Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Highlights

PoNet: A universal physical optimization-based spectral super-resolution network for arbitrary multispectral images

Jiang He, Qiangqiang Yuan*, Jie Li*, Liangpei Zhang

Information Fusion xxx (xxxx) xxx

- Generalized spectral super-resolution is presented for all multispectral imaging.
- Considering physical degradation into CNN modeling improves model interpretability.
- The proposed attention learns channel-to-channel parameters and speeds up model.
- Utilizing both deep and shallow features can further improve model performance.

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

PoNet: A universal physical optimization-based spectral super-resolution network for arbitrary multispectral images

Jiang He [a], Qiangqiang Yuan [a,*], Jie Li [a,*], Liangpei Zhang [b]

[a] School of Geodesy and Geomatics, Wuhan University, Hubei, 430079, China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, Hubei, 430079, China

ARTICLE INFO

Keywords:
Spectral super-resolution
Multispectral images
Hyperspectral images
Physical interpretability
Deep learning

ABSTRACT

Spectral super-resolution is a very important technique to obtain hyperspectral images from only multispectral images, which can effectively solve the high acquisition cost and low spatial resolution of hyperspectral images. However, in practice, multispectral channels or images captured by the same sensor are often with different spatial resolutions, which brings a severe challenge to spectral super-resolution. This paper proposed a universal spectral super-resolution network based on physical optimization unfolding for arbitrary multispectral images, including single-resolution and cross-scale multispectral images. Furthermore, two new strategies are proposed to make full use of the spectral information, namely, cross-dimensional channel attention and cross-depth feature fusion. Experimental results on five data sets show superiority and stability of PoNet addressing any spectral super-resolution situations.

1. Introduction

Hyperspectral (HS) imaging is a technique used to acquire radiation characteristics of the observed objects with a fine spectral resolution. With rich spectral information, hyperspectral images are used in many applications, such as semantic segmentation [1,2], scene classification [3–7], object detection [8,9], and target tracking [10,11]. With continuous spectrum in pixels, hyperspectral images can improve the discriminability of objects, have attracted increasing attention in many fields, for example, food science [12,13], atmosphere monitoring [14–16], medical science [17,18], and remote sensing [2–7].

Although hyperspectral images have been greatly used, the high acquisition cost and low spatial resolution hinder their development of finer applications, owing to the increase in the spatial size of sensors for each pixel when generating spectra with high signal-to-noise ratios [19]. In contrast, multispectral sensors usually capture high-spatial-resolution images with only several spectral channels, which means rich spatial detail but less spectral information. Thus, how to acquire high-resolution hyperspectral images from high-resolution multispectral images at low cost has attracted more attention. In other words, given a multispectral image, a hyperspectral image with the same spatial resolution and high spectral resolution can be obtained by increasing the channel number of the multispectral, which is called spectral super-resolution.

To solve this ill-posed problem, many researchers firstly restore hyperspectral images by utilizing sparse recovery and dictionary learning to extract the hyperspectral dictionary and sparse coefficients. Nguyen et al. [20] proposed a new training-based spectral recovery method by improving a radial basis function network with RGB white-balancing to normalize the illumination. Then, Robles-Kelly [21] employed color and appearance information to achieve spectral super-resolution through a prototype set extracted from training samples using constrained sparse coding. Arad and Ben-Shahar [22] learned a hyperspectral dictionary by K-means Singular Value Decomposition (K-SVD) and described RGB images using the projected dictionary. Jia et al. [23] applied a nonlinear dimensionality reduction technique to natural spectra and map an RGB vector to its corresponding hyperspectral vector via a manifold-based method. Inspired by their previous work in spatial super-resolution, Aeschbacher and Wu et al. [24] proposed a new shallow method for enhancing the spectral resolution of RGB images. Akhtar et al. [25] employed gaussian processes to improve the extracted dictionary through sparse representation recently. The main idea behind all the mentioned methods is extracting hyperspectral dictionary from a set of hyperspectral images and recover spectra with the coefficients calculated on multispectral images. The modeling of these methods is similar to spectral unmixing, in which the variables are all with physical meanings, the dictionary is equal to spectral

* Corresponding authors.
E-mail addresses: jiang_he@whu.edu.cn (J. He), yqiang86@gmail.com (Q. Yuan), jli89@sgg.whu.edu.cn (J. Li), zlp62@whu.edu.cn (L. Zhang).

https://doi.org/10.1016/j.inffus.2021.10.016
Received 26 March 2021; Received in revised form 22 September 2021; Accepted 31 October 2021

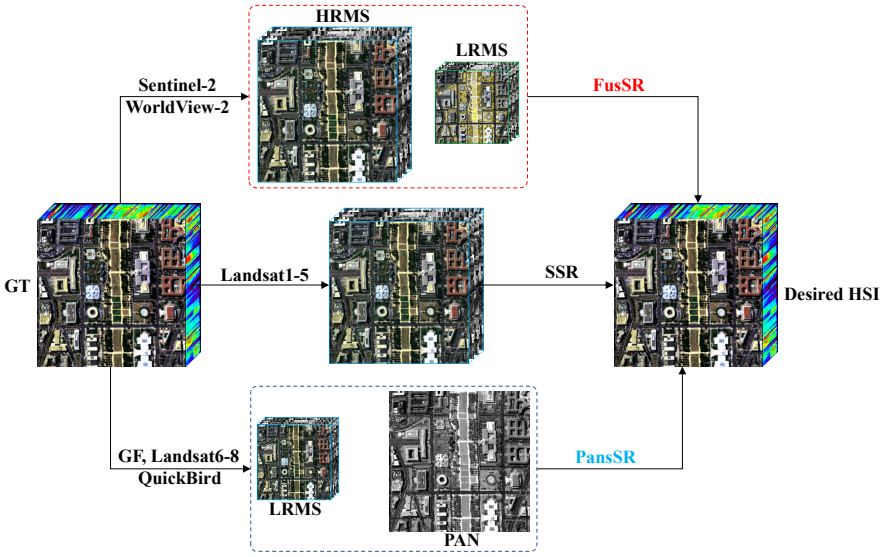


Fig. 1. Multiple cases of the degradation and recovery of hyperspectral images.

endmember and sparse coefficients are equivalent to fractional abundances. Nevertheless, spectra of observed objects cannot be represented perfectly by finite hyperspectral dictionary [26].

Instead of extracting hyperspectral dictionary, another category of methods aims to exploit the relationship between multispectral and hyperspectral images directly. Because the mapping is severely nonlinear, deep learning-based methods are employed to achieve it [27–38]. Inspired by the semantic segmentation architecture Tiramisu [39], Galliani et al. [40] proposed a deep DenseUnet with 56 convolutional layers. Further, Rangnekar et al. [41] applied a conditional adversarial framework to train CNN. Learned from the spatial super-resolution task, Xiong et al. [42] proposed an adapted network from a very deep CNN for super-resolution (VDSR) to recover hyperspectral images. To further improve results, Shi et al. [43] used dense blocks with path-widening feature fusion. Fu et al. [44] designed a spatial–spectral CNN-based method, which can jointly select the camera spectral sensitivity and learn to enhance the spectral resolution of RGB images. Also noted the importance of spectral response functions, Nie et al. [45] employed a 1×1 convolutional layer to learn it and help achieve spectral super-resolution. To show that moderate deep learning can also achieve spectral super-resolution, Can et al. [46] proposed a 9-layer residual CNN with parametric ReLU. Zhang et al. [47] currently proposed a pixel-aware deep function-mixture network with multi-scale kernels to increase the network flexibility, which can adaptively determine the receptive field size for each pixel. Notwithstanding good performance the deep learning-based methods mentioned above can achieve, they can only deal with the input images without spatial degradation. In practical application, however, channels or images captured by the same satellite are often with different spatial resolutions, which owes to the imaging process of devices, such as Sentinel-2, WorldView-2, Gaofen-1, and Gaofen-2, namely cross-scale multispectral images. On the contrary, images consisting of channels with same spatial resolution are called single-resolution images in this paper.

There has been very little research considering spatial degradation in spectral super-resolution. Mei et al. [48] obtain high-spatial-resolution (HR) and high-spectral-resolution images from low-spatial-resolution (LR) multispectral images using two similar CNNs for two stages. Although their method can enhance the spatial details as well as spectral resolution, nevertheless they directly stack convolutional layers without any physical meanings to find a mapping function between input and output images. Furthermore, they only used images with one spatial resolution, while as mentioned before, there are many images with different spatial resolutions (lower or higher than that of used channels) obtained even by the same satellite.

To make full use of images at different scales and consider physical degradation in spectral super-resolution based on deep learning, a deep physical optimization-based CNN (PoNet) which can address generalized spectral super-resolution for single-resolution or cross-scale multispectral images is proposed. As shown in Fig. 1, generalized spectral super-resolution includes traditional spectral super-resolution (SSR) and spectral super-resolution on cross-scale multispectral images. Concentrating on spectral super-resolution with cross-scale spectral information, there are two subproblems need to be solved, one is how to use auxiliary lower-resolution channels with spectral information covering different wavelength to optimize spectral super-resolution results (FusSR), and the other is how to address joint enhancement of spatial and spectral resolution by introducing more high-resolution spatial details from panchromatic images (PansSR). The proposed PoNet can handle them well. The main contributions of this paper can be summarized as follows:

- For the first time, we define and address generalized spectral super-resolution for single-resolution or cross-scale spectral information, in other words, SSR, FusSR, and PansSR. All the cases mentioned will be discussed in experiments.
- Considering physical degradation into CNN modeling, PoNet can better exploit cross-scale spectral information and reconstruct hyperspectral images more finely. Model construction following the dataflow of optimization algorithms gives networks physical interpretability to help people better understand how CNN works.
- To further improve the model performance, cross-dimensional channel attention achieved the parameter learning channel-to-channel as well as reducing the number of parameters is proposed in this paper. Furthermore, we also employ cross-depth feature fusion to ensure the effective utilization of deep and shallow features.

The remaining part of this paper is organized as follows. Section 2 derives the spectral super-resolution algorithm considering spatial degradation, and then introduces the proposed PoNet in detail. Data used in this article are introduced in Section 3. Section 4 presents some experimental results on a natural image data set to verify the reliability of the proposed model. Then, PoNet is compared with other methods in two cases mentioned above. Finally, we draw some conclusions in Section 5.

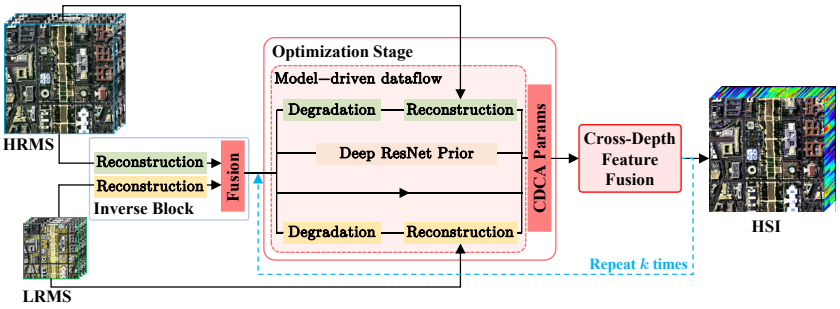


Fig. 2. Framework of the proposed PoNet.

