sensors

MDPI

Article

Deep Unsupervised Fusion Learning for Hyperspectral Image Super Resolution

Zhe Liu [1], Yinqiang Zheng [2] and Xian-Hua Han [1,*]

[1] Graduate School of Science and Technology for Innovation, Yamaguchi University, Yamaguchi 753-8511, Japan; a501wbu@yamaguchi-u.ac.jp
[2] National Institute of Informatics, Tokyo 101-8430, Japan; yqzheng@nii.ac.jp
* Correspondence: hanxhua@yamaguchi-u.ac.jp

Abstract: Hyperspectral image (HSI) super-resolution (SR) is a challenging task due to its ill-posed nature, and has attracted extensive attention by the research community. Previous methods concentrated on leveraging various hand-crafted image priors of a latent high-resolution hyperspectral (HR-HS) image to regularize the degradation model of the observed low-resolution hyperspectral (LR-HS) and HR-RGB images. Different optimization strategies for searching a plausible solution, which usually leads to a limited reconstruction performance, were also exploited. Recently, deep-learning-based methods evolved for automatically learning the abundant image priors in a latent HR-HS image. These methods have made great progress for HS image super resolution. Current deep-learning methods have faced difficulties in designing more complicated and deeper neural network architectures for boosting the performance. They also require large-scale training triplets, such as the LR-HS, HR-RGB, and their corresponding HR-HS images for neural network training. These training triplets significantly limit their applicability to real scenarios. In this work, a deep unsupervised fusion-learning framework for generating a latent HR-HS image using only the observed LR-HS and HR-RGB images without previous preparation of any other training triplets is proposed. Based on the fact that a convolutional neural network architecture is capable of capturing a large number of low-level statistics (priors) of images, the automatic learning of underlying priors of spatial structures and spectral attributes in a latent HR-HS image using only its corresponding degraded observations is promoted. Specifically, the parameter space of a generative neural network used for learning the required HR-HS image to minimize the reconstruction errors of the observations using mathematical relations between data is investigated. Moreover, special convolutional layers for approximating the degradation operations between observations and the latent HR-HS image are specifically to construct an end-to-end unsupervised learning framework for HS image super-resolution. Experiments on two benchmark HS datasets, including the CAVE and Harvard, demonstrate that the proposed method can is capable of producing very promising results, even under a large upscaling factor. Furthermore, it can outperform other unsupervised state-of-the-art methods by a large margin, and manifests its superiority and efficiency.

Keywords: hyperspectral image; super-resolution; unsupervised fusion learning; image priors

check for updates

Citation: Liu, Z.; Zheng, Y.; Han, X.-H. Deep Unsupervised Fusion Learning for Hyperspectral Image Super Resolution. *Sensors* **2021**, *21*, 2348. https://doi.org/10.3390/s21072348

Academic Editor: Pedram Ghamisi

Received: 5 February 2021
Accepted: 25 March 2021
Published: 28 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In hyperspectral (HS) imaging, three-dimensional cubic data with decades or hundreds of wavelength bands are captured. Each spatial point (pixel) contains a high dimensional vector for recording the light intensity at different wavelengths. Thus, the images acquired using the HS imaging technology contain not only abundant spatial structures but also detailed spectral signature and well suited to the substantial and high-performance analysis of imaged scenes. Having the advantages of detailed spectral distribution, the HS images were successfully used in various applications, such as remote sensing [1], food inspection [2–4], image classification [5–7] and object detection [8–10], and

medicine [11–13], and are capable of achieving high-performance gain compared with other common RGB images. However, due to the radiant collection for each narrow-spectrum band in HS imaging sensors, less radiant energy per pixel and per spectral band measurement of an image scene is anticipated compared with the RGB imaging sensors. To ensure sufficient signal-to-noise ratio, the photo collection must be conducted in a much larger spatial region. This implies that the spatial resolution must be sacrificed to obtain detailed spectral information. Therefore, there is a trade-off between spatial and spectral resolution in real imaging sensors. This means that an HS sensor usually captures low spatial resolution and detailed spectral distribution (high spectral resolution) images. In contrast, common RGB sensors can provide much higher spatial resolution images but only with RGB color information. There are still some difficulties in acquiring high-resolution data in both spatial and spectral domains from commercial imaging sensors. Therefore, extensive research is necessary to fuse a low-resolution HS image (LR-HS) with the corresponding HR-RGB (multispectral) image for generating an HR-HS image using image processing and machine learning techniques. These fusion methods for generating HR-HS images are in general referred to as hyperspectral image super-resolution (HSI SR) methods [14].

Due to its ill-posed nature (the unknown number of variables in a latent HR-HS image is much larger than the known number of variables in the observations), multispectral and hyperspectral image fusion is a very challenging task. Most previously reported methods generally leverage various hand-crafted image priors to regularize the mathematical degradation model between observations and the latent HR-HS image. They also explore different optimization strategies to achieve the optimal solution. Based on the physical properties of the observed spectrum in HS images, one research direction is the investigation of effective representation of high-dimensional spectral vectors such as matrix factorization and spectral unmixing [15]. The spectral representation models were evolved from the fact that HS observations can generally be expressed as a weighted linear combination of the reflectance function basis and their corresponding fraction coefficients [16]. These models achieve acceptable reconstruction performance. Subsequently, based on the significant success of sparse representation in natural image processing, several research studies imposed a sparsity constraint on the spectral representation [17] as prior knowledge, and attempted to model the spatial structure and local spectral characteristics by automatically learning the spectral dictionary from the observed HR-RGB and LR-HS images. Additionally, based on possible low-dimensionality of the spectral space, a low-rank image prior technique was also adopted for exploring the intrinsic spectral correlation of a latent HR-HS image. This technique proved capable of reducing the spectral distortion to some extent [18]. Moreover, some recent research studies extensively exploited the similarity between image priors of global spatial structures and local spectral structures to further boost reconstruction performance [19,20]. Although the integration of various hand-crafted image priors, such as physical spectral mixing, mathematical sparsity of spectral representation, low-rank property, and similarity resulted in significant progress regarding the HSI SR performance gain, discovering an optimal image prior for a specific scene is still an extremely difficult task due to the configuration and texture diversity in both the spatial and spectral domains.

The successful application of deep convolutional neural networks (DCNN) in various computer vision tasks allowed HSI SR to rely on the DCNN's powerful learning capabilities [21] for robustly reconstructing a latent HR-HS image, and demonstrated impressive performance using various neural network architectures [22]. In contrast to the traditional optimization methods, the deep learning (DL) method is capable of automatically learning the underlying image priors in a latent HR-HS image instead of exploiting hand-crafted image priors. Using previously collected training samples consisting of LR-HS, HR-RGB images, and their corresponding HR-HS images, an HS image prediction model in the training phase can be constructed, and the corresponding HR-HS image can be efficiently estimated from its low-quality observations. However, most of the current DL methods are implemented in a fully supervised manner. For optimal network parameter learning,

large-scale training triplets must be collected in advance. However, this is a difficult task, especially in the HSI SR scenario due to the high-cost of capturing the HR-HS images. Moreover, the fully supervised DL paradigm usually suffers from insufficient generalization in real applications and separate HS image prediction models for different HS datasets must be learned. More recently, Ulyanov et al. [23] proposed a deep image prior (DIP) learning neural network and stated that a convolutional neural network itself is capable of capturing a large number of low-level image statistics for well reconstructing a natural image, which can be successfully applied to several image restoration tasks. Sidorov et al. [24] extended the idea of DIP into deep hyperspectral prior (DHP), and adopted denoising, inpainting and super-resolution for hyperspectral images. However, the DHPs only leverage the observed LR-HS image for network training and cannot efficiently learn both the spatial structure and spectral attribute priors for reconstructing a latent HR-HS image.

To deal with the limitations of current DL-based methods, a novel deep unsupervised fusion-learning (DUFL) framework for HSI SR is proposed in this work. Specifically, a generative neural network capable of automatically learning the spatial and spectral image priors in a latent HR-HS image is exploited. Instead of directly searching the target (HR-HS image) space, we attempt to determine the optimal parameters by searching the neural network parameter space to minimize the reconstruction errors of the observed HR-RGB and LR-HS images using a degradation procedure. For a robust neural network training to avoid falling into a local minimum state, the fixed initialized noisy input is perturbed with a small randomly generated noise at each training step, and the HR-HS image is predicted by employing the deep-learned network and the fixed noisy input. Moreover, special convolutional layers are designed for implementing the blurring/down-sampling operation and spectral transformation in the degradation model. By employing the deep-learned spatial and spectral image priors, the proposed DUFL method is capable of efficiently recovering the underlying spatial structure and spectral attribute of a latent HR-HS image using only the observed HR- RGB and LR-HS images. Also, massive triplets of the HR-RGB, LR-HS, and HR-HS images are not required. Experiments on two benchmark HS datasets demonstrate that the proposed framework is capable of achieving superior HS image reconstruction performance compared with most state-of-the-art (SoTA) methods in terms of both quantitative evaluation and visual effect. In summary, our main contributions are three-fold and summarized as follows:

1. Propose a novel deep unsupervised fusion-learning framework, which can effectively learn the potential HR-HS image from the observed LR-HS and HR-RGB images only, for HSI SR problem;
2. Leverage to search the parameter space of an encoder/decoder generative neural network instead of the raw target (HR-HS image) space to well reconstruct a latent HR-HS image, which has the prospects to learn both spatial structure and spectral attribute image priors;
3. Construct an end-to-end unsupervised fusion-learning framework for the HSI SR problem by designing special convolutional layers for implementing the mathematical degradation model between the latent HR-HS image and the observations.

