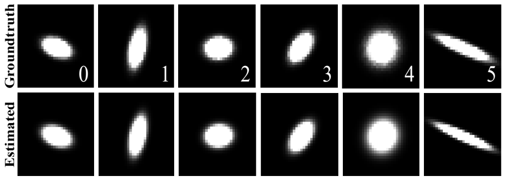


Figure 6. Random generated anisotropic Gaussian blur kernels and the corresponding estimated blur kernels by our method.

As for the SRF, DL-based estimation lose the essential features of SRF. Evidently, the superfluous learning capability causes over-fitting. By contrast, our explicit estimation method achieve a more precise estimation which captures the peak in each band, as shown in Fig. 8.

Computational Cost Analysis

Further, we report the FLOPs, model size, training time, and testing time of competing methods in Table 2. Evidently, our method surpasses other methods in lightweight and fast processing.

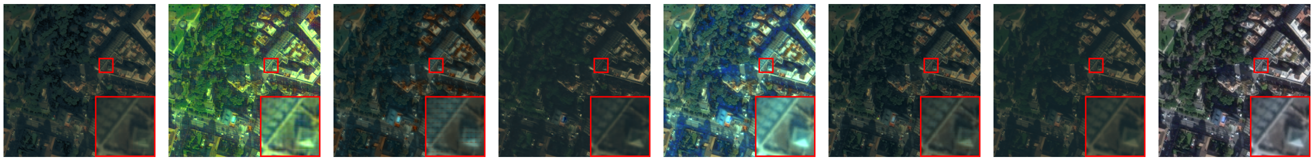

(a) CNMF [17] (b) CSTF [26] (c) HySure [18] (d) MHFNet [39] (e) CUCaNet [22] (f) UAL [23] (g) Ours (h) HR-MSI

Figure 7. Visual SR results in WV2 dataset. We show the images with bands 5-3-2 as R-G-B.

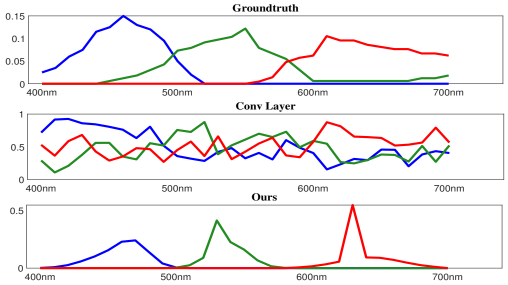


Figure 8. Groundtruth SRF and the estimated SRFs by DL toolkits and our explicit estimation strategy.

Table 2. Computational and time consumption on competing methods.

Methods	FLOPs (G)	Param (M)	Training Time (min)	Testing Time (s)
CNMF	/	/	/	101.2
CSTF	/	/	/	123.8
HySure	/	/	/	153.8
MHFNet	21226	129.4	1255	23.95
CUCaNet	7226	2.28	/	34.2(min)
UAL	9275	27.05	351	29.7
Ours	**1.67**	**0.019**	**86.8**	**20.5**

Table 3. Effort of each component in the proposed method.

Methods	PSNR	SAM	Train Time (min)	Test Time (s)
w/o coord	32.56	7.717	86.8	20.5
w/o DE	32.72	7.713	83.1	18.7
Conv DE	31.42	8.320	89.5	21.0
w/o fine-tune	27.93	13.61	86.8	0.235
w/o pre-train	30.97	11.57	0	20.5
w/o fix enc	35.79	7.198	86.8	28.2
Ours	39.48	6.825	86.8	20.5

Ablation Study

We analyze three critical components in our method: (1) coordination optimization between prior and likelihood, (2) the explicit degradation estimation, and (3) fast fine-tune strategy. To verify the effort of the **coordination optimization**, we break the feedback circle between likelihood and prior through replacing the estimated PSF in the fusion module with fixed PSF (isotropic Gaussian kernel with variance $[0.1, 0.1]$), represented as 'w/o coord'. The dropped performance in Table 3 confirms the effort of coordination strategy. For the **explicit degradation estimation**, we conduct two variances that remove degradation estimation and replace explicit estimation with convolution layers, termed 'w/o DE' and 'Conv DE' , respectively. As shown in Table 3, their performance is inferior to the original setting, which verifies the effort of degradation estimation and explicit estimation. As for the **fast fine-tune strategy**, we build three variances in which product fusion results without fine-tuning, adapt the whole network only on tested samples without pre-train, and optimize the entire network in fine-tune stage, named 'w/o fine-tune', 'w/o pre-train', and 'w/o fix enc', respectively. In Table 3, we can confirm the effort of each component in the corresponding experiment. The setting 'w/o fix enc' shows more time consumption and lower metrics resulted from the abundant learnable parameters and over-fitting. All of the variances are tested in CAVE dataset under the degradation of $PSF5$.