2. Method

Considering spatial degradation between multispectral and hyperspectral imaging modes, the observation model is proposed at first. Based on this model, SSR in cross-scale images is formulated and optimized in the variational model-based algorithm. By unrolling optimization algorithms to deep learning, PoNet is depicted as shown in Fig. 2. Cross-scale multispectral images are fed into the inverse imaging block and fusion layer to reconstruct initial results. Then, several recurrent optimization stages are employed to obtain better spectral information. Furthermore, among different stages, a strategy named cross-depth feature fusion is proposed to use deep and shallow features.

2.1. Model formulation

Let $X \in \mathbb{R}^{W\times H\times C}$ represent the desired hyperspectral images, where C is the number of the spectral channels, and W and H are the width and height. $M_H \in \mathbb{R}^{W\times H\times c_1}$ represents the multispectral images with the same spatial resolution as X but only c_1 bands. $M_L \in \mathbb{R}^{w\times h\times c_2}$ means the low-spatial-resolution multispectral images with the size of $w \times h$ and c_2 channels. Varying in spectral response functions, sensors obtain different MS or HS data with different bands. A transformation matrix Φ can be used to describe the spectral degradation between MS and HS imaging, $\Phi_1 \in \mathbb{R}^{C\times c_1}$ for HRMS and $\Phi_2 \in \mathbb{R}^{C\times c_2}$ for LRMS. Furthermore, regarding the spatial degradation between HSI and LRMS as matrix $D^{(w\times h)\times(W\times H)}$, the observation model is shown as follows.

$$M_H = X\Phi_1 \tag{1}$$

$$M_L = DX\Phi_2 \tag{2}$$

According to (1) and (2), the relationship between MSIs and HSIs is illuminated. The spectral recovery task obtaining HR HSIs from LRMS and HRMS image pairs is an inverse problem, which is ill-posed and often formulated as an optimization problem that minimizes an energy function, i.e.,

$$\hat{X} = \arg\min_X \frac{1}{2}\left\|M_H - X\Phi_1\right\|_2^2 + \frac{1}{2}\left\|M_L - DX\Phi_2\right\|_2^2 + \lambda\mathcal{R}(X) \tag{3}$$

where $\mathcal{R}$ is a regularizer that imposes prior knowledge, $\|\cdot\|_2$ means the Euclidean norm of data-fidelity terms, and λ is a trade-off parameter between the regularizer and data-fidelity terms. Although the energy function can ensure consistency between the reconstructed image and observed images as well as introduce image prior, it is hard to directly solve this optimization problem. To separate the regularizer and data-fidelity terms in (3) and further solve this minimization problem, variable splitting technique is employed. Introducing auxiliary variable Z, the optimization problem (3) is constrained as

$$\hat{X} = \arg\min_X \frac{1}{2}\left\|M_H - X\Phi_1\right\|_2^2 + \frac{1}{2}\left\|M_L - DX\Phi_2\right\|_2^2 + \lambda\mathcal{R}(Z),\ \text{s.t. } Z = X \tag{4}$$

Employing the half-quadratic splitting method, a new cost function is derived:

$$\mathcal{L}_\mu(X,Z) = \frac{1}{2}\left\|M_H - X\Phi_1\right\|_2^2 + \frac{1}{2}\left\|M_L - DX\Phi_2\right\|_2^2 + \mu\left\|Z - X\right\|_2^2 + \lambda\mathcal{R}(Z) \tag{5}$$

where μ is a penalty parameter. By half-quadratic splitting method, optimization problem (3) can be split into two subproblems which could be solved more easily and efficiently,

$$\begin{cases} \hat{X} = \arg\min_X \frac{1}{2}\left\|M_H - X\Phi_1\right\|_2^2 + \frac{1}{2}\left\|M_L - DX\Phi_2\right\|_2^2 + \mu\left\|Z - X\right\|_2^2, \\ \hat{Z} = \arg\min_Z \frac{1}{2}\left\|Z - X\right\|_2^2 + \frac{\lambda}{\mu}\mathcal{R}(Z) \end{cases} \tag{6}$$

Considering the X-subproblem, an approximate solution can be updated by following the gradient descent algorithm:

$$\begin{aligned} X_{k+1} &= X_k - \epsilon\left(D^T DX_k\Phi_2\Phi_2^T - D^T M_L\Phi_2^T + X_k\Phi_1\Phi_1^T\right. \\ &\quad \left. - X_H\Phi_1^T + \mu X_k - \mu Z_k\right) \\ &= (1-\epsilon\mu)X_k - \epsilon D^T DX_k\Phi_2\Phi_2^T - \epsilon X_k\Phi_1\Phi_1^T \\ &\quad + \epsilon D^T M_L\Phi_2^T + \epsilon M_H\Phi_1^T + \epsilon\mu Z_k \end{aligned} \tag{7}$$

where ϵ is the optimization stride. As for the Z-subproblem, proximal operators that impose prior knowledge can all handle it, which is defined by

$$\hat{Z}_{k+1} = \mathrm{Prox}(X_{k+1}) = \arg\min_Z \frac{1}{2}\left\|Z - X_{k+1}\right\|_2^2 + \frac{\lambda}{\mu}\mathcal{R}(Z) \tag{8}$$

With the help of half-quadratic splitting and gradient descent algorithm, problem (3) can be solved by alternately updating (7) and (8).

2.2. The proposed PoNet

As mentioned above, we can solve spectral super-resolution problem by following (7) and (8). Nonetheless, alternating updating two variables requires manual intervention that affects the quality of the reconstruction results greatly, which limits the application of optimization algorithms in SSR.

Instead of solving a specific physical model, deep learning-based methods are famous for their strong ability to learn a nonlinear mapping implicitly. In this paper, we are devoted to unrolling the physical optimization algorithm to a learnable end-to-end CNN, namely, the PoNet.

As shown in Fig. 3, given M_H and M_L, the initial X_0 is first restored in the inverse block through different reconstruction streams:

$$M_0^{H\uparrow} = \mathrm{ReLU}(W_0^H * M_H + b_0^H) \tag{9}$$

$$M_0^{L\uparrow} = \mathrm{ReLU}(W_0^L * \mathrm{Up}(M_L) + b_0^L) \tag{10}$$

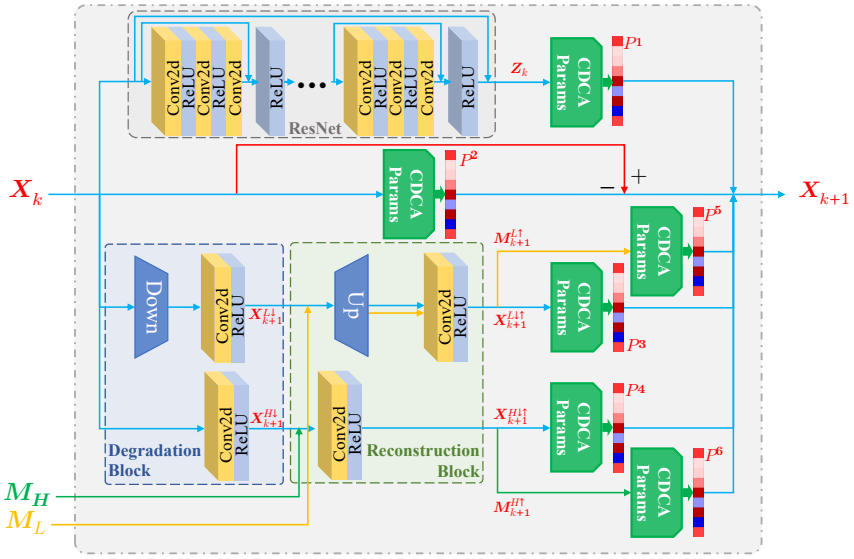


Fig. 3. Optimization stage of the proposed PoNet.

where $\mathrm{ReLU}(x) = \max\{0, x\}$, denoting the rectified linear unit. $\mathrm{Up}(\cdot)$ means the up-sampling operator. W_0^H and b_0^H represent the kernel weights and biases for high-resolution multispectral imaging in initialization, and W_0^L and b_0^L are for low-resolution multispectral images. Then, to improve the coarse reconstruction results by simultaneously using cross-scale spectral information, a convolutional layer with the kernel size of 1×1 is employed to integrate two rough results:

$$X_0 = \mathrm{ReLU}(W_0^F * \mathrm{Concat}(M_0^{H\uparrow}, M_0^{L\uparrow}) + b_0^F) \tag{11}$$

where W_0^F and b_0^F presents the kernel weights and biases for the fusion layer. $\mathrm{Concat}(\cdot)$ means the operation to concatenate input features. Subsequently, X_0 will be fed into optimization stages to obtain the final optimized results.

In every stage, we design four subnetworks consists of four modules to learn spectral optimization, as shown in Fig. 3, each of them is consistent with the matrix in the optimization-based algorithm. To show the optimization stages more clearly, (7) is reformulated into:

$$\hat{X}_{k+1} = (1 - \epsilon\mu)\, X_k + \epsilon\mu Z_k - \epsilon X^{L\downarrow\uparrow}_{k+1} - \epsilon X^{H\downarrow\uparrow}_{k+1} + \epsilon M^{L\uparrow}_{k+1} + \epsilon M^{H\uparrow}_{k+1} \tag{12}$$

where $X^{L\downarrow\uparrow}_{k+1} = D^T D X_k \Phi_2 \Phi_2^T$ and $X^{H\downarrow\uparrow}_{k+1} = X_k \Phi_1 \Phi_1^T$ represent the degradation–reconstruction processes corresponding to LR and HR multispectral imaging. $M^{H\uparrow}_{k+1} = M_H \Phi_1^T$ and $M^{L\uparrow}_{k+1} = D^T M_L \Phi_2^T$ mean the reconstructions for the original multispectral images.