The rest of this paper is organized as follows. The related research work, including traditional optimization methods and deep-learning-based methods, is reviewed in Section 2. The proposed deep unsupervised fusion-learning framework is introduced in Section 3. Extensive experiments and comparison of results with two benchmark datasets are presented in Section 4. The concluding remarks are summarized in Section 5.

2. Related Research Work

In recent years, hyperspectral image reconstruction has been actively investigated in the computer vision and computational photography research community, and substantial improvement has been achieved. This work mainly focuses on hyperspectral image super resolution (HSI SR) by fusing the available LR-HS and HR-RGB images obtained from commercial imaging sensors. In this section, the related research work is briefly reviewed.

2.1. Traditional Optimization Method

The HSI SR problem, following the fusion paradigm of the observed LR-HS and HR-RGB images (fusion-based HSI SR), is closely related to a multi-spectral (MS) image pan-sharpening task, where the goal is to merge an LR-MS image with its corresponding HR wide-band panchromatic image [25]. Numerous methods for MS pan-sharpening, mainly including multi-resolution analysis approaches [26] and component substitution methods [27], were proposed. The fusion-based HSI SR problem can be treated as several pan-sharpening sub-problems, where each band of the HR-MS (RGB) image can be considered to be a panchromatic image. However, this simplification cannot fully use the spectral correlation and usually leads to significant spectral distortion in the recovered HR-HS image.

Recently, existing fusion-based HSI SR methods have widely leveraged the hand-crafted image priors in a latent HR-HS image for robustly solving the inverse optimization problem. The investigated image priors play a key role in obtaining a plausible solution in the optimization problem. The popularly used image priors are mainly used to explore the hidden knowledge in spatial and spectral representation such as physical spectral mixing, sparsity, low-rank, and similarity [28]. Yokoya et al. [15] proposed to decompose the latent HR-HS image into a non-negative end-member matrix and an abundance matrix called negative matrix factorization (NMF). Then, they exploited a coupled NMF version (CNMF) to fuse a pair of HR-MS and LR-HS images, whereas Lanaras et al. [29] used a proximal alternating linearized-minimization method to optimize the coupled spectral unmixing model for HSI SR. Subsequently, the sparsity reguralized decomposition was extensively investigated by imposing sparse constraints on the abundance matrix [30]. Akhtar et al. [31] proposed a Bayesian sparse representation scheme, aiming to infer the underlying probability distributions of spectra and their proportions, which were decomposed from the latent HR-HS image. Grohnfeldt et al. [32] employed a joint sparse representation for separately modeling the spatial structure (local patch) in each individual band image. Based on the inherent low-dimensionality of the spectral space and the 3D structure in a latent HR-HS image, tensor factorization and low-rank image priors were actively integrated for the HSI SR problem [33], and the feasibility of the reconstructed HR-HS image was demonstrated. More recently, Dong et al. [34] considered the spatially non-local similarity of a latent HSI and proposed a non-negative structured sparse representation (NSSR) method. Han et al. [35] further integrated both the global spatial and local spectral similarity to boost reconstruction performance. Although these hand-crafted image prior algorithms demonstrated impressive performance, finding a suitable image prior for a specific scene is still a challenging task.

2.2. Deep Learning Based Methods

Based on the successful application of deep convolutional neural network in nature RGB image super-resolution, deep learning was also investigated for fusion-based HSI SR tasks. Instead of exploiting the hand-crafted image priors, in this method, the inherent image priors hidden in a latent HR-HS image are automatically learned, and a superior reconstruction performance can be achieved. The current fusion-based HSI SR employing deep learning is mainly divided into the fully supervised learning method and the unsupervised learning method.

2.2.1. Fully Supervised Learning Method

To train the HS image prediction model, it is necessary to previously collect the training triplets, including the LR-HS, HR-RGB images, and corresponding ground-truth (label), i.e., the HR-HS images in the fully supervised method, and provided an elaborate design for fusing multiple modalities of observations with different spatial and spectral structures. Han et al. [36] conducted an initial investigation by directly inputting the fused data of an HR-RGB image and an up-sampled LR-HS image to a simple 3-layer CNN, and used a more complex CNN with a residual structure to achieve better performance [37].

Palsson et al. [38] explored a 3D CNN-based MS/HS fusion scheme, and integrated the principal component analysis (PCA) to reduce the computational cost. Dian et al. [39] proposed a multi-stage method, and used a simple optimization strategy for initial HS reconstruction and final refinement, while adopting a 20-layer CNN for learning a latent HR-HS image from the initial one. More recently, Wang et al. [40] exploited an efficient hyperspectral image fusion network by iteratively integrating the representation relations between the target and observations into the deep-learning network to achieve superior performance. Despite the high performance gain, these methods fail to generalize well among different datasets, and they need to separately train the reconstruction models regarding the datasets under investigation, even for small imaging condition changes.

2.2.2. Unsupervised Learning Method

To deal with the generalization limitation in the fully supervised learning method, the unsupervised neural network was proposed as a good solution to the HSI SR problem. Qu et al. [41] investigated an unsupervised encoder-decoder architecture to solve the MS/HS fusion problem. With the potential applicability of using a CNN-based end-to-end unsupervised neural network, this method needs to be carefully optimized for the two sub-networks in an alternating way, and its performance can be further improved. Based on the deep image prior (DIP) and the fact that the convolutional neural network itself can capture a large number of low-level image statistics to achieve a well-reconstructed natural image, Sidorov et al. [24] extended the DIP concept for automatically learning the underlying priors for HS images (DHPs) and applied it to the spatial resolution enhancement of an hyperspectral image. However, the DHPs can only leverage the observed LR-HS image for network training and cannot efficiently learn both the spatial structure and spectral attribute image priors for reconstructing a latent HR-HS image. Furthermore, Zhang et al. [42] leveraged the generated training triplets (the LR-HS, HR-RGB, and HR-HS images) with different degradation models to learn a common deep model for predicting an initial HR-HS image, and then exploited unsupervised adaptation learning for fine-turning the initial estimation and automatically learning the degradation operations of the under-studying observations. Although remarkable performance gain has been achieved with different degradation models compared with most state-of-the-art methods, the performance of the fine-turning HR-HS image in the adaptation learning is greatly affected by the initially estimation in the common model, and is easy to fall into a local minimum solution. In addition, Nie et al. [43] proposed two steps of learning method, where the spatial and spectral degradation models were first predicted via modeling the relation between the HR-HS image and the observations: LR-HS and HR-RGB images, and then the latent HR-HS image is reconstructed with the previously estimated degradation models. Though, feasibility and potential of the HR-HS image super resolution with unsupervised strategy is verified, current methods usually explore different steps of learning for obtaining acceptable performance for this challenge unsupervised learning. This study exploits a generative network for predicting the underlaying priors in the latent HR-HS image and the simply modified versions of the vanilla constitutional layer for approximating the degradation models, and thus implement the unsupervised HR-HS image super resolution in one stage. Our proposed method is mostly related to the DHP method, although there is a substantial difference between them. In this work, we effectively adopt both the observed LR-HS and HR-RGB images and propose an unsupervised fusion-learning framework for more robust HS image reconstruction.

3. Deep Unsupervised Fusion-Learning Method

In this section, the mathematical formulae of the fusion-based HSI SR problem are initially explained. Then the proposed deep unsupervised fusion-learning method, including the generative neural network architecture for automatically learning the underlying priors of a latent HR-HS image and the implementation of the constructed loss function for network training are introduced.

3.1. Problem Formulation

Given the observed image pair (LR-HS image $\mathbf{X} \in \mathbf{R}^{w \times h \times L}$ and HR-RGB image $\mathbf{Y} \in \mathbf{R}^{W \times H \times 3}$), where w and h represent the width and height, the goal of HSI SR is to recover a HR-HS image $\mathbf{Z} \in \mathbf{R}^{W \times H \times L}$, where W and H represent the width and height of $\boldsymbol{Y}$ and $\boldsymbol{Z}$, and L is the number of spectral channels in the HR-HS image. Generally, the mathematical relation for formulating degradation operations between the observed images: $\mathbf{X}$, $\mathbf{Y}$ and the target HR-HS image $\mathbf{Z}$ can be expressed as follows:

$$\mathbf{X} = \mathbf{k}^{(\mathbf{Spa})} \otimes \mathbf{Z}^{(\mathbf{Spa})} \downarrow + \mathbf{n}, \mathbf{Y} = \mathbf{Z} * \mathbf{C}^{(\mathbf{Spec})} + \mathbf{n}, \tag{1}$$

where $\otimes$ denotes the 2D convolution operator, $\mathbf{k}^{(\mathbf{Spa})}$ represents the spatial 2D blur kernel, $(Spa) \downarrow$is the spatial decimation (down-sampling) operator, $\mathbf{C}^{(\mathbf{Spec})}$ is the spectral sensitivity function of the RGB camera (three 1D spectral filters) for converting the L spectral band to the RGB band, and $\boldsymbol{n}$ is the additive white Gaussian noise (AWGN) of the noise level σ. The mathematical expression of the above degradation models can be expressed in the following simplified matrix format:

$$\mathbf{X} = \mathbf{DBZ} + \mathbf{n}, \mathbf{Y} = \mathbf{ZC} + \mathbf{n}, \tag{2}$$

where $\mathbf{B}$ and $\mathbf{D}$ stand for the blurring matrix in the spatial domain and down-sampling matrix, respectively, for transforming $\mathbf{Z}$ to $\mathbf{X}$. $\mathbf{C}$ denotes the spectral sensitivity function (CSF) of an RGB sensor. Assuming that the degradation parameters $\mathbf{B}$, $\mathbf{D}$, and $\mathbf{C}$ (which can be obtained from the hardware design of the HS and RGB sensors) are known, a heuristic approach to intuitively minimize the following reconstruction errors using the observed $\mathbf{X}$ and $\mathbf{Y}$ for estimating $\mathbf{Z}$ is given as follows:

$$\mathbf{Z}^* = \arg\min_{\mathbf{Z}} \alpha\beta_1 ||\mathbf{X} - \mathbf{DBZ}||_F^2 + (1-\alpha)\beta_2 ||\mathbf{Y} - \mathbf{ZC}||_F^2, \tag{3}$$

where $||\cdot||_F$ stands for the Frobenium norm. Since the element numbers in the HR-RGB and LR-HS images are different, it generally needs to introduce the normalization weights such as $\beta_1 = 1/N_1$ and $\beta_2 = 1/N_2$, where N_1 and N_2 are the products of the pixel numbers and the spectral bands in LR-HS and HR-RGB images, respectively. Beside, we further exploit a hyperparameter α ($0 \leq \alpha \leq 1$)to adjust the contributions between these two reconstruction errors. (3) aims to obtain an optimal $\mathbf{Z}^*$ for minimizing the weighted reconstruction error of the observations. Using the assumed AWGN in (2), (3) is completely equivalent to maximize the likelihood of a latent HR-HS image given the observation of $\mathbf{X}$ and $\mathbf{Y}$. It is known that in the HSI SR problem the total number of unknown variables in $\boldsymbol{Z}$ is much greater than the known variables in $\mathbf{X}$ and $\mathbf{Y}$, and this constitutes a severely ill-posed problem. Recovering a robust HR-HS image based on the observations is an extremely difficult task. To overcome this problem, most existing methods explore various hand-crafted image priors to model the underlying structure of the HR-HS image, and then impose a regularization term on the reconstruction error minimization problem, which can be formulated as follows:

$$\mathbf{Z}^* = \arg\min_{\mathbf{Z}} \alpha\beta_1 ||\mathbf{X} - \mathbf{DBZ}||_F^2 + (1-\alpha)\beta_2 ||\mathbf{Y} - \mathbf{ZC}||_F^2 + \lambda\phi(\mathbf{Z}), \tag{4}$$

where $\phi(\mathbf{Z})$ is a term for modeling the underlying structure of $\boldsymbol{Z}$, and $\boldsymbol{\lambda}$ represents a hyperparameter, which balances the regularization term and reconstruction error distributions. By introducing the prior probability function $Pr(\mathbf{Z})$, where $\phi(\mathbf{Z}) = -\log(\mathbf{Pr}(\mathbf{Z}))$, (4) can be explained as the widely used maximum a posterior (MAP) framework. Although high performance gain can be achieved using various hand-crafted image prior in the HSI SR scenario, finding an appropriate image priors for a specific scene remains a challenging task. This work aims to use the powerful learning capability of deep-learning networks for automatically learning the underlying image priors in latent HR-HS images. Based on a

DIP work, where the deep network architecture itself possesses a large number of low-level image statistics (image priors), a generative network for learning the spatial and spectral priors in a latent HR-HS image is employed. Then a reliable HR-HS image constrained by the learned priors using only the low-quality observations is reconstructed.

3.2. Proposed Deep Unsupervised Fusion Learning Method

In this section, the deep unsupervised fusion-learning framework for the HSI SR problem is introduced. Recent deep-learning-based HS reconstruction methods proved that a DCNN is capable of effectively capturing the underlying spatial and spectral structures (common prior information) of latent HS images and demonstrated promising performance. However, these methods are generally implemented in a fully supervised manner, and require large-scale training triplets containing LR-HS, HR-RGB, and HR-HS images, which are difficult to be specifically collected for obtaining the training labels (HR-HS images). Extensive research on natural image generation (DCGAN [44] and its variants) proved that high-definition and high-quality images with some defined characteristics and attributes can be successfully generated from a random noisy input without the supervision of high-quality ground-truth. This indicates that the inherent structure (priors) of a latent image with the defined characteristics can be captured by searching the neural network parameter space, starting from a random initial state. Moreover, DIPs [23] were exploited to model the more specific structures of an under-studying scene with the guidance of its degradation version only and successfully applied to several natural image restoration tasks such as image de-noising, impainting, and super-resolution. In this paper, this un-supervision paradigm is followed, aiming to learn the specific spatial and spectral structures (priors) of a latent HR-HS image with the guidance of its degraded observations (LR-HS and HR-RGB images). The conceptual scheme of the proposed DUFL framework is illustrated in Figure 1. Specifically, a generative neural network G_θ (θ is the network parameter to be learned) is leveraged to model the underlying spatial and spectral structures of a latent HR-HS image $\boldsymbol{Z}$. By replacing $\boldsymbol{Z}$ with G_θ in (4) and removing the regularization term $\phi(\mathbf{Z})$ due to the automatically captured priors in the generative network, the fusion-based HSI SR model can be reformulated as follows:

$$\theta^* = \arg\min_{\theta} \alpha\beta_1 ||\mathbf{X} - \mathbf{DB}G_\theta(\mathbf{z_{in}})||_F^2 + (1-\alpha)\beta_2 ||\mathbf{Y} - G_\theta(\mathbf{z_{in}})\mathbf{C}||_F^2, \tag{5}$$

where z_{in} is the input of the generative neural network, and $G_\theta(z_{in})_i$ represents the i-th element of the estimated HR-HS image. Instead of directly optimizing a raw HR-HS image, which is extremely large and not unique, (5) aims to search the parameter space of the generative neural network G_θ for leveraging the possessed priors in it. To solve the objective function using the searching parameter space of the generative neural network, a detailed description of the used input data and the substantial architecture of the generative neural network is provided. The implementation of the degradation operations, including the spatial blurring/down-sampling operation and spectral transformation, are also provided.

Input Data to the Generative Neural Network: A popular generative neural network is trained to create a target image with the predefined characteristics, where randomly sampled noisy vectors based on a distribution function, such as Gaussian or uniform distribution, are usually used as inputs to ensure sufficient variety and diversity in the generated images. In our HSI SR task, the corresponding HR-HS images of the observed degradation version (LR-HS and HR-RGB images) must be acquired. Therefore, an intuitive way is to adopt an initially sampled noisy vector z_{in}^0 and fix it to search the optimal network parameter space for a specific HR-HS image. However, the fixed noisy input possibly leads to the generative neural network falling into a local minimum state. This results in an un-plausible estimation of the latent HR-HS image. Thus, a perturbation in the fixed initial input with a small randomly generated noisy vector at each training step is proposed to

avoid a local minimum state. The input vector for the i-th training step can be expressed as follows:

$$z_{in}^{i} = z_{in}^{0} + \beta \mathbf{n}_i, \tag{6}$$

where $\mathbf{n}_i$ is the randomly sampled noise vector at the i-th training step, and β denotes the perturbation degree (a small scalar value). After learning the generative neural network G_θ with the perturbed input, a prediction using the fixed noisy vector $\mathbf{Z}^* = G_\theta(z_{in}^0)$ as the final estimated HR-HS image is performed.