6. Conclusion

In this paper, we developed a coordination optimization HSI-SR framework which established a virtous cycle between degradation estimation and fusion. Under the framework, we further explored an explicit degradation estimation strategy with stable performance. Meanwhile, we established a tiny autoencoder under spectral mixing prior for interpretable and lightweight fusion. Six different PSFs were generated to verify the efficiency and stability of our method under unknown scenarios with various degradations. Comprehensive experiments show that our framework outperforms SOTA methods by a large margin among accuracy metrics and computational cost. We leave two research direction as our future work, including implementation on real-world devices and extension to more tasks in image enhancement.

Acknowledgement. This research was supported by the China NSFC projects 62121001, U22B2014, and the Young Elite Scientist Sponsorship Program by the China Association for Science and Technology 2020QNRC001.

References

[1] Jose M. Bioucas-Dias, Antonio Plaza, Gustavo Camps-Valls, Paul Scheunders, Nasser Nasrabadi, and Jocelyn Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):6–36, 2013. 1

[2] Lujendra Ojha, Mary Beth Wilhelm, Scott L. Murchie, Alfred S. McEwen, James J. Wray, Jennifer Hanley, Marion Massé, and Matt Chojnacki. Spectral evidence for hydrated salts in recurring slope lineae on Mars. *Nature Geoscience*, 8(11):829–832, November 2015. 1

[3] Ying Fu, Antony Lam, Yasuyuki Kobashi, Imari Sato, Takahiro Okabe, and Yoichi Sato. Reflectance and fluorescent spectra recovery based on fluorescent chromaticity invariance under varying illumination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014. 1

[4] Ying Fu, Antony Lam, Imari Sato, Takahiro Okabe, and Yoichi Sato. Separating reflective and fluorescent components using high frequency illumination in the spectral domain. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2013. 1

[5] Bing Lu, Phuong D. Dao, Jiangui Liu, Yuhong He, and Jiali Shang. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing*, 12(16), 2020. 1

[6] A-K Mahlein, Matheus T Kuska, Jan Behmann, Gerrit Polder, and Achim Walter. Hyperspectral sensors and imaging technologies in phytopathology: state of the art. *Annual review of phytopathology*, 56:535–558, 2018. 1

[7] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. *Journal of biomedical optics*, 19(1):010901, 2014. 1

[8] Pedram Ghamisi, Naoto Yokoya, Jun Li, Wenzhi Liao, Sicong Liu, Javier Plaza, Behnood Rasti, and Antonio Plaza. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):37–78, 2017. 1

[9] Shutao Li, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and Jón Atli Benediktsson. Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6690–6709, 2019. 1

[10] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017. 1

[11] Hien Van Nguyen, Amit Banerjee, and Rama Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 44–51. IEEE, 2010. 1

[12] Yuliya Tarabalka, Jocelyn Chanussot, and Jón Atli Benediktsson. Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(5):1267–1279, 2009. 1

[13] Qi Wang, Jianzhe Lin, and Yuan Yuan. Salient band selection for hyperspectral image classification via manifold ranking. *IEEE transactions on neural networks and learning systems*, 27(6):1279–1289, 2016. 1

[14] Burak Uzkent, Matthew J Hoffman, and Anthony Vodacek. Real-time vehicle tracking in aerial video using hyperspectral features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 36–44, 2016. 1

[15] Burak Uzkent, Aneesh Rangnekar, and Matthew Hoffman. Aerial vehicle tracking by adaptive fusion of hyperspectral likelihood maps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 39–48, 2017. 1

[16] Yunsong Li, Weiying Xie, and Huaqing Li. Hyperspectral image reconstruction by deep convolutional neural network for classification. *Pattern Recognition*, 63:371–383, 2017. 1

[17] Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):528–537, 2012. 2, 6, 7, 8

[18] Miguel Simões, José Bioucas-Dias, Luis B. Almeida, and Jocelyn Chanussot. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3373–3388, 2015. 2, 6, 7, 8

[19] Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. Fast fusion of multi-band images based on solving a sylvester equation. *IEEE Transactions on Image Processing*, 24(11):4109–4121, 2015. 2