Specifically, given X_k, two streams in degradation block are employed to learn degenerated images:

$$X^{H\downarrow}_{k+1} = \mathrm{ReLU}(W^{H,1}_{k+1} * X_k + b^{H,1}_{k+1}) \tag{13}$$

$$X^{L\downarrow}_{k+1} = \mathrm{ReLU}(W^{L,1}_{k+1} * \mathrm{Down}(X_k) + b^{L,1}_{k+1}) \tag{14}$$

where $W_k^{H,n}$ and $b_k^{H,n}$ mean the kernel weights and biases for high-resolution multispectral imaging in the kth stage. Moreover, $W_k^{L,n}$ and $b_k^{L,n}$ are for low-resolution multispectral imaging. $\mathrm{Down}(\cdot)$ presents the downsampling operator. The convolutional kernel size in degradation block is set to 1×1. According to the optimization-based algorithm, they will be restored again through the reconstruction block, which is similar to the inverse block:

$$X^{H\downarrow\uparrow}_{k+1} = \mathrm{ReLU}(W^{H,2}_{k+1} * X^{H\downarrow}_{k+1} + b^{H,2}_{k+1}) \tag{15}$$

$$X^{L\downarrow\uparrow}_{k+1} = \mathrm{ReLU}(W^{L,2}_{k+1} * \mathrm{Up}(X^{L\downarrow}_{k+1}) + b^{L,2}_{k+1}) \tag{16}$$

For the two convolutional layers employed to recover high-resolution spectral information, kernel size is set as 3×3 to gain a large receptive field. Then, the reconstruction blocks are also applied to the original multispectral images to compute $M^{H\uparrow}_{k+1}$ and $M^{L\uparrow}_{k+1}$, which is similar to (9) and (10):

$$M^{H\uparrow}_{k+1} = \mathrm{ReLU}(W^{4,2}_{k+1} * M_H + b^{4,2}_{k+1}) \tag{17}$$

$$M^{L\uparrow}_{k+1} = \mathrm{ReLU}(W^{3,2}_{k+1} * \mathrm{Up}(M_L) + b^{3,2}_{k+1}) \tag{18}$$

Having dealt with the X-subproblem, the question now is how to update Z_k with hyperspectral image prior, in other words, how to solve (8). Traditional optimization-based algorithms can explicitly represent prior, which is too limited to explain hyperspectral image completely. To solve Z-subproblem implicitly is equivalent to solve:

$$\hat{Z}_k = \mathrm{Prox}(X_k) = \arg\min_Z \frac{1}{2} \|Z - X_k\|_2^2 + \frac{\lambda}{\mu} \mathcal{R}(Z) \tag{19}$$

As described in [49], (19) can be rewritten as

$$\mathrm{Prox}(X_k) = \arg\min_X \frac{1}{2(\sqrt{\lambda/2\mu})^2} \|Z - X_k\|_2^2 + \mathcal{R}(Z) \tag{20}$$

With the mathematical equivalence to the regularized denoising, the proximal operator $\mathrm{Prox}(\cdot)$ can be replaced by any existing denoisers $\mathrm{Denoiser}(\cdot)$ with a noise level of $\sqrt{\lambda/2\mu}$. In this paper, we employed deep residual networks as the proximal operator:

$$\hat{Z}_k = \mathrm{Prox}(X_k) = \mathrm{Denoiser}(X_k) = ResNet(X_k) \tag{21}$$

In this formulation, the hyperspectral image prior $\mathcal{R}(\cdot)$ can be implicitly learned by CNN, which opens a convenient and efficient door to optimize spectral information for spectral recovery.

2.2.1. Cross-dimensional channel attention

After updating intermediate variables, there are many other hyperparameters to be defined, such as ϵ and μ. In traditional physical optimization-based algorithms, hyperparameters need to be defined manually and adjust to the optimal through a large number of experiments. Furthermore, in spectral super-resolution, differential treatment should be performed for the hyperparameters of different channels due to the different radiation characteristics.

In modern deep learning-based algorithms, channel attention mechanism [50] has attracted many sights of scholars for strong channel adaptive weighting ability. However, pooling is a common operation used in traditional channel attention, which is popular for fast computation and no parameter requirement at the cost of high information

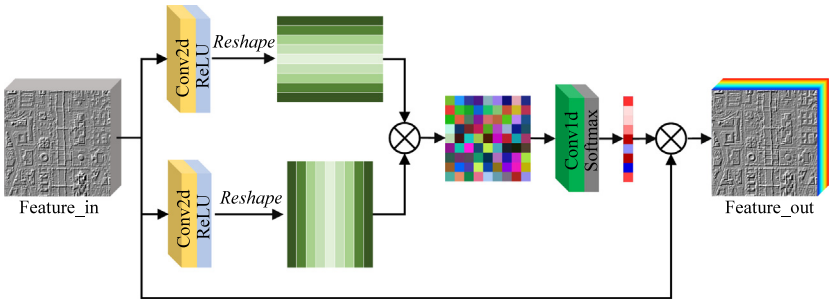


Fig. 4. Cross-dimensional channel attention.

loss. Furthermore, traditional channel attention weights the different channels of features separately ignoring the interaction between channels. Within the knowledge based on [51], it states that building relationships between any two channels is much of importance. Besides, when the number of channels is large and attention mechanisms are frequently employed, the problem of computational burden should also be focused on.

Inspired by the above-mentioned points, we proposed a strategy named *Cross-Dimensional Channel Attention* (CDCA) employing 1D and 2D convolutional layers to manage the hyperparameter learning in this paper. 2D convolutional layers are used to extract pixel-by-pixel attention maps. On the other hand, 1D convolutional layers are employed to integrate attention maps for fast computational speed. Details of the proposed module are shown in Fig. 4.

We adopt two 2D convolutional layers with the kernel size of 1×1 to extract different spectral features $R, S \in \mathbb{R}^{W \times H \times C}$. Attention map $A \in \mathbb{R}^{C \times C}$ between any two channels can be calculated as follows:

$$a_{ij} = R_i S_j^T \tag{22}$$

where a_{ij} measures the attention between the ith and jth bands. R_i and S_j is the reshaped channel. To boost the computational speed, we employ a 1D convolutional layer with the kernel size of k to integrate channel-to-channel attention map A. Then, the final cross-dimensional channel attention-based hyperparameter $P \in \mathbb{R}^{1 \times C}$ will be obtained after a softmax layer:

$$p_j = \frac{\exp\left(W^{1\mathrm{d}} * A_j + b^{1\mathrm{d}}\right)}{\sum_{j=1}^{C} \exp\left(W^{1\mathrm{d}} * A_j + b^{1\mathrm{d}}\right)} \tag{23}$$

where W^{1d} and b^{1d} mean the kernel weights and biases for the 1D convolutional layer, and $p_j \in P$ is the parameter for the jth band. Obtaining the cross-dimensional channel attention-based hyperparameters for each intermediate variable, we update $\hat{X}_{k+1}$ as follows:

$$\begin{aligned}\hat{X}_{k+1}^{j} = {} & X_k^j + p_{j,k+1}^1 Z_k^j - p_{j,k+1}^2 X_k^j - p_{j,k+1}^3 X_{k+1}^{L\downarrow\uparrow,j} \\ & - p_{j,k+1}^4 X_{k+1}^{H\downarrow\uparrow,j} + p_{j,k+1}^5 M_{k+1}^{L\uparrow,j} + p_{j,k+1}^6 M_{k+1}^{H\uparrow,j}\end{aligned} \tag{24}$$

where $\hat{X}_{k+1}^{j}$ presents the jth band of $\hat{X}_{k+1}$ and $p_{j,k+1}^{n}$ means the hyperparameter learned for the jth channel of 6 variables in (24).

In this way, we build a learnable end-to-end CNN by unrolling the physical optimization algorithm, which keeps the advantages of deep learning and physical model-based algorithm.

2.2.2. Cross-depth feature fusion

In deep learning-based algorithms, the depth makes much sense for the network effect, in other words, the deeper networks get the better results. However, there have been much research to confirm that shallow features are also very important [52–54]. In the proposed PoNet, we get multiple updated results at different depths. To improve the model memory of shallow features, a strategy named *Cross-Depth Feature Fusion* (CDFF) is proposed as shown in Fig. 5.

Given a set of intermediate results $\{X_0, X_1, X_2, \ldots, X_{k-1}\}$, the PoNet firstly concatenates results at different depths:

$$F_{k-1}^{C} = \mathrm{Concat}(X_0, X_1, X_2, \ldots, X_{k-1}) \tag{25}$$

Similar to the generation of X_0, as indicated in (11), a convolutional layer with ReLU is also employed to fuse cross-depth features to obtain the input for the next stage.

$$X_k^{In} = \mathrm{ReLU}(W_{k-1}^{F} * F_{k-1}^{C} + b_{k-1}^{F}) \tag{26}$$

where X_k^{In} means the input feature for the kth optimization stage. Acquiring various information from cross-depth features, X_k^{In} can represent more comprehensive spectral information from shallow and deep features, which is beneficial to the subsequent optimization.

3. Data

In this section, we will introduce two situations of the spectral super-resolution considering spatial degradation, including FusSR and PansSR. FusSR is utilizing more auxiliary lower-resolution spectral channels to obtain better spectral recovery, and PansSR is to handle the joint enhancement of spatial resolution and spectral resolution with the help of high-resolution panchromatic images.

3.1. FusSR

Because there are multiple spatial resolutions in the images captured by the same sensor, we can get different channels with various spatial resolutions, such as Sentinel-2 and WorldView series, etc. In this paper, to deal with FusSR problem, we simulated four types of spectral differences between sensors, namely RgB2CAVE, Sen2OHS, Sen2CHRIS, and RgB2CASI, which involve cooled CCD camera, Sentinel-2, *Orbita Hyperspectral Satellites* (OHS), *Compact High Resolution Imaging Spectrometer* (CHRIS), and *Compact Airborne Spectrographic Imager* (CASI). Details are shown below.

3.1.1. RgB2CAVE

For this case, we used the CAVE data set, as shown in Fig. 6, which comprises 32 scenes with a size of 512×512, which is a popular hyperspectral image data set in HSI processing. All the hyperspectral images in CAVE data set are captured by a cooled CCD camera named *Apogee Alta U260* and cover the spectral range from 400 nm to 700 nm with a 10 nm spectral resolution containing 31 bands. Moreover, the RGB images covering the same scene as hyperspectral data are available.

To simulate the FusSR situation, the Green channel of the original RGB image is spatially downsampled with a ratio of 1/2. Thus, in RgB2CAVE, solving FusSR problem is to obtain 31-band hyperspectral images from the degraded RGB images consist of a Red channel, a Blue channel, and a low-resolution Green channel.

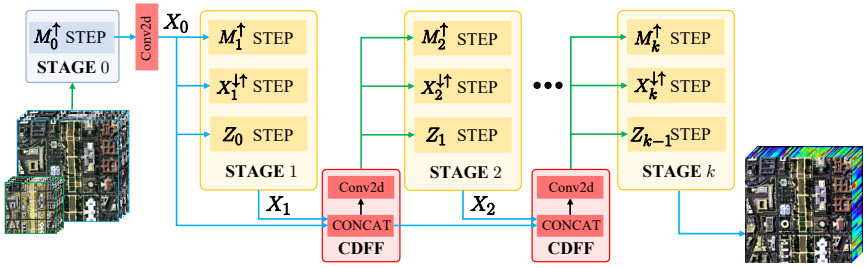


Fig. 5. Cross-depth feature fusion strategy.