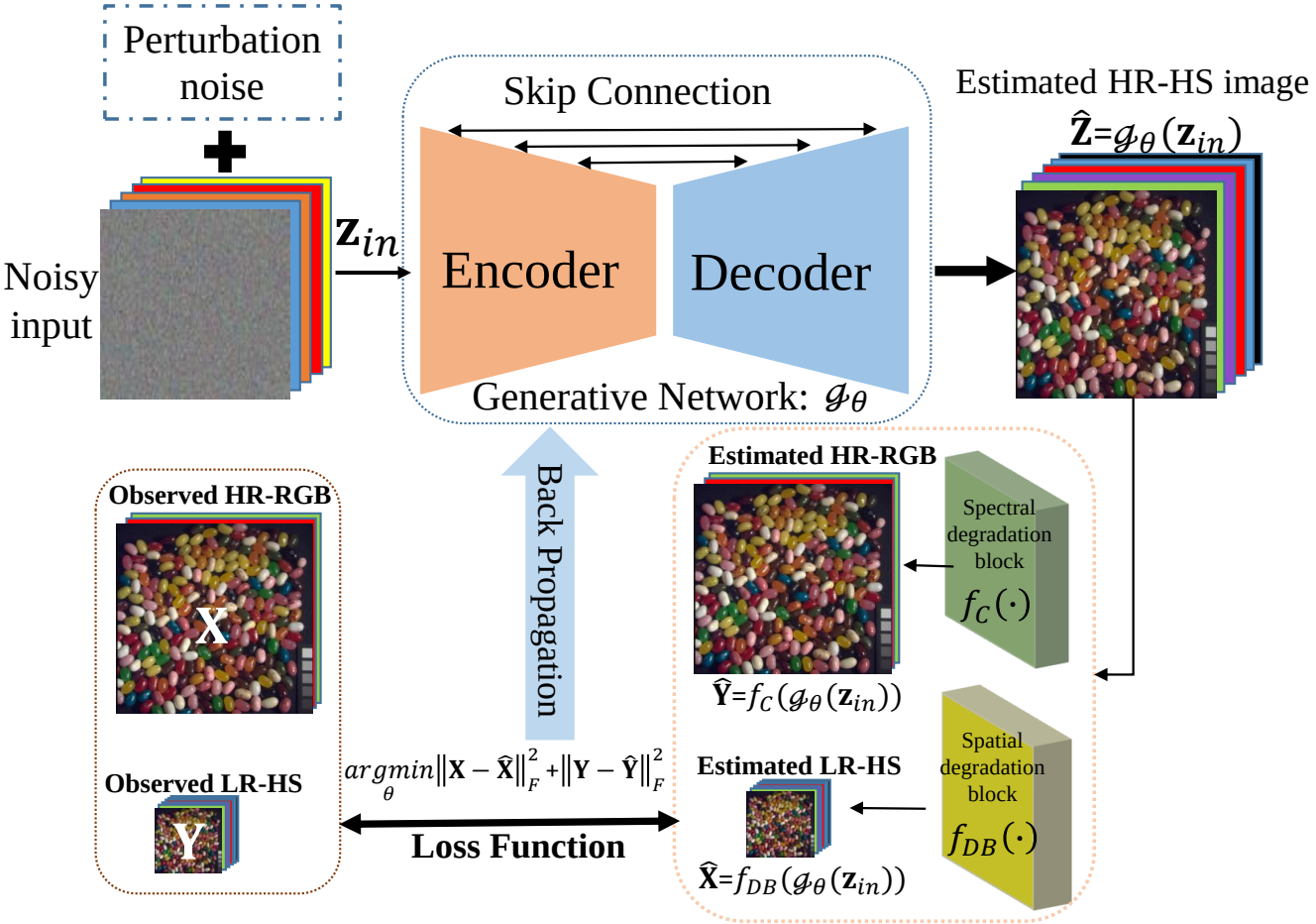


Figure 1. Conceptual diagram of the proposed DUFL method. An encoder-decoder-based generative neural network is used for modeling the rich spatial and spectral structures of a latent HR-HS image, and two parallel specifically designed convolutional blocks are used for approximating the spatial and spectral degradation models. Then the proposed framework is trained with the observations (LR-HS and HR-RGB images).

Detailed Architecture of the Generative Neural Network:An arbitrary DCNN architecture can be adopted to serve as the generative neural network G_θ. Due to diverse information, such as salient structures, rich textures, and complex spectra in a latent HR-HS image, a generative neural network G_θ is required to provide sufficient modeling capabilities. Various generative neural networks, such as that in the adversarial learning scenario [Pix2pix and others], were already proposed and demonstrated their great potential to generate high-quality natural images [45]. In this work, an encoder-decoder architecture with its multi-level feature-learning property and simplification is leveraged, and the skip connections between the encoder and decoder paths are leveraged for feature reusing. A detailed generative neural network is illustrated in Figure 1. Both the encoder and decoder consist of 5 blocks, which can learn the representative features in different scales, and the outputs of all 5 blocks in the encoder side are transferred to the corresponding

decoder with skip connection for reusing the extracted detailed features. Each block is composed of 3 convolutional layers, following the RELU activation function, where the max-pooling layer with a 2×2 kernel is used to reduce the feature map size between the encoder blocks, whereas the up-sampling layer is employed to doubly recover the feature map size between the decoder blocks. Finally, a convolutional output layer is adopted for estimating the latent HR-HS image. However, in the unsupervised learning setting, there is no ground-truth HR-HS image for guiding or evaluating the training states of the generative neural network. Then, the observed HR-RGB and LR-HS images are leveraged to construct the evaluation criterion described in (5).

Implementation of the Degradation Operations: From the predicted HR-HS images of the generative neural network, it is possible to use mathematical degradation operations to approximate the LR-HS and HR-RGB images and then formulate quantitative criteria (loss function) for network learning (5). However, mathematical operations outside the generative neural network may result in difficulties in the training procedure. In this work, without any loss of generalization, two parallel special convolutional blocks are leveraged to implement the degradation model following the generative neural network and construct an end-to-end learnable framework. Specifically, the vanilla convolution layer is modified to be adapted for approximating the blurring/down-sampling operations and spectral transformation. Since, in a real scenario, each spectral band undergoes the same blurring/down-sampling transformations, the same kernel is defined for different spectral bands (channels) in a depth-wise convolutional layer with the stride parameter as the spatial expanding factor, and the bias term is set to 'False'. The formula for the blurring/dawn-sampling transformation can be expressed as follows:

$$\hat{\mathbf{X}} = f_{\mathbf{DB}}(G_\theta(z_{in}^0)), \tag{7}$$

where $f_{\mathbf{DB}}()$ represents the transformation of an especially designed depth-wise convolutional layer.

For implementing the spectral transformation from the generated HR-HS image $\hat{\mathbf{Z}}$ to the approximation of $\mathbf{X}$, a point-wise convolutional layer with 3 output channels is applied, and the bias term as set to 'False'. The formula for this spectral transformation can be expressed as follows:

$$\hat{\mathbf{Y}} = f_{\mathbf{C}}(G_\theta(z_{in}^0)), \tag{8}$$

where $f_{\mathbf{C}}()$ represents the spectral transformation with the point-wise convolutional layer. Using these two parallel blocks, an end-to-end learnable framework can be realized. For the known spatial blurring/down-sampling degradation, the kernel weight of the especially designed depth-wise convolutional layer with the known ones is initialized, and the trainable parameter is set to 'False'. Similarly, kernel weights are set for the point-wise convolutional layer as the known CSF (spectral transformation matrix) of the RGB camera. Therefore, the proposed fusion-learning framework is considerably flexible and can be easily adapted to various degradation models. Moreover, it also has the prospect of automatically learning the transformation parameters in the embedded convolutional blocks for unknown degradations, which is left for future research. By replacing the spatial and spectral degradation operations with the designed convolutional blocks, the loss function for training the proposed deep unsupervised fusion learning network can be rewritten as follows:

$$\theta^* = \arg\min_\theta \alpha\beta_1 ||\mathbf{X} - f_{\mathbf{DB}}(G_\theta(\mathbf{z_{in}}))||_F^2 + (1-\alpha)\beta_2 ||\mathbf{Y} - f_{\mathbf{C}}(G_\theta(\mathbf{z_{in}}))||_F^2, \tag{9}$$

The optimization of (9) for obtaining the optimal parameter set of the generative neural network can be considered to be a kind of "zero shot" learning [46]. During the training procedure, only the low-quality image pairs (that is, the observed LR-HS and HR-RGB images) are used without the corresponding label (the HR-HS images) and

thus the proposed method is completely unsupervised of being generalized for any real observations. The detail implementation for the proposed DUFL method is summarized in Algorithm 1.

Algorithm 1 Algorithm of the proposed deep unsupervised fusion learning method.

Input: The observed LR-HS image $\mathbf{X}$ and HR-RGB image $\mathbf{Y}$
Output: Latent HR-HS image $\mathbf{Z}$

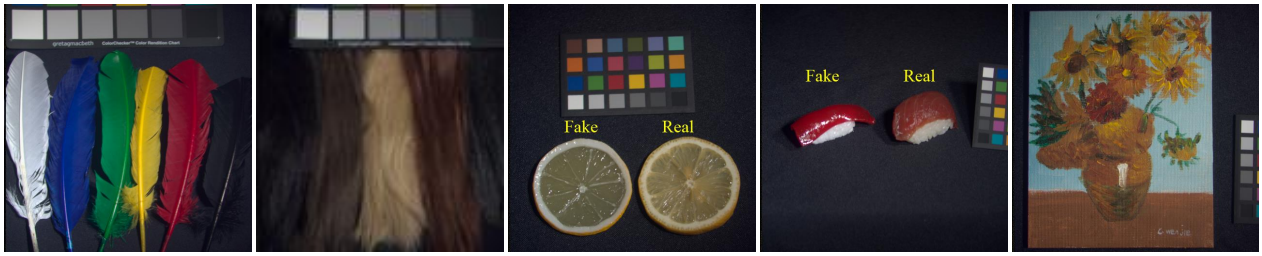
Sample $\mathbf{z}_{in}^{0}$ from uniform distribution with seed 0
for $i = 0$ to max. iter. (I) **do**
  Sample $\mathbf{n}_{(0,1)}^{i}$ from uniform distribution
  Perturb $\mathbf{z}_{in}^{0}$ with $\mathbf{n}_{(0,1)}^{i}$: $\mathbf{z}_{in}^{i} = \mathbf{z}_{in}^{0} + \beta \mathbf{n}_{(0,1)}^{i}$
  $\hat{\mathbf{Z}} = G_{\theta}(\mathbf{z}_{in}^{i}, \theta^{i-1})$
  $\hat{\mathbf{X}} = f_{DB}(\hat{\mathbf{Z}})$
  $\hat{Y} = f_{C}(\hat{\mathbf{Z}})$
  Loss function: $\alpha\beta_1||\mathbf{X} - \hat{\mathbf{X}}||_F^2 + (1-\alpha)\beta_2||\mathbf{Y} - \hat{\mathbf{Y}}||_F^2$
  Compute the gradients regarding G_{θ}
  Update θ using the ADAM algorithm [47] as θ^{i}
end for
$\mathbf{Z} = G_{\theta}(\mathbf{z}_{in}^{0})$