[20] Weisheng Dong, Chen Zhou, Fangfang Wu, Jinjian Wu, Guangming Shi, and Xin Li. Model-guided deep hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 30:5754–5768, 2021. 2

[21] Ying Qu, Hairong Qi, and Chiman Kwan. Unsupervised sparse dirichlet-net for hyperspectral image super-resolution. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2511–2520, 2018. 2, 3, 5

[22] Jing Yao, Danfeng Hong, Jocelyn Chanussot, Deyu Meng, Xiaoxiang Zhu, and Zongben Xu. Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution. In *ECCV 2020*, pages 208–224, 2020. 2, 3, 4, 5, 6, 7, 8

[23] L. Zhang, J. Nie, W. Wei, Y. Zhang, and L. Shao. Unsupervised adaptation learning for hyperspectral imagery super-resolution. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2, 3, 5, 6, 7, 8

[24] Renwei Dian, Shutao Li, and Leyuan Fang. Learning a low tensor-train rank representation for hyperspectral image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9):2672–2683, 2019. 2

[25] Renwei Dian and Shutao Li. Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization. *IEEE Transactions on Image Processing*, 28(10):5135–5146, 2019. 2

[26] Shutao Li, Renwei Dian, Leyuan Fang, and José M. Bioucas-Dias. Fusing hyperspectral and multispectral images via coupled sparse tensor factorization. *IEEE Transactions on Image Processing*, 27(8):4118–4130, 2018. 2, 6, 7, 8

[27] Naveed Akhtar, Faisal Shafait, and Ajmal Mian. Sparse spatio-spectral representation for hyperspectral image super-resolution. In *European conference on computer vision*, pages 63–78. Springer, 2014. 2

[28] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer, 2016. 2

[29] Renwei Dian, Leyuan Fang, and Shutao Li. Hyperspectral image super-resolution via non-local sparse tensor factorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5344–5353, 2017. 2

[30] Weisheng Dong, Fazuo Fu, Guangming Shi, Xun Cao, Jinjian Wu, Guangyu Li, and Xin Li. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Transactions on Image Processing*, 25(5):2337–2352, 2016. 2

[31] Naveed Akhtar, Faisal Shafait, and Ajmal Mian. Bayesian sparse representation for hyperspectral image super resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3631–3640, 2015. 2

[32] Weisheng Dong, Fazuo Fu, Guangming Shi, Xun Cao, Jinjian Wu, Guangyu Li, and Xin Li. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Transactions on Image Processing*, 25(5):2337–2352, May 2016. 2

[33] Ke Zheng, Lianru Gao, Wenzhi Liao, Danfeng Hong, Bing Zhang, Ximin Cui, and Jocelyn Chanussot. Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3):2487–2502, 2021. 2, 3

[34] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10581–10590, 2021. 3

[35] Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang, and Tieniu Tan. Learning the degradation distribution for blind image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6063–6072, 2022. 3

[36] Zongsheng Yue, Qian Zhao, Jianwen Xie, Lei Zhang, Deyu Meng, and Kwan-Yee K. Wong. Blind image super-resolution with elaborate degradation modeling on noise and kernel. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2118–2128, 2022. 3, 4

[37] Ying Fu, Tao Zhang, Yinqiang Zheng, Debing Zhang, and Hua Huang. Hyperspectral image super-resolution with optimized rgb guidance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11661–11670, 2019. 3

[38] Lucien Wald. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002. 3, 7

[39] Qi Xie, Minghao Zhou, Qian Zhao, Zongben Xu, and Deyu Meng. Mhf-net: An interpretable deep network for multispectral and hyperspectral image fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3):1457–1473, 2022. 4, 6, 7, 8

[40] D. P. Kingma and M. Welling. Auto-encoding variational bayes. In *ICLR*, 2014. 5

[41] D. Iso F. Yasuma, T. Mitsunaga and S.K. Nayar. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Technical report, Nov 2008. 6, 7

[42] A. Chakrabarti and T. Zickler. Statistics of Real-World Hyperspectral Images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 193–200, 2011. 6

[43] Naoto Yokoya, Claas Grohnfeldt, and Jocelyn Chanussot. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 5(2):29–56, 2017. 6

[44] Roberta H Yuhas, Joseph W Boardman, and Alexander FH Goetz. Determination of semi-arid landscape endmembers and seasonal trends using convex geometry spectral unmixing techniques. In *JPL, Summaries of the 4th Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop*, 1993. 7

[45] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE signal processing letters*, 9(3):81–84, 2002. 7