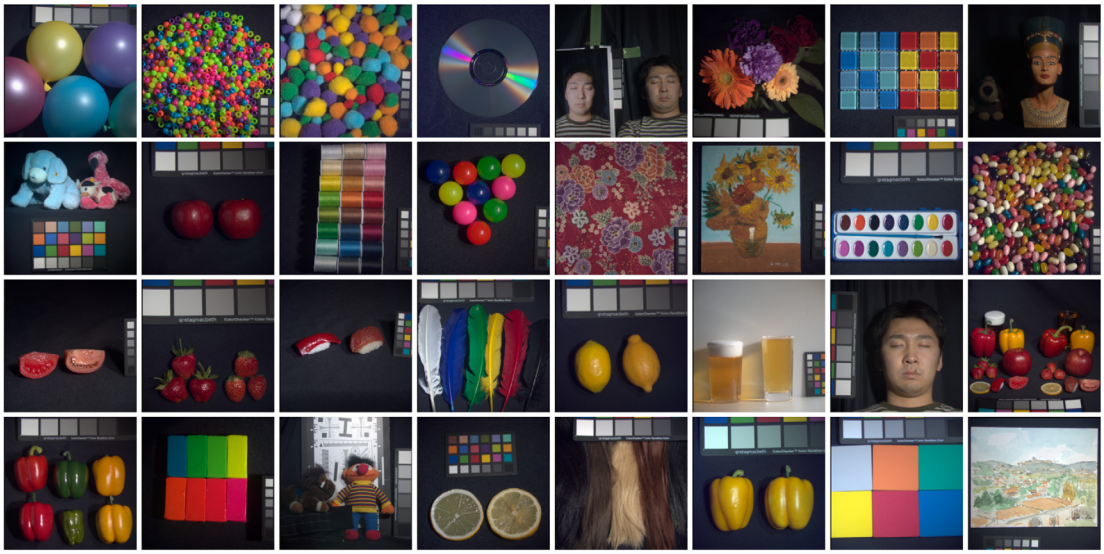


Fig. 6. Original RGB images in CAVE data set.

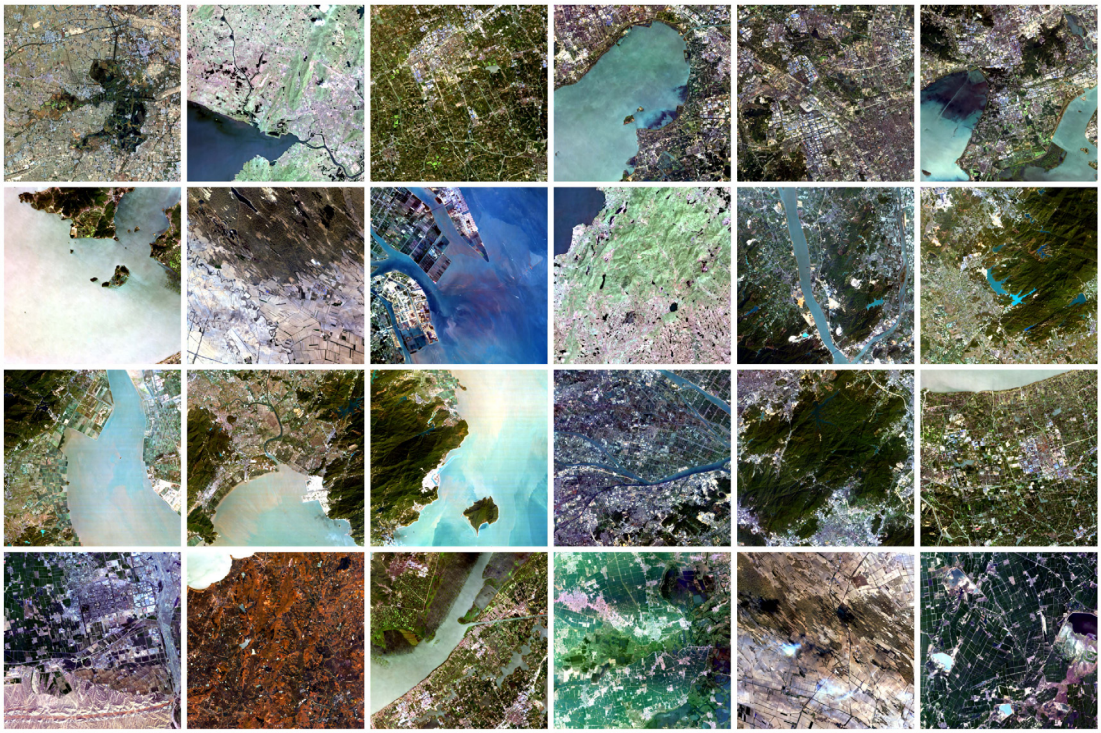
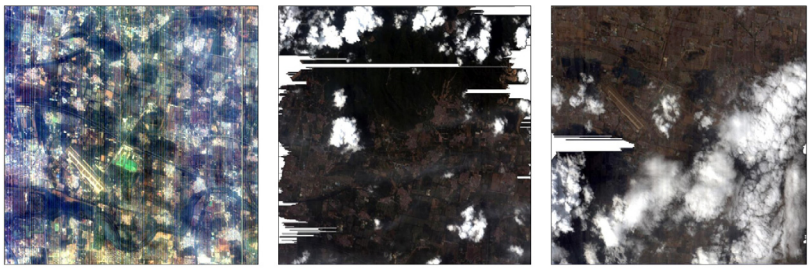


Fig. 7. Parts of multispectral images in Sen2OHS data set.



(a) (b) (c)

Fig. 8. Three limits of hyperspectral images acquired by CHRIS. (a) Hybrid noises. (b) Stripe gaps. (c) High cloud cover.

3.1.2. Sen2OHS

The second situation is recovering hyperspectral images with the same spectral resolution as captured by OHS from multispectral images sensed by Sentinel-2. Images from Chinese Orbita hyperspectral satellites are with 10 m spatial resolution, which are captured in the spectral range from 400 nm to 1000 nm with 2.5 nm increments. Note that the hyperspectral data sent to users are sampled to 32 bands. The rich spectral information in OHS data is of considerable importance for application. However, free OHS data are mostly unavailable because of commerciality. The unavailability limits the hyperspectral data sources for researchers.

Sentinel-2 provides images with spatial resolutions of 10, 20, and 60 m. Besides, the revisit time in the same measured area is 5 days. With good spatial resolution and availability for free, multispectral images sensed by Sentinel-2 become more and more popular. In this case, we aim to achieve FusSR between multi-resolution Sentinel-2 and OHS data. Note that the 60 m channels captured by Sentinel-2 dedicated to atmospheric correction and cloud screening were excluded from the experiments.

In this paper, Sentinel-2 multispectral images are simulated from OHS data by using Hysure [55] to reduce the errors caused by geographic registration and the inconsistency of acquiring time. Furthermore, 6000 hyperspectral images with a size of 128×128 are selected for training from the *Competition in Hyperspectral Remote Sensing Image Intelligent Processing Application*.[1] Parts of them are shown in Fig. 7.

3.1.3. Sen2CHRIS

Compact High Resolution Imaging Spectrometer, namely CHRIS, is carried on *PRoject for On Board Autonomy 1* (PROBA-1) satellite and provides hyperspectral images with five modes. The first mode acquires images with 62 spectral channels from 406 nm to 1003 nm with a spatial resolution of 34 m, which can obtain the most spectral information but the poorest spatial details. Hence, the third case is obtaining CHRIS data from Sentinel-2 data.

However, as shown in Fig. 8, images sensed by CHRIS are always accompanied by some problems, including hybrid noise, stripe gap, and high cloud cover. Furthermore, all satellite passes are systematically acquired according to a fixed acquisition plan. Observation over a new specific area should be performed by submitting the request to add a new site to the acquisition plan, which costs much money and time.

To overcome these problems, experiments are carried on with three freely available data subsets, including Xiong'an, Washington DC Mall, and Chikusei, by downsampling the spectral channels of free hyperspectral data to the same of CHRIS and Sentinel-2 using Hysure. Three data subsets are shown in Fig. 9.

Xiong'an data subset is an aerial image covered a $1.8\ \text{km} \times 0.8\ \text{km}$ rural area in Matiwan Village, Xiong'an New Area, China [56]. The spectral range of Xiong'an data subset is 400 to 1000 nm, with a spatial resolution of 0.5 meters and 250 bands. Moreover, the coverage of it is 3750×1580. Washington DC Mall data subset [57] was acquired using HYDICE airborne sensor and with a size of $1280 \times 307 \times 210$, covering the spectral wavelength from 400 to 2500 nm, and the spatial resolution is lower than Xiong'an and close to Chikusei. The image in Chikusei data subset [58] was taken by the Headwall Hyperspec-VNIR-C imaging sensor over agricultural and urban areas in Chikusei, Japan, with a size of 2517×2335. It contains 128 spectral bands ranging from 363 to 1018 nm with a spatial resolution of 2.5 m.

With the help of spectral resampling, we constructed our Sen2CHRIS data sets using the free hyperspectral images mentioned above, where hyperspectral images are with 62 bands while multispectral images are resampled to four HR bands and four LR bands in the range of 400 to 1000 nm similar to Sen2OHS data set.

3.1.4. RgB2CASI

Compact Airborne Spectrographic Imager is an airborne hyperspectral all-in-one imaging system produced by the Canadian company ITRES, which is a visible near-infrared (VNIR, range from 380 to 1050 nm) sensor that offers up to 288 bands of crisp visible–near infrared goodness for a wide variety of environmental, forestry, agricultural, optical water quality and wetlands applications.

Fig. 10 displays an image in the data set of 2018 IGARSS Data Fusion Contest[2] acquired by CASI, which includes a hyperspectral image with 48 bands and a corresponding RGB image.

Similar to RgB2CAVE data set, the Green channel of the original RGB image is also spatially downsampled in RgB2CASI data set. Thus, for this data set, the goal of FusSR is to restore a 48-band hyperspectral image from the degraded RGB images consist of a Red channel, a Blue channel, and a low-resolution Green channel.

3.2. PansSR

As mentioned before, there is another situation that using cross-scale multispectral images to address spectral super-resolution, i.e., PansSR. PansSR is to achieve the joint enhancement of spatial resolution and spectral resolution using auxiliary higher-resolution spectral channels, usually, panchromatic images. In this case, we constructed a data set named GF2Hyper to show the generalization ability of the proposed model, which involves Panchromatic/Multispectral imager carried on Gaofen-1 and Hyperion sensor on EO-1 satellite.

Gaofen-1 is a Chinese satellite employed two imagers, which can obtain 8 m resolution multispectral images with four bands and a 2 m resolution PAN image that covered the full spectral ranges of corresponding multispectral images, from 450 to 900 nm. Images obtained by Hyperion are with a spatial resolution of 30 meters and obtain 242 bands from 357 to 2567 nm. However, only 196 bands of them can be used, as 44 bands are set to zero and two bands redundant.

Furthermore, in this data set, because of the large spectral range gap between Gaofen-1 and Hyperion, we select 63 hyperspectral bands covering the same spectral range from 450 to 900 nm. Fig. 11 shows

[1] The data set can be download at http://doi.org/10.5281/zenodo.5642597.

[2] Data can be download at http://dase.grss-ieee.org/.

Fig. 9. True color synthesis of multispectral images in Sen2CHRIS data set. (a) Xiong'an data subset. (b) Chikusei data subset. (c) Washington DC Mall data subset.

Fig. 10. RGB images in 2018 IGARSS data fusion contest.