4. Experiment Results

4.1. Experimental Settings

Datasets:Two benchmark HSI datasets, including the CAVE [48] and Harvard [49] datasets, were adopted to evaluate the efficiency of the proposed method. The CAVE dataset contains 32 HS images with a spatial resolution of 512 × 512 pixels and a spectral resolution of 31 bands in the spectral range 400 nm–700 nm. The images in the CAVE dataset contain various real-world materials and objects, such as human faces, fruit, etc. The Harvard dataset contains 50 images of various natural scenes with a spatial resolution of 1392 ×1040 pixels and a spectral resolution of 31 bands in the spectral range 420 nm–720 nm. We illustrate some example images of the CAVE and Harvard datasets in Figure 2. In the experiments conducted, the upper-left sub-image portions of size 1024 × 1024 pixels were cropped from the original HS images in the Harvard dataset and downsampled them to 512 × 512 pixel images, which constitute the ground-truth HS images. The observed LR-HS images were synthesized from the ground-truth HS images in both datasets using bicubic down-sampling with different spatial decimation factors (8, 16, and 32) to create images with a size of 64 × 64 × 31, 32 × 32 × 31 and 16 ×16 × 31, respectively. Meanwhile, the observed HR-RGB images were generated by multiplying the spectral response function of the Nikon D700 camera with the ground truth HR-HS images.

Comparison Methods: There are three main research paradigms for HS image super-resolution, containing (1) traditional optimization methods with manually engineered image priors based on accumulated experience or clarified physical knowledge; (2) unsupervised methods with automatically learned image priors; (3) fully supervised deep-learning methods with learned external image priors (training samples). First, the proposed method (unsupervised method with learned image priors) with a spatial up-scale factor of 32 of the CAVE dataset was compared with different paradigms of SoTA methods, including GOMP [50], MF [16], SNNMF [51], DHIP [24], uSDN [41], PNN [52], 3D-CNN [53], and CMHF-net [54] methods. Since the fully supervised deep-learning methods (PNN [52], 3D-CNN [53], and CMHF-net [54]) must be separately trained to construct individual models for different spatial up-scale factors of different datasets and a long time is required to train just one model, the results for other up-scale factors (8, 16) of the CAVE dataset and all factors of the Harvard dataset were compared with those obtained from the unsupervised SoTA methods.



(a) CAVE dataset



(b) Harvard dataset

Figure 2. Example images in the CAVE and Harvard datasets.

Evaluation Metrics: To quantitatively evaluate the performance of the reconstructed HS images, the following five commonly used metrics were adopted: the root-mean-square-error (RMSE), peak signal-to-noise ratio (PSNR), structural similarity (SSIM), spectral angle mapper (SAM) and relative size global error (ERGAS). The RMSE, PSNR, and ERGAS were individually calculated per pixel and per spectral band and averaged over all spatial positions and spectral bands. The SSIM was computed for spatial structure evaluation. All these metrics exhibit the spatial fidelity of the reconstructed HSIs. The SAM exhibits spectral fidelity, which is calculated for each 1-D spectral vector and averaged over all spatial positions. Larger values of the PSNR and SSIM indicate better performance, whereas smaller values of the RMSE, ERGAS, and SAM imply better SR results.

Implementation Details: The proposed method was implemented in Pytorch. The input noise was initially set to the same dimension as the to-be-estimated HR-HS image. To train the generative network, the Adam optimizer [47] with the simple L_2 norm-based loss was adopted. The learning rate was initialized to 1e-3 and reduced by 0.7 every 1000 steps. Moreover, the perturbation degree β was initially set to 0.05 and reduced by 0.5 every 1000 steps. The optimization process was terminated after 12,000 steps and fixed for all images with different upscale factors from different datasets. All experiments were conducted in the learning environment with Tesla K80 GPU. In our experiment, the learning time for an image with size of 512×512 is about 20 min.

4.2. Numerical Results Comparison

Table 1 shows the compared quantitative evaluation of the super-resolved HS images for an up-scale factor of 32 of the CAVE dataset with different paradigms, including the traditional methods with manually engineered image priors (GOMP [50], MF [16], SNNMF [51]), the unsupervised methods with learned priors (DHIP [24], uSDN [41]), and fully supervised deep-learning methods (PNN [52], 3D-CNN [53], and CMHF-net [54]). Since an HR-HS image in the proposed method can be predicted by the generative neural network using the initial fixed noise w/o perturbation, two predictions with different inputs (denoted as DUFL and $DUFL_+$) can be achieved. In Table 1, it can be observed that the proposed method significantly outperforms the SoTA methods, belonging to the same unsupervised paradigm with the learned image priors (DHIP [24] and uSDN [41]). For the results obtained using the uSDN [41] method, the released code (https://github.com/

mbaddeley/usdn, accessed on 18 October 2020), was re-run with default hyper-parameters, and only 8 samples out of 32 images provided correct super-resolved results. Thus, the average performance using only 8 correct samples is presented in Table 1. Although the proposed DUFL method can be robustly learned with the provided hyper-parameters mentioned in the previous Section, it is not necessary to adjust these parameters for different datasets and up-scale factors. It should also be noted that it is possible to improve its performance to conduct hyper-parameter turning for different datasets and up-scale factors. Moreover, the proposed method exhibits better performance than the fully supervised deep-learning methods and most traditional methods. One possible reason for this performance is the small number of samples used for training the HSI SR model, which is a challenge issue in the HSI SR scenario. The proposed method exhibits comparable results (without integrating any prior knowledge) with the SNNMF [51] method by leveraging the manually engineered image priors, which evolved and proved to be efficient by conducting a long-time research effort. Additionally, the proposed DUFL method aims to generate latent HR-HS images from the noise input and does not leverage any existing rich spectral information and spatial structure in the observed LR-HS and HR-RGB images for regularizing the generative neutral network learning. Thus, the HSI SR performance is expected to be improved by imposing some constraints on the network training using the observations. This is left for future investigation.

A comparison between the results obtained using the unsupervised SoTA methods on the CAVE dataset (with up-scale factors of 16 and 8) and the Harvard dataset is prensented in Tables 2 and 3, respectively. In these tables, it can be observed that the proposed method achieves better or comparable results with those obtained using SoTA methods. Furthermore, it must be clarified that the uSDN method cannot produce accurate results for the same samples, especially for large up-scale factors, and the average evaluations computed only with the correct outputs are presented in Tables 2 and 3. In Tables 1–3, the up-arrow indicates better result with larger value while the down-arrow denotes better result with smaller value. Meanwhile, in order to evaluate the effect of different data terms on the loss function, we set the hyper-parameter α value as 0.3, 0.5 and 0.7, respectively, and provide the compared result in Table 4. Table 4 manifests the quantitative metrics of PSNR, SAM and ERGAS with our DUFL+ method, which illustrates that there are no large impact on the super-resolution performance via fine-tuning the hyper-parameter α. We are going to investigate in detail this phenomenon in the future.

4.3. Perceptual Quality

To visualize the reconstruction results, two representative images from the CAVE and Harvard datasets, respectively, with different upscale factors (8, 16, and 32) are shown in Figures 3–8 using different deep unsupervised methods (DHIP [24], uSDN [41], the proposed (DUFL method, and the $DUFL_+$ method w/o a perturbation term). In all figures, the first row shows the original HR image and the super-resolved results with spatial and spectral fidelity indexes (PSNR Sam for the recovered images using different methods), whereas the second row shows the difference images between the recovered HR-HS image and the ground-truth HR-HS image. In these figures, it can be observed that the proposed DUFL method is capable of recovering the HR-HS image with a smaller difference to the ground-truth HR-HS image and more reliable spatial/spectral indexes for most cases, excluding the results of uSDN in Figure 8. As mentioned in Section 4.2, although the uSDN method is capable of achieving impressive performance for some specific images using the default hyper-parameters, it cannot provide accurate results for most of the images, especially for those with large upscale factors. Despite the better results obtained for some images using the uSDN method, accurate recovered results cannot be achieved for 20 images out of the total 50 images. Moreover, the proposed $DUFL_+$ method is capable of further improving the performance of the DUFL-version method for all images with different upscale factors.

Table 1. Comparison of the quantitative evaluation results obtained using the proposed method with those obtained by state-of-the-art methods, including traditional methods with manually engineered image priors (GOMP [50], MF [16], and SNNMF [51]), unsupervised methods with learned image priors (DHIP [24] and uSDN [41]), and fully supervised deep-learning methods (PNN [52], 3D-CNN [53], and CMHF-net [54]).

Type of Priors	Method	RMSE↓	PSNR↑	SSIM↑	Sam↓	ERGAS↓
Manually Engineered Priors	GOMP	6.47	32.48	-	14.19	0.77
	MF	3.03	39.37	-	6.12	0.40
	SNNMF	3.26	38.73	-	6.50	0.44
Learned Priors	uSDN	4.217	35.79	0.9216	15.75	0.5171
	DHIP	16.01	24.73	0.7449	13.0761	2.1490
	DUFL (Our)	3.47	38.17	0.9512	8.31	0.4597
	DUFL+ (Our)	3.34	38.47	0.9548	8.12	0.4436
Supervised External Priors	PNN	6.103	32.42	0.962	14.73	1.3451
	3D-CNN	4.629	34.82	0.937	8.96	1.0920
	CMHF-net	3.507	37.23	0.962	7.30	0.8187

Table 2. Comparison of the evaluation results using the proposed methods with those obtained using the unsupervised SoTA methods, including the manually engineered image priors (GOMP [50], MF [16], SNNMF [51]) and automatically learned image priors (DHIP [24] and uSDN [41]) in the CAVE dataset for up-scale factors of 16 and 8.