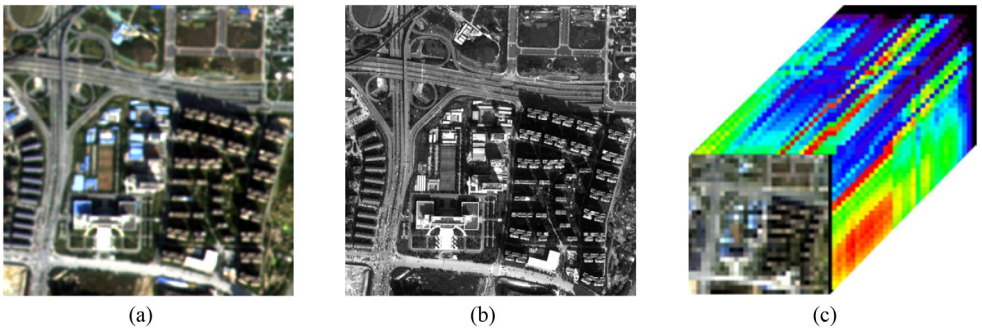


Fig. 11. Difference between Gaofen-1 data and Hyperion data. (a) MSI acquired by Gaofen-1. (b) PAN acquired by Gaofen-1. (c) HSI acquired by Hyperion.

Table 1
Details of all proposed data sets, including the size of input, the size of output, and the percentage of images for training and test, where $n \times c \times w \times h$ means there are n image patches have c channels with the size of $w \times h$.

	HR Input	LR Input	Output	Training	Test
RgB2CAVE	$32 \times 2 \times 128 \times 128$	$32 \times 1 \times 64 \times 64$	$32 \times 31 \times 128 \times 128$	75%	25%
Sen2OHS	$6000 \times 4 \times 128 \times 128$	$6000 \times 4 \times 64 \times 64$	$6000 \times 32 \times 128 \times 128$	75%	25%
Sen2CHRIS	$2838 \times 4 \times 128 \times 128$	$2838 \times 4 \times 64 \times 64$	$2838 \times 62 \times 128 \times 128$	75%	25%
RgB2CASI	$1904 \times 2 \times 128 \times 128$	$1904 \times 1 \times 64 \times 64$	$1904 \times 50 \times 128 \times 128$	75%	25%
GF2Hyper	$1152 \times 1 \times 128 \times 128$	$1152 \times 4 \times 32 \times 32$	$1152 \times 63 \times 128 \times 128$	75%	25%

the difference between Gaofen-1 data and Hyperion data. PAN and multispectral images are used as input, and the objective of PansSR in this data set is to recover 63-band hyperspectral Hyperion data with the high spatial resolution of panchromatic images. To ensure the data set can be quantitatively evaluated, multispectral data are simulated from Hyperion data. Thus, the spatial resolution of PAN images is 30 m and that of multispectral images is degraded to 120 m. In this way, we could calculate quantitative results using the original Hyperion data as ground truth, which meets the Wald protocol [59].

4. Experiments & results

In this section, we show some experiments to verify the superiority of the proposed PoNet, including experiments at different multispectral imaging situations, analysis of sampling operators, and ablation study. Details are as follows.

4.1. Experimental setting

4.1.1. Comparison methods

Because it is an early attempt to propose a universal spectral super-resolution model for arbitrary multispectral imaging situations, including SSR, FusSR, and PansSR, we draw the comparison experiments with the existing SSR algorithms, including DenseUnet [40], CanNet [46], HSCNN+ [43], and our previous work, HSRnet [60], which are all based on deep learning. Although there are some dictionary learning-based algorithms, CNN shows significant improvement over sparse recovery [47].

4.1.2. Quantitative metrics

Five quantitative quality metrics, including correlation coefficient (CC), mean peak signal-to-noise ratio (mPSNR) in decibel units, mean structural similarity (mSSIM) [61], spectral angle mapper (SAM) [62] in degree, and Erreur Relative Global Adimensionnelle de Synthèse (ERGAS), are used to evaluate the performance of all comparison methods quantitatively. CC, mPSNR, and mSSIM are indexes assess the spatial fidelity of the reconstructed hyperspectral images, which are computed on each channel and averaged over all bands. Results with large values indicate that the method is more effective for maintaining spatial detail. Moreover, SAM evaluates the spectral distortion, the better spectral fidelity and the lower SAM. ERGAS computes the errors of cross-scale image reconstruction as an omnibus index, and the lower ERGAS also illustrates the better performance.

4.1.3. Implementation detail

Details of all used data sets are shown in Table 1, including the size of input and output, the percentage of images for training and test. As shown in Table 1, the input data involves HR images and LR images. For SSR, we use only HR images in RgB2CAVE, Sen2OHS, Sen2CHRIS, and RgB2CASI data sets, while for FusSR, we use both HR and LR images in the four data sets mentioned above. Moreover, GF2Hyper is only for PansSR.

We set the number of optimization stages to 9, which shows the best restoration effect among the following experiments. Moreover, the optimization algorithm based on adaptive estimates of low-order moments (Adam) is employed to train PoNet with the learning rate set to 0.001. The models are trained by Pytorch framework running in the Windows 10 environment with 16 GB RAM and one Nvidia RTX 2080 GPU. Besides, we have also tried our best to adjust hyperparameters to train the comparative models optimal.

4.2. Results on RgB2CAVE data set

On RgB2CAVE data set, we compared all methods in two situations of spectral super-resolution, SSR and FusSR. FusSR is exploiting HR Red, LR Green, and HR Blue channels to recover 31-channel images. On the contrary, SSR means utilizing HR channels only.

Table 2 lists quantitative comparisons of different methods, where physical optimization-based models show certain improvement than traditional CNN. Besides, with more spectral information, in FusSR, models all recover hyperspectral images with a better quality showing an average of 0.0173 increase in CC, 11.05% improvement in mPSNR, 0.0198 increase in mSSIM, and above 25.68% reduction in SAM. It illustrates that more spectral information can truly help spectral super-resolution, even they are low-resolution. In SSR, although PoNet gets a decrease in CC and mSSIM, the spectral maintaining is still the best, however, traditional CNNs are not far behind. With the help of the LR Green channel, PoNet can restore more hyperspectral features in optimization stages and integrate them with the help of physical degradation. The tremendous improvement can be realized no matter on spectral maintaining or spatial fidelity, where the amelioration of SAM is even up to 26.94%.

The error maps between the restored results and ground-truth in two situations are also visualized in Figs. 12 and 13. Seeing only Fig. 12, PoNet gets the lowest error maps than other models, which shows its advantage in addressing traditional spectral super-resolution. CanNet and HSCNN+ get higher error on background. For all algorithms, the problem is results of 540 nm are not good, which may be due to the spectral information loss of Green channel.

With the help of LR Green channel, error maps shown in Fig. 13 all decrease, which validates the effectiveness of low-resolution spectral information. Especially, in 590 nm band, the error map of peppers is down to almost 0. Compared with HSRnet, PoNet can better hold the balance between background and the observed object showing lower errors on both peppers and small squares. Finally, among all results, error maps calculated on the PoNet's results show the least values, which illustrates the significant superiority of the proposed PoNet.

4.3. Results on Sen2OHS data set

Performance in SSR and FusSR are also compared on Sen2OHS data set and the quantitative comparisons are listed in Table 3. Using only multispectral channels with high resolution including band 2, 3, 4, and 8, PoNet and HSRnet, which consist of physical interpretability-based optimization stages, get much better results than other CNNs.

Making full use of the multispectral information in Sen2OHS data set, that is, using band 2, 3, 4, 5, 6, 7, 8, and 8a, we report quantitative results in FusSR in Table 3. The proposed PoNet achieves the best performance not only in CC, mPSNR, and mSSIM, but also in SAM and ERGAS. Note that, the selected text images cover different objects including various vegetations, buildings, rivers, and farmlands.

The outcomes obtained by performing SSR on Sen2OHS data set are displayed in Fig. 14, where six images are selected. The error maps of

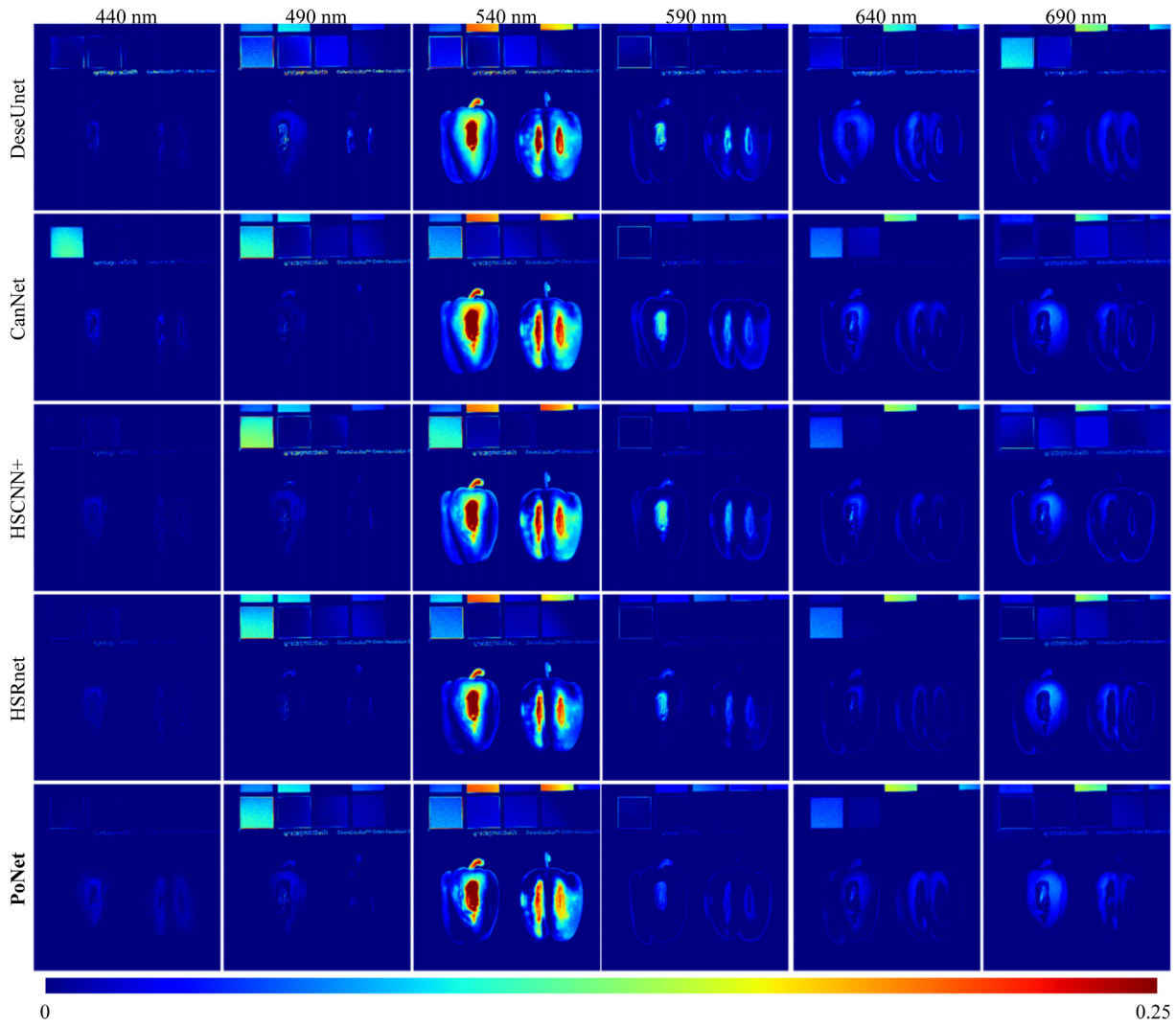


Fig. 12. Visualization of the error maps between ground-truth and model results on RgB2CAVE data set with only Red and Blue channels, i.e., plain SSR. From left to right, reconstructed bands of wavelength 440, 490, 540, 590, 640, and 690 nm, are displayed.

Table 2
Quantitative results on RgB2CAVE data set, including SSR and FusSR, which could be differentiated by the use of LR Green channel. The best is highlighted in bold and the second best is underlined.

	SSR				FusSR				
	CC	mPSNR	mSSIM	SAM	CC	mPSNR	mSSIM	SAM	ERGAS
DenseU	0.9762	29.2668	0.9336	11.4231	0.9911	32.0511	0.9549	8.1459	8.9606
CanNet	0.9742	29.0431	0.9469	11.2296	0.9917	32.9176	0.9695	8.4988	8.6572
HSCNN+	0.9737	30.1048	0.9514	11.3278	0.9920	33.0453	0.9698	8.9057	8.7036
HSRnet	**0.9755**	30.4968	**0.9551**	11.0565	0.9929	33.6890	0.9720	8.0627	8.0939
PoNet	0.9748	**30.5074**	0.9528	**10.8340**	**0.9933**	**34.2215**	**0.9744**	**7.9154**	**7.6652**

Table 3
Quantitative results on Sen2OHS data set, including SSR and FusSR, which could be differentiated by the use of 20 m-resolution channels. The best is highlighted in bold and the second best is underlined.

	SSR				FusSR				
	CC	mPSNR	mSSIM	SAM	CC	mPSNR	mSSIM	SAM	ERGAS
DenseU	0.9498	26.7262	0.8769	8.3135	0.9543	27.2900	0.8723	6.7571	10.1827
CanNet	0.9621	28.1981	0.8901	7.4233	0.9690	28.9335	0.8964	5.4653	6.7348
HSCNN+	0.9593	28.8117	0.9164	6.9076	0.9672	29.6968	0.9253	5.4139	6.5567
HSRnet	0.9725	28.9801	0.9344	6.8410	0.9749	29.9565	0.9329	5.2295	7.1032
PoNet	**0.9748**	**29.3074**	**0.9428**	**6.5788**	**0.9816**	**30.5290**	**0.9521**	**4.9404**	**5.4513**

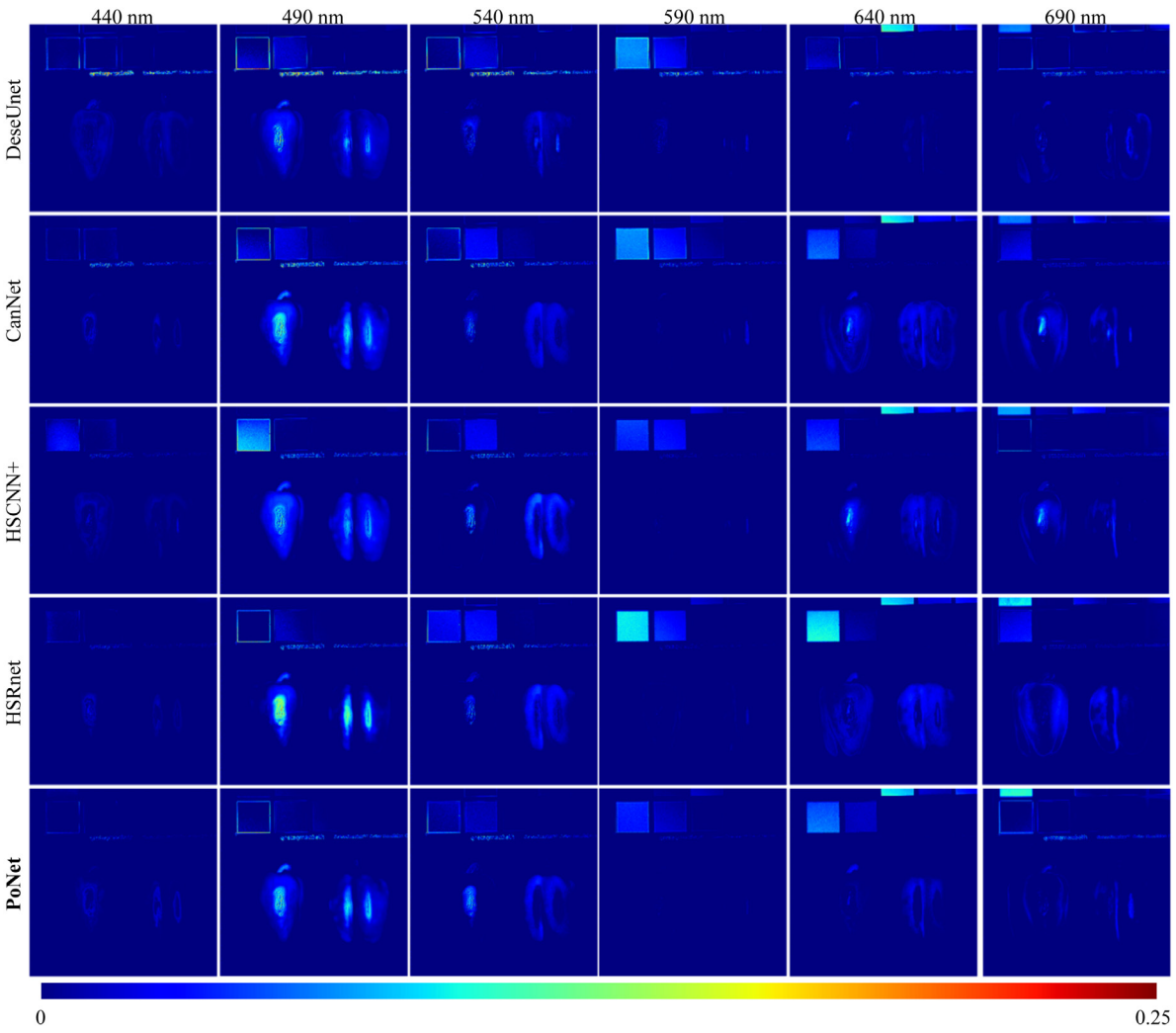


Fig. 13. Visualization of the error maps between ground-truth and model results on RgB2CAVE data set with all multispectral channels, including the LR Green channel. From left to right, reconstructed bands of wavelength 440, 490, 540, 590, 640, and 690 nm, are displayed.

PoNet visually show less local variation than other compared methods and DenseUnet restores images with the highest error. Furthermore, because of insufficient spectral information, all the models show poor spectral enhancement at the wavelength range of 520, 670, 746, and 896 nm. Moreover, The compared methods show poor results on farmlands and bare lands, while the proposed PoNet can better handle this problem.

Fig. 15 displays the visual comparison between models for addressing FusSR using Sen2OHS data set. Compared with the results in Fig. 14, all bands get lower errors, especially for the wavelength of 520, 746, and 896 nm. Analyzing the error maps in 670 and 896 nm, there are a wide variety of crops with different spectral information in farmlands, which brings a big challenge to spectral restoration. However, PoNet can still achieve perfect spectral recovery for all crops in images, which shows the great superiority of the proposed physical interpretability-based optimization stages and cross-dimensional channel attention.

4.4. Results on Sen2CHRIS data set

Using Sen2CHRIS data set, including Xiong'an, Washington DC Mall, and Chikusei, We evaluate the model validity of recovering new hyperspectral information obtained by different spectral response functions.

Quality metrics reported in Tables 4 and 5 reveals that PoNet have a stable advantage to achieve FusSR with different spectral response functions, while it shows performance degradation in normal SSR.

Fig. 16 displays the error maps of the proposed method as well as four compared methods on Sen2CHRIS. To compare the SSR results between data subsets, different bands in different images are selected. DenseUnet still gets the worst spatial details among three data subsets showing more edge details in error maps. CanNet shows surprising good performance in visual performance. Besides, there are two shortages for all models. The first one is the performance in Washington DC Mall data subset is not good enough. Secondly, the spectral restoration to the bands at the spectrum edge is insufficient.

Visual comparisons of FusSR are shown in Fig. 17. Error maps of all models decrease except DenseUnet showing more highlighted values in bands of 561 and 833 nm, which is caused by the introduction of low-resolution spectral information. Similar to results in SSR, PoNet and CanNet get the best performance showing low error maps especially for the buildings in 703 nm band. Furthermore, PoNet shows superiority in bands of 561, 883, and 905 nm.

4.5. Results on RgB2CASI data set

To verify the model performance at very high resolution, images in RgB2CASI data set are employed and the quantitative results in SSR and FusSR are reported in Table 6. In SSR, although the spectral