Up-Scale Factor	Type of Priors	Method	RMSE↓	PSNR↑	SSIM↑	Sam↓	ERGAS↓
16	Manually Engineered Priors	GOMP	6.08	32.96	-	12.60	43
		MF	2.71	40.43	-	4.82	0.73
		SNNMF	2.45	41.21	-	4.61	0.66
	Learned Priors	uSDN	4.389	35.74	0.9111	18.7928	1.3469
		DHIP	11.31	27.76	0.8024	10.6639	3.0838
		DUFL (Our)	2.6053	40.71	0.9665	6.6163	0.6990
		DUFL+ (Our)	2.497	41.03	0.9694	6.434	0.6741
8	Manually Engineered Priors	GOMP	5.69	33. 64	-	11.86	2.99
		MF	2.34	41.83	-	3.88	1.26
		SNNMF	1.89	43.53	-	3.42	1.03
	Learned Priors	uSDN	5.808	33.402	0.8779	24.2909	7.0549
		DHIP	7.5995	31.4040	0.8708	9.2542	4.2025
		DUFL (Our)	2.078	42.50	0.9750	5.355	1.155
		DUFL+ (Our)	1.957	42.98	0.9774	5.223	1.098

Table 3. Comparison of the evaluation results using the proposed methods with those obtained using the unsupervised SoTA methods, including the manually engineered image priors (GOMP [50], MF [16], SNNMF [51]) and automatically learned image priors (DHIP [24] and uSDN [41]) in the Harvard dataset for up-scale factors of 32, 16, and 8.

Up-Scale Factor	Type of Priors	Method	RMSE↓	PSNR↑	SSIM↑	Sam↓	ERGAS↓
32	Manually Engineered Priors	GOMP	4.08	38.02	-	4.79	0.41
		MF	1.96	43.19	-	2.93	0.23
		SNNMF	2.20	42.03	-	3.17	0.26
	Learned Priors	uSDN	3.328	37.75	0.9729	6.091	0.5162
		DHIP	13.25	26.23	0.7186	5.6758	1.4121
		DUFL (Our)	2.8203	40.119	0.9571	3.9550	0.4268
		DUFL+ (Our)	2.6199	40.7461	0.9627	3.9124	0.4173
16	Manually Engineered Priors	GOMP	3.83	38.56	-	4.16	0.77
		MF	1.94	43.30	-	2.85	0.47
		SNNMF	1.93	43.31	-	2.82	0.45
	Learned Priors	uSDN	2.409	40.529	0.9827	3.834	0.6308
		DHIP	10.38	38.44	0.754	4.57	2.08
		DUFL (Our)	2.814	40.767	0.9527	3.0081	0.7479
		DUFL+ (Our)	2.556	41.656	0.9593	2.9529	0.7118
8	Manually Engineered Priors	GOMP	3.79	38.89	-	4.00	1.65
		MF	1.83	43.74	-	2.66	0.87
		SNNMF	1.79	43.86	-	2.63	0.85
	Learned Priors	uSDN	2.423	42.11	0.9848	3.882	1.1707
		DHIP	7.9489	30.8609	0.8029	3.5295	3.1509
		DUFL (Our)	2.3842	42.16	0.9646	2.3545	1.0869
		DUFL+ (Our)	2.1167	43.23	0.9709	2.3045	1.007

Table 4. Comparison of the evaluation results using our proposed method: DUFL+ with different weights α values of 0.3, 0.5 and 0.7 in the CAVE and Harvard datasets for up-scale factors of 16 and 8. Bold values indicate the best results.

		CAVE			Harvard		
Up-Scale Factor	α	**PSNR↑**	**SAM↓**	**ERGAS↓**	**PSNR↑**	**SAM↓**	**ERGAS↓**
32	0.3	38.8653	7.0733	0.3312	39.4625	3.8458	0.4368
	0.5	38.0345	7.2645	0.3708	40.0153	3.5062	0.3910
	0.7	39.1060	6.6191	0.3334	39.0670	3.3812	0.3879
16	0.3	40.7483	5.7067	0.5395	40.9471	2.8952	0.6622
	0.5	40.7500	5.8747	0.5518	40.7912	2.7029	0.6181
	0.7	40.4173	5.6399	0.5829	41.9040	2.4800	0.5177
8	0.3	42.1921	5.0925	0.9452	43.0742	2.1642	0.9310
	0.5	42.9066	4.4038	0.8570	41.6796	2.1902	1.0589
	0.7	42.1635	4.7505	0.9171	41.8517	2.1843	1.0885

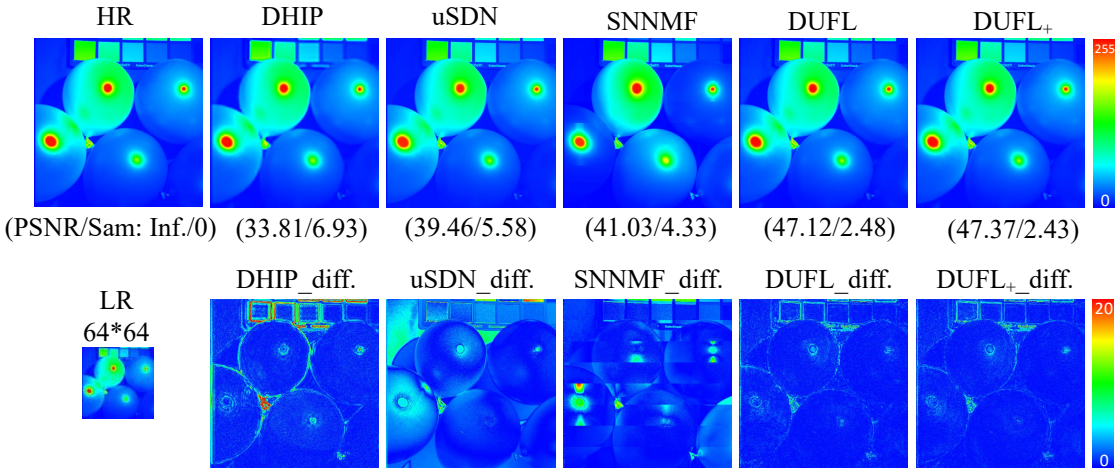


Figure 3. Recovered HR-HS image of the 'Balloon' sample in the CAVE dataset using the DHIP [24], uSDN [41], SNNMF [51], and the proposed methods and corresponding difference images between the ground-truth and recovered images with an upscale factor of 8.

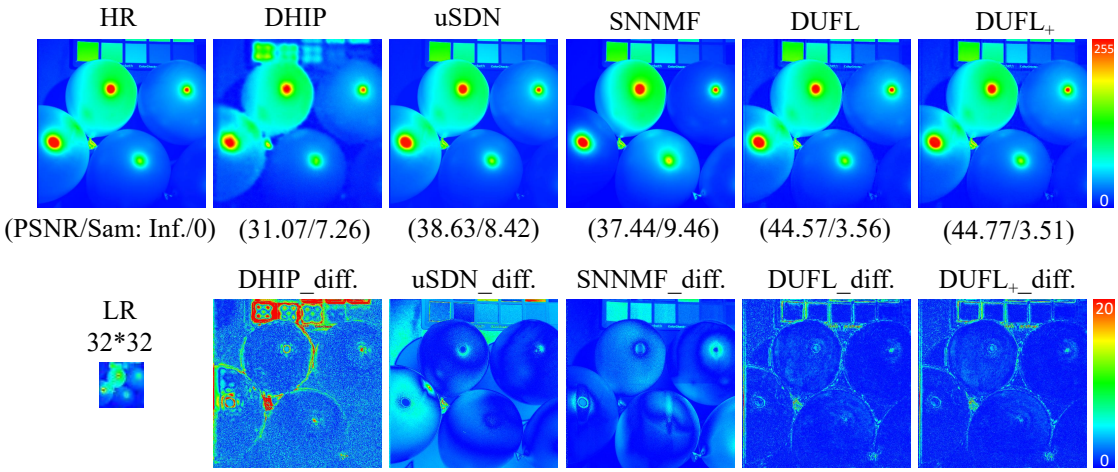


Figure 4. Recovered HR-HS image of the 'Balloon' sample in the CAVE dataset using the DHIP [24], uSDN [41], SNNMF [51], and the proposed methods and corresponding difference images between the ground-truth and recovered images with an upscale factor of16.

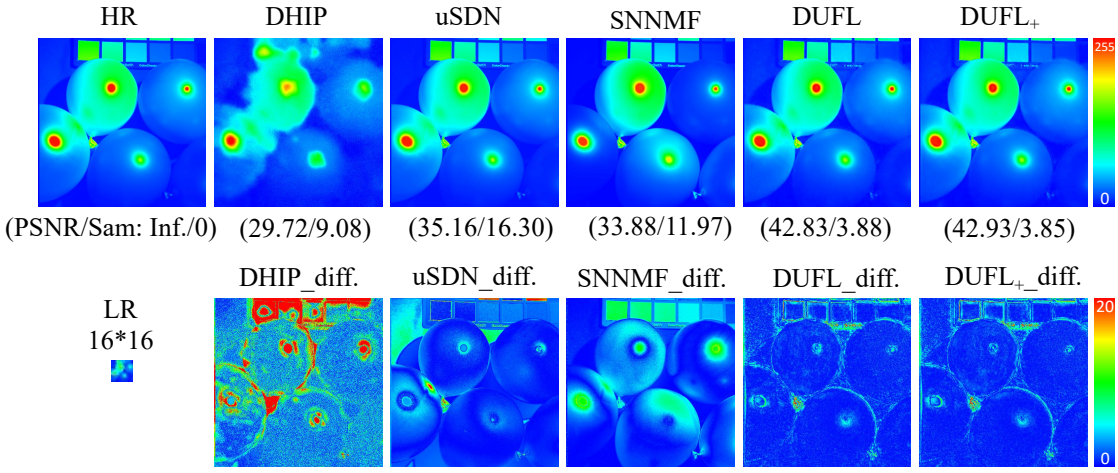


Figure 5. Recovered HR-HS image of the 'Balloon' sample in the CAVE dataset using the DHIP [24], uSDN [41], SNNMF [51], and the proposed methods and corresponding difference images between the ground-truth and recovered images with an upscale factor of 32.

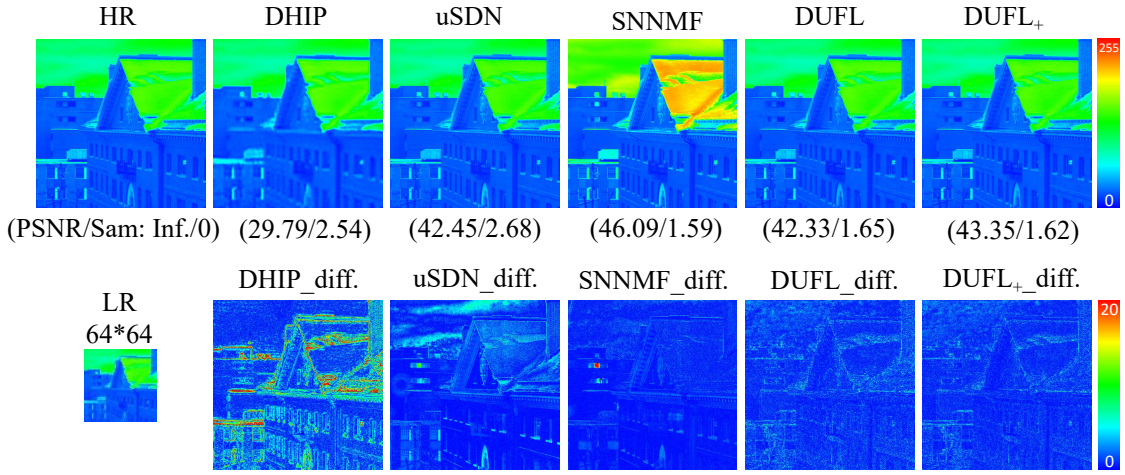


Figure 6. Recovered HR-HS image of the 'img1' sample in the Harvard dataset using the DHIP [24], uSDN [41], SNNMF [51], and the proposed methods and corresponding difference images between the ground-truth and recovered images with an upscale factor of 8.

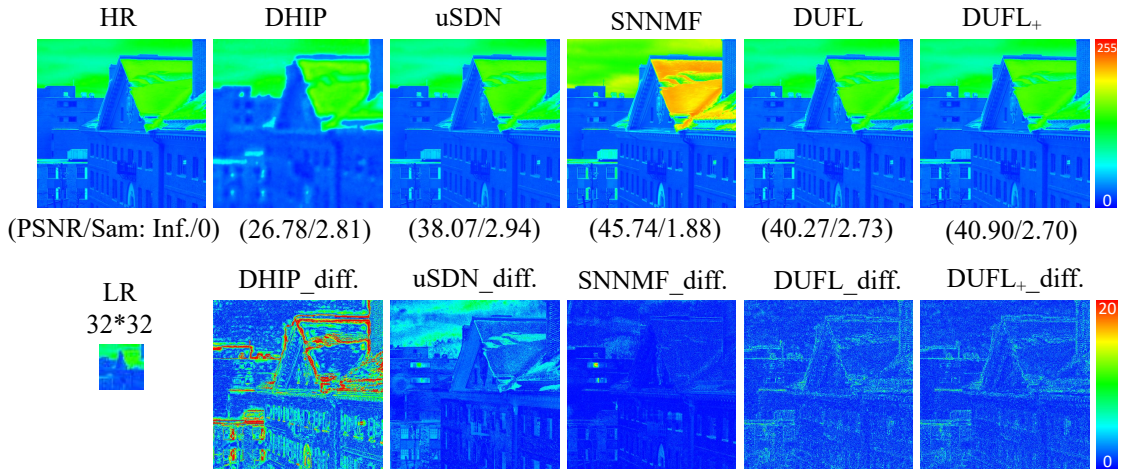


Figure 7. Recovered HR-HS image of the 'img1' sample in the Harvard dataset using the DHIP [24], uSDN [41], SNNMF [51], and the proposed methods and corresponding difference images between the ground-truth and recovered images with an upscale factor of 16.

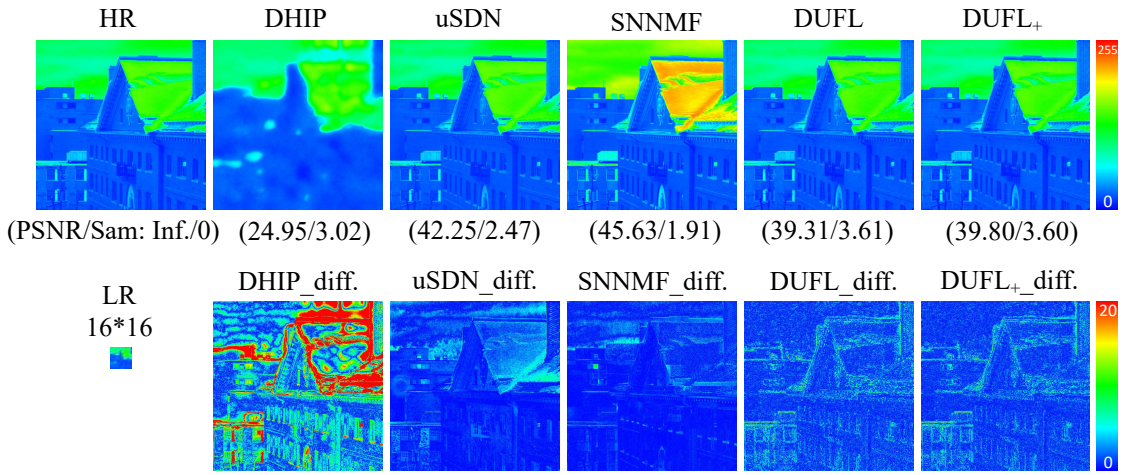


Figure 8. Recovered HR-HS image of the 'img1' sample in the Harvard dataset using the DHIP [24], uSDN [41], SNNMF [51], and the proposed methods and corresponding difference images between the ground-truth and recovered images with an upscale factor of 32.

5. Conclusions

In this work, a deep unsupervised fusion-learning framework for the hyperspectral image super-resolution problem was proposed. Based on the fact that the neural network architecture itself can capture a large number of low-level image statistics, a deep convolutional neural network for automatically learning the underlying spatial and spectral image priors in latent HR-HS images from a perturbed noisy input was leveraged. Specifically, an encoder-decoder architecture with skip connection was adopted for modeling the diverse context in a latent HR-HS image, and the special depth-wise and point-wise convolutional blocks were designed to implement the degradation transformations between observations and the required target, thus establishing an end-to-end learnable framework using only low-quality observations. The proposed unsupervised learning neural network is capable of effectively leveraging the HR spatial structure in HR-RGB images and the detailed spectral properties in LR-HS images to provide more reliable HS image reconstruction without using any training samples. Extensive experiments on two hyperspectral datasets, including the CAVE and Harvard datasets, demonstrated that the proposed method is capable of achieving impressive performance in terms of quantitative evaluation and visual effect.