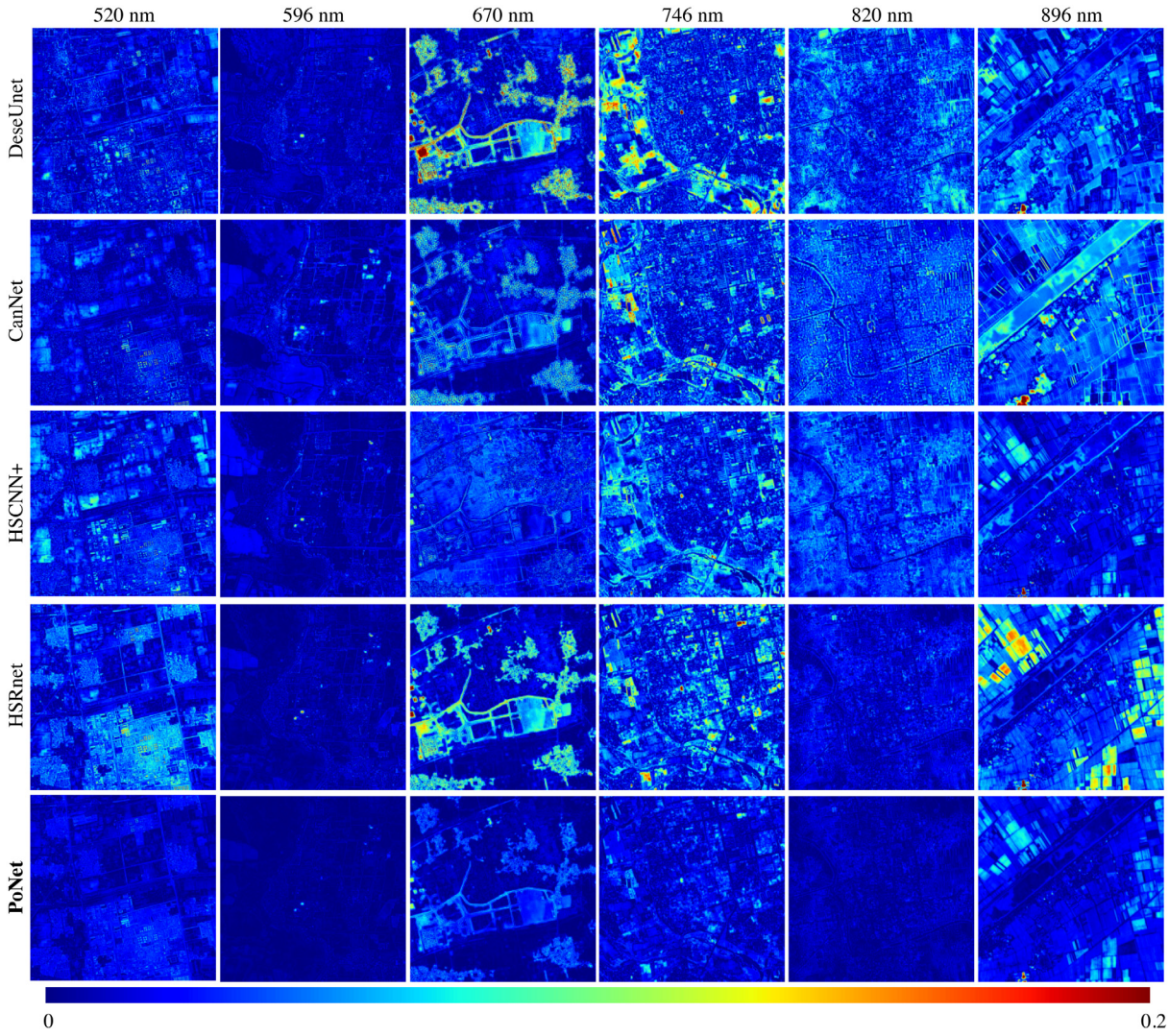


Fig. 14. Visualization of the error maps between ground-truth and model results only using four high-resolution multispectral channels on Sen2OHS data set. Six images selected from test data are displayed. From left to right, the wavelengths of reconstructed bands are 520, 596, 670, 746, 820, and 896 nm.

Table 4
Quantitative results on Sen2CHRIS data set, including Xiong'an, Washington DC Mall, and Chikusei data subsets, where only channels with 10 m resolution in multispectral images are used as input, i.e., band 2, 3, 4, and 8. The best is highlighted in bold and the second best is underlined.

	Xiong'an				Washington DC Mall				Chikusei			
	CC	mPSNR	mSSIM	SAM	CC	mPSNR	mSSIM	SAM	CC	mPSNR	mSSIM	SAM
DenseU	0.9776	31.0294	0.9544	4.3396	0.9882	41.4252	0.9805	2.2813	0.9877	39.0021	0.9773	4.6744
CanNet	0.9945	37.5414	0.9830	2.9026	0.9984	51.4529	0.9979	1.0816	0.9937	43.4565	0.9872	4.3289
HSCNN+	0.9924	35.7680	0.9846	3.4531	0.9984	51.6680	0.9979	1.0519	0.9941	42.6957	0.9880	4.3493
HSRnet	**0.9963**	37.4468	**0.9897**	3.4239	0.9986	52.1912	0.9982	0.9935	**0.9947**	**44.0038**	**0.9889**	**4.2829**
PoNet	0.9952	**39.2589**	0.9879	**2.1122**	**0.9987**	**52.9510**	**0.9984**	**0.9724**	0.9939	43.8969	0.9878	4.3337

Table 5
Quantitative results on Sen2CHRIS data set, including Xiong'an, Washington DC Mall, and Chikusei data subsets, where only channels with 10 m resolution in multispectral images are used as input, namely, band 2, 3, 4, and 8. The best is highlighted in bold and the second best is underlined.

	Xiong'an					Washington DC Mall					Chikusei				
	CC	mPSNR	mSSIM	SAM	ERGAS	CC	mPSNR	mSSIM	SAM	ERGAS	CC	mPSNR	mSSIM	SAM	ERGAS
DenseU	0.9801	31.2499	0.9557	3.7330	3.5526	0.9883	41.4672	0.9809	2.2485	3.2449	0.9883	39.2332	0.9773	4.6973	8.1293
CanNet	0.9958	39.7155	0.9836	1.6726	1.5151	0.9984	51.3461	0.9979	1.0409	1.1562	0.9961	44.4941	0.9862	4.1710	6.2716
HSCNN+	0.9953	38.2452	0.9889	1.9854	1.8595	0.9981	51.2922	0.9978	1.0624	1.2053	0.9953	42.6871	0.9885	4.2578	6.4048
HSRnet	0.9972	40.2616	0.9924	1.7292	1.5238	0.9987	52.7957	0.9983	0.9238	1.0271	0.9968	44.4520	**0.9902**	4.0105	6.1612
PoNet	**0.9974**	**41.9316**	**0.9928**	**1.4185**	**1.2398**	**0.9989**	**53.7637**	**0.9986**	**0.8865**	**0.9739**	**0.9969**	**45.1265**	0.9901	**3.9189**	**6.0190**

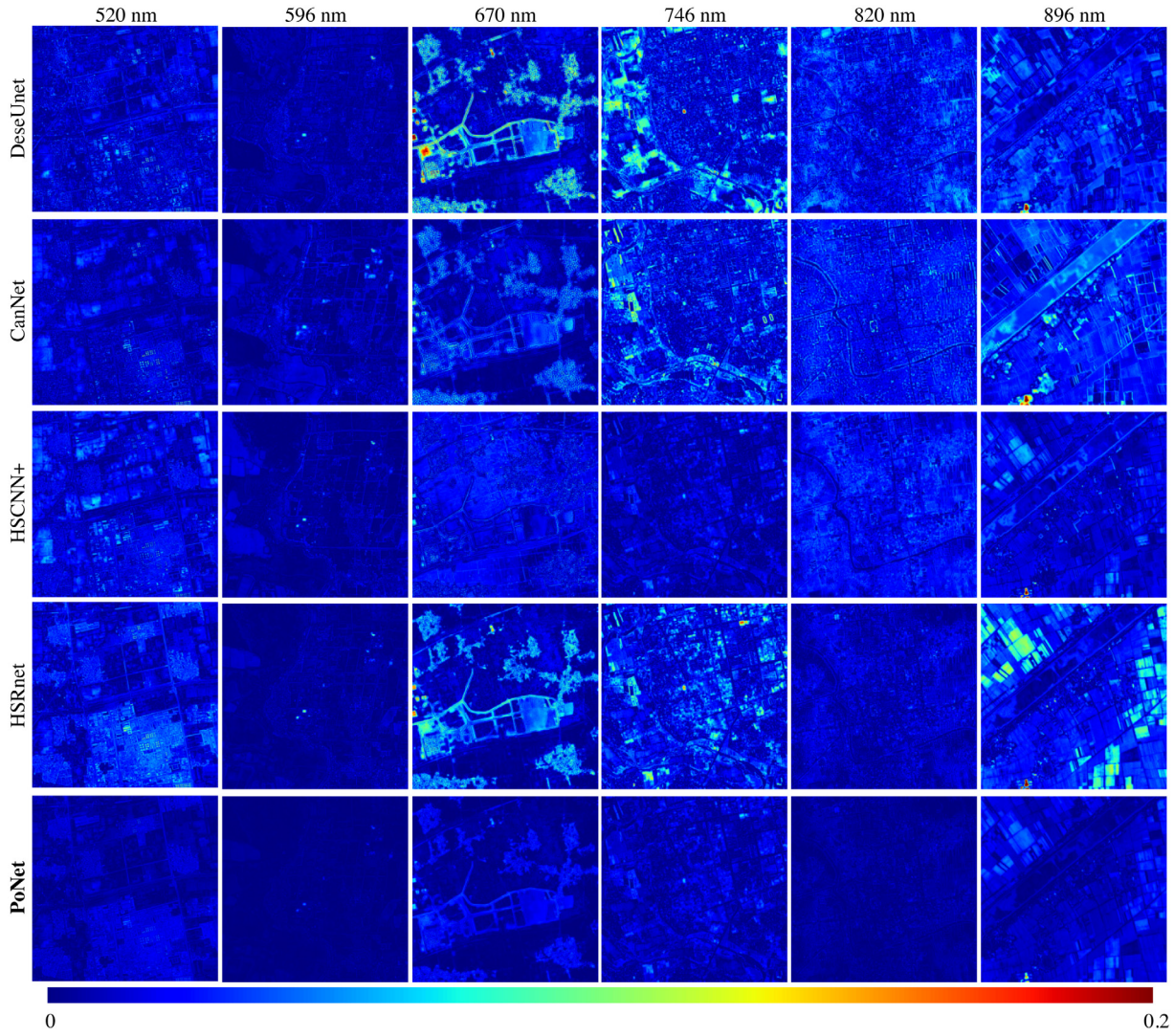


Fig. 15. Visualization of the error maps between ground-truth and model results on Sen2OHS data set with four high-resolution and three low-resolution multispectral channels. Similarly, six images selected from test data are displayed. From left to right, the wavelengths of reconstructed bands are 520, 596, 670, 746, 820, and 896 nm.

Table 6
Quantitative results on RgB2CASI data set, including SSR and FusSR, which could be differentiated by the use of LR Green channel. The best is highlighted in bold and the second best is underlined.

	SSR				FusSR				
	CC	mPSNR	mSSIM	SAM	CC	mPSNR	mSSIM	SAM	ERGAS
DenseU	**0.8010**	29.7784	<u>0.8424</u>	14.7484	0.8433	30.4219	0.8631	13.7883	11.8762
CanNet	0.6266	27.0725	0.7956	15.8461	0.7775	28.2414	0.8418	14.0059	14.0456
HSCNN+	0.7874	29.5541	0.8395	14.7601	0.8415	30.3498	0.8652	<u>13.7815</u>	<u>11.8079</u>
HSRnet	<u>0.7996</u>	**29.8161**	**0.8446**	<u>14.6122</u>	<u>0.8454</u>	**30.5858**	<u>0.8657</u>	13.8958	11.8192
PoNet	0.7951	<u>29.5439</u>	0.8420	**13.4246**	**0.8479**	<u>30.5309</u>	**0.8683**	**12.5997**	**9.2775**

maintaining of PoNet achieves the best performance while the improvement of spatial fidelity is not much. Moreover, DenseUnet gets a better rank in this data set benefits by the deep structure compared with CanNet. As for FusSR, PoNet is still superior to other models with the optimal values for CC, mSSIM, SAM, and ERGAS, which evidences the competitive performance of the proposed method to deal with FusSR. Moreover, with the help of more spectral information, HSCNN+ gets better results than DenseUnet. CanNet shows poor performance due to its shallow architecture that cannot be qualified to extract more high-resolution spectral information.