Author Contributions: Conceptualization, methodology, writing—original draft preparation, review and editing, Z.L. and X.-H.H.; funding acquisition, Y.Z. and X.-H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 20K11867, JSPS KAKENHI Grant Number JP19K20307, and ROIS NII Open Collaborative Research 2020-20FC02.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The data can be found here: [https://www.cs.columbia.edu/CAVE/databases/multispectral/] (accessed on 28 March 2021) and [http://vision.seas.harvard.edu/hyperspec/d2x5g3/] (accessed on 28 March 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote. Sens. Mag.* **2013**, *1*, 6–36. [CrossRef]
2. Feng, Y.Z.; Sun, D.W. Application of hyperspectral imaging in food safety inspection and control: A review. *Crit. Rev. Food Sci. Nutr.* **2012**, *52*, 1039–1058. [CrossRef] [PubMed]
3. Qin, J.; Kim, M.S.; Chao, K.; Chan, D.E.; Delwiche, S.R.; Cho, B.K. Line-scan hyperspectral imaging techniques for food safety and quality applications. *Appl. Sci.* **2017**, *7*, 125. [CrossRef]
4. ElMasry, G.; Gou, P.; Al-Rejaie, S. Effectiveness of specularity removal from hyperspectral images on the quality of spectral signatures of food products. *J. Food Eng.* **2021**, *289*, 110148. [CrossRef]
5. Camps-Valls, G.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1351–1362. [CrossRef]
6. Mou, L.; Ghamisi, P.; Zhu, X.X. Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3639–3655. [CrossRef]
7. Uddin, M.P.; Mamun, M.A.; Afjal, M.I.; Hossain, M.A. Information-theoretic feature selection with segmentation-based folded principal component analysis (PCA) for hyperspectral image classification. *Int. J. Remote Sens.* **2021**, *42*, 286–321. [CrossRef]
8. Liang, J.; Zhou, J.; Bai, X.; Qian, Y. Salient object detection in hyperspectral imagery. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; pp. 2393–2397.
9. Kang, X.; Duan, P.; Xiang, X.; Li, S.; Benediktsson, J.A. Detection and correction of mislabeled training samples for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5673–5686. [CrossRef]
10. Tian, Y.; Li, Z.; Lin, Y.; Xiang, L.; Li, X.; Shao, Y.; Tian, J. Metal object detection for electric vehicle inductive power transfer systems based on hyperspectral imaging. *Measurement* **2021**, *168*, 108493. [CrossRef]
11. Calin, M.A.; Parasca, S.V.; Savastru, D.; Manea, D. Hyperspectral imaging in the medical field: Present and future. *Appl. Spectrosc. Rev.* **2014**, *49*, 435–447. [CrossRef]
12. Zhang, D.; Zhang, J.; Wang, Z.; Sun, M. Tongue colour and coating prediction in traditional Chinese medicine based on visible hyperspectral imaging. *IET Image Process.* **2019**, *13*, 2265–2270. [CrossRef]

13. Dremin, V.; Marcinkevics, Z.; Zherebtsov, E.; Popov, A.; Grabovskis, A.; Kronberga, H.; Geldnere, K.; Doronin, A.; Meglinski, I.; Bykov, A. Skin complications of diabetes mellitus revealed by polarized hyperspectral imaging and machine learning. *IEEE Trans. Med. Imaging* **2021**. [CrossRef] [PubMed]
14. Zhang, H.; Zhang, L.; Shen, H. A super-resolution reconstruction algorithm for hyperspectral images. *Signal Process.* **2012**, *92*, 2082–2096. [CrossRef]
15. Yokoya, N.; Yairi, T.; Iwasaki, A. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.* **2011**, *50*, 528–537. [CrossRef]
16. Kawakami, R.; Matsushita, Y.; Wright, J.; Ben-Ezra, M.; Tai, Y.W.; Ikeuchi, K. High-resolution hyperspectral imaging via matrix factorization. In Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 2329–2336.
17. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3973–3985. [CrossRef]
18. Zhang, Y.; Du, B.; Zhang, L.; Wang, S. A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 1376–1389. [CrossRef]
19. Wei, W.; Zhang, L.; Jiao, Y.; Tian, C.; Wang, C.; Zhang, Y. Intracluster structured low-rank matrix analysis method for hyperspectral denoising. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 866–880. [CrossRef]
20. Mei, S.; Hou, J.; Chen, J.; Chau, L.P.; Du, Q. Simultaneous spatial and spectral low-rank representation of hyperspectral images for classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2872–2886. [CrossRef]
21. Han, X.H.; Zheng, Y.; Chen, Y.W. Multi-Level and Multi-Scale Spatial and Spectral Fusion CNN for Hyperspectral Image Super-Resolution. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Seoul, Korea, 27 October–2 November 2019.
22. Han, X.H.; Shi, B.; Zheng, Y. Ssf-cnn: Spatial and spectral fusion with cnn for hyperspectral image super-resolution. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 2506–2510.
23. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep image prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9446–9454.
24. Sidorov, O.; Yngve Hardeberg, J. Deep Hyperspectral Prior: Single-Image Denoising, Inpainting, Super-Resolution. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Seoul, Korea, 27 October–2 November 2019.
25. Chen, Z.; Pu, H.; Wang, B.; Jiang, G.M. Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1418–1422. [CrossRef]
26. Licciardi, G.A.; Khan, M.M.; Chanussot, J.; Montanvert, A.; Condat, L.; Jutten, C. Fusion of hyperspectral and panchromatic images using multiresolution analysis and nonlinear PCA band reduction. *Eurasip J. Adv. Signal Process.* **2012**, *2012*, 1–17. [CrossRef]
27. Vivone, G.; Restaino, R.; Licciardi, G.; Dalla Mura, M.; Chanussot, J. Multiresolution analysis and component substitution techniques for hyperspectral pansharpening. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 2649–2652.
28. Han, X.H.; Wang, J.; Shi, B.; Zheng, Y.; Chen, Y.W. Hyper-spectral image super-resolution using non-negative spectral representation with data-guided sparsity. In Proceedings of the 2017 IEEE International Symposium on Multimedia (ISM), Taichung, Taiwan, 11–13 December 2017; pp. 500–506.
29. Lanaras, C.; Baltsavias, E.; Schindler, K. Hyperspectral super-resolution by coupled spectral unmixing. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3586–3594.
30. Chen, Y.; He, W.; Yokoya, N.; Huang, T.Z. Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition. *IEEE Trans. Cybern.* **2019**. [CrossRef]
31. Akhtar, N.; Shafait, F.; Mian, A. Bayesian sparse representation for hyperspectral image super resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3631–3640.
32. Zhu, X.X.; Grohnfeldt, C.; Bamler, R. Exploiting joint sparsity for pansharpening: The J-SparseFI algorithm. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 2664–2681. [CrossRef]
33. Wei, K.; Fu, Y. Low-rank Bayesian tensor factorization for hyperspectral image denoising. *Neurocomputing* **2019**, *331*, 412–423. [CrossRef]
34. Dong, W.; Fu, F.; Shi, G.; Cao, X.; Wu, J.; Li, G.; Li, X. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Trans. Image Process.* **2016**, *25*, 2337–2352. [CrossRef] [PubMed]
35. Han, X.H.; Shi, B.; Zheng, Y. Self-similarity constrained sparse representation for hyperspectral image super-resolution. *IEEE Trans. Image Process.* **2018**, *27*, 5625–5637. [CrossRef] [PubMed]
36. Han, X.H.; Shi, B.; Zheng, Y. Residual hsrcnn: Residual hyper-spectral reconstruction cnn from an rgb image. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 2664–2669.
37. Han, X.H.; Chen, Y.W. Deep residual network of spectral and spatial fusion for hyperspectral image super-resolution. In Proceedings of the 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), Singapore, 11–13 September 2019; pp. 266–270.

38. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O.; Benediktsson, J.A. Model-based fusion of multi-and hyperspectral images using PCA and wavelets. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 2652–2663. [CrossRef]
39. Dian, R.; Li, S.; Guo, A.; Fang, L. Deep hyperspectral image sharpening. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 5345–5355. [CrossRef]
40. Wang, Z.; Chen, B.; Lu, R.; Zhang, H.; Liu, H.; Varshney, P.K. FusionNet: An Unsupervised Convolutional Variational Network for Hyperspectral and Multispectral Image Fusion. *IEEE Trans. Image Process.* **2020**, *29*, 7565–7577. [CrossRef]
41. Qu, Y.; Qi, H.; Kwan, C. Unsupervised sparse dirichlet-net for hyperspectral image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2511–2520.
42. Zhang, L.; Nie, J.; Wei, W.; Zhng, Y.; Liao, S.; Shao, L. Unsupervised Adaptation Learning for Hyperspectral Imagery Super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
43. Nie, J.; Zhang, L.; Wei, W.; Lang, Z.; Zhang, Y. Unsupervised Alternating Optimization for Blind Hyperspectral Imagery Super-resolution. *arXiv* **2020**, arXiv:2012.01745.
44. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
45. Seeliger, K.; Güçlü, U.; Ambrogioni, L.; Güçlütürk, Y.; van Gerven, M.A. Generative adversarial networks for reconstructing natural images from brain activity. *NeuroImage* **2018**, *181*, 775–785. [CrossRef] [PubMed]
46. Liu, B.; Yu, X.; Yu, A.; Zhang, P.; Wan, G.; Wang, R. Deep few-shot learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 2290–2304. [CrossRef]
47. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
48. Yasuma, F.; Mitsunaga, T.; Iso, D.; Nayar, S.K. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE Trans. Image Process.* **2010**, *19*, 2241–2253. [CrossRef]
49. Chakrabarti, A.; Zickler, T. Statistics of real-world hyperspectral images. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 193–200.
50. Wang, J.; Kwon, S.; Shim, B. Generalized orthogonal matching pursuit. *IEEE Trans. Signal Process.* **2012**, *60*, 6202–6216. [CrossRef]
51. Wycoff, E.; Chan, T.H.; Jia, K.; Ma, W.K.; Ma, Y. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 1409–1413.
52. Scarpa, G.; Vitale, S.; Cozzolino, D. Target-Adaptive CNN-Based Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5443–5457. [CrossRef]
53. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O. Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 639–643. [CrossRef]
54. Xie, Q.; Zhou, M.; Zhao, Q.; Meng, D.; Zuo, W.; Xu, Z. Multispectral and hyperspectral image fusion by MS/HS fusion net. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 1585–1594.