Figs. 18 and 19 display the error maps of results in SSR and FusSR, where different bands of seven images are selected in RgB2CASI data set. Unlike the previous experimental results, the error maps of 718 nm are the lowest, and bands from 818 to 1018 nm get high residuals. The reason may be the spectral response functions of Red and Blue channels in RgB2CASI data set cover the wider wavelength.

However, adding the LR Green channel into the input still can improve the performance of spectral enhancement, which can be confirmed by lower residuals in error maps. Moreover, in the band of 818 nm, the errors on the playground increase which is the opposite of overall performance. For this data set, the auxiliary low-resolution channel is Green channel, which cannot bring more spectral information to help models distinguish between real and fake grass. Even so, the proposed PoNet performs the better spectral recovery among all models.

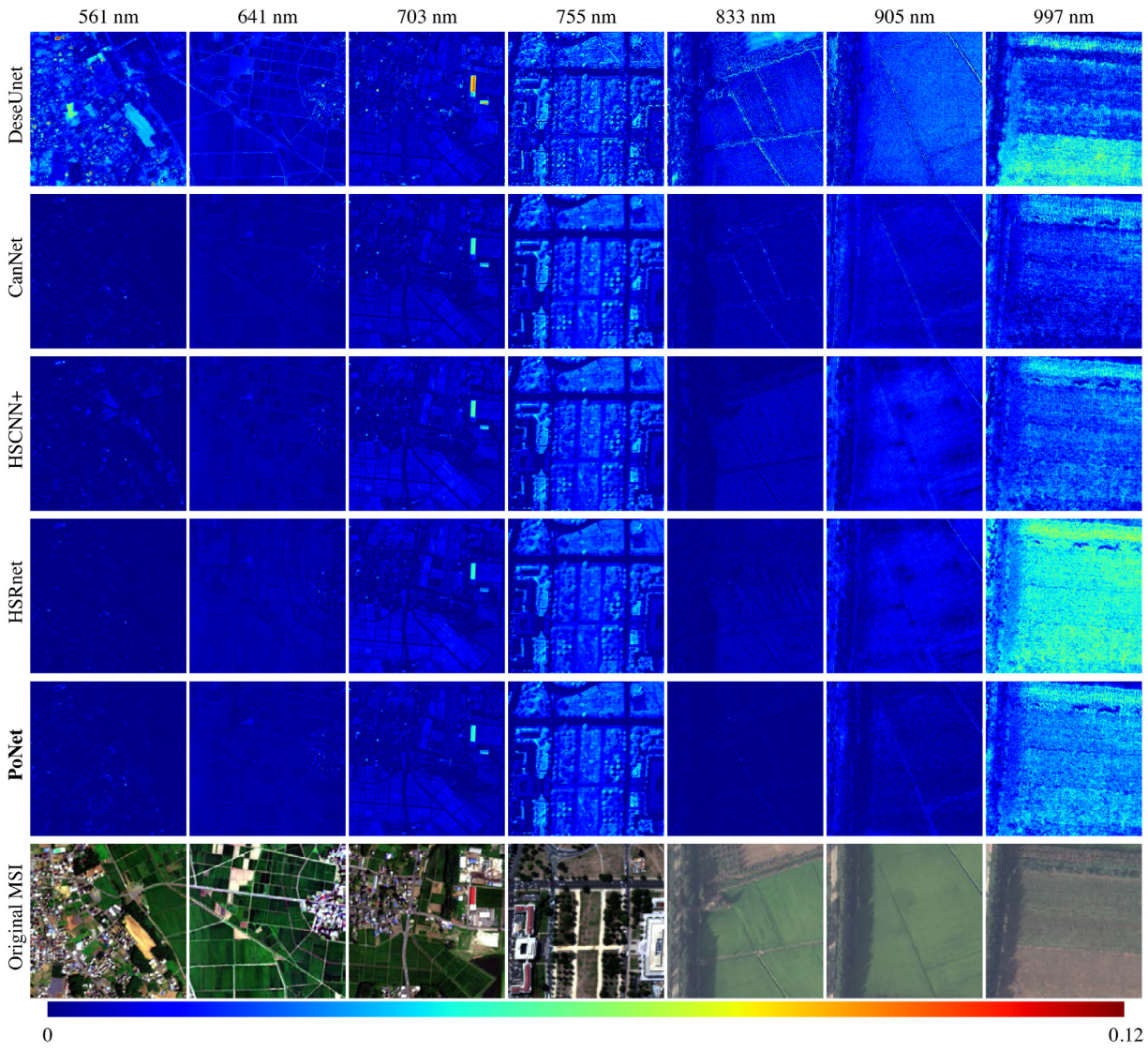


Fig. 16. Visualization of the error maps between ground-truth and model results only using four high-resolution multispectral channels on Sen2CHRIS data set. There are also three data subsets in Sen2OHS, thus, the other seven images different from Sen2OHS data set selected from test data in Chikusei (Column 1–3), Washington DC Mall (Column 4), and Xiong'an (Column 5–7), are displayed. From left to right, the wavelengths of reconstructed bands are 561, 641, 703, 755, 833, 905, and 997 nm.

Table 7
Quantitative results on GF2Hyper data set for addressing the other situation of spectral super-resolution, namely, PansSR. After the results of seven single images, the averages are also calculated at the last row.

Images	CC	mPSNR	mSSIM	SAM	ERGAS
1	0.9284	44.2446	0.9828	4.2478	3.3479
2	0.8972	48.5929	0.9914	2.2095	2.0790
3	0.9325	45.8955	0.9849	3.9832	3.0857
4	0.9071	44.3540	0.9522	6.1078	4.6244
5	0.8770	49.3000	0.9937	3.0700	4.9934
6	0.8315	45.0522	0.9870	2.0565	2.8733
7	0.9691	43.9667	0.9767	2.8755	1.8548
AVG	0.9284	44.2446	0.9828	4.2478	3.3479

4.6. PansSR

It is a big challenge to use the same model to address PansSR. Utilizing the proposed GF2Hyper data set, we can test the performance of PoNet in PansSR while the other models cannot directly deal with this problem. Hence, Table 7 only lists the PoNet's quantitative results of seven selected images. Note that although the resolution ratio between PAN and multispectral images is up to 4, the averages of CC, mPSNR, and mSSIM are comparatively ideal, which can verify the validity of PoNet to addressing PansSR.

To presents the performance of PoNet more clearly, Fig. 20 displays the PansSR results of four randomly selected images. The first and second rows show the input data, including up-sampled multispectral images using Bilinear interpolation and PAN images. The false-color images using bands of 884.7, 640.5, and 548.9 nm are shown in the third row. In the last row, the reflectance of PAN, multispectral images, reference hyperspectral images, and PoNet results at randomly selected locations shown in (i)–(l) with yellow marks are displayed to compare the spectral maintaining.

Comparing 1st with 3rd row, rich spatial details are fused into the reconstruction results although there is a wide resolution gap between PAN and multispectral images. Seeing from (m)–(p), more spectral information can be explored by PoNet to recover ideal hyperspectral images. In these figures, several interesting results should be noticed. One is that (o) shows an abnormal sharp decline in band 48 and 49, however, this area is covered by vegetation, which should be similar to (m). Checked the specific channels, we found it is caused by the insufficient denoising to strong noise. In this situation, the proposed

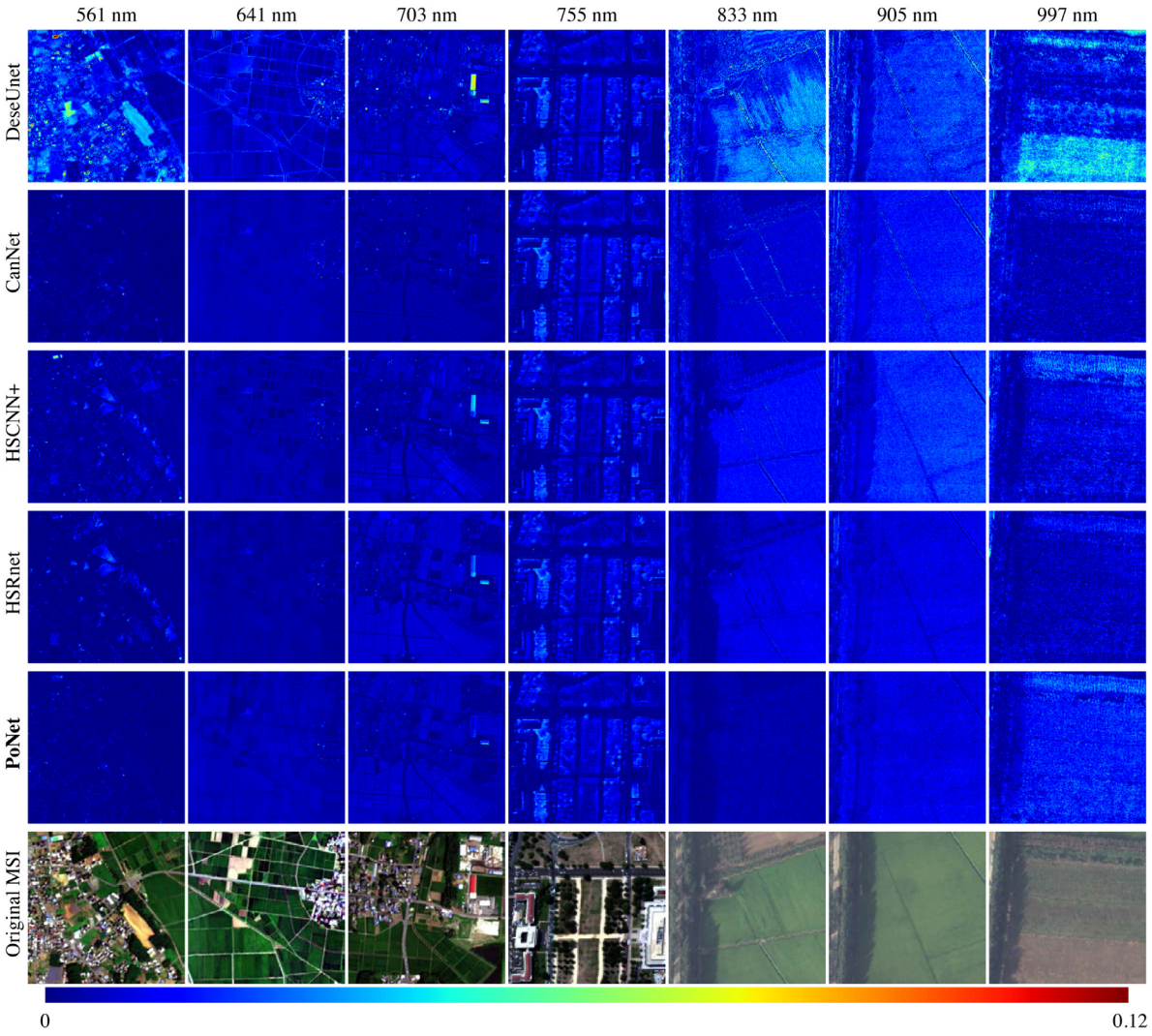


Fig. 17. Visualization of the error maps between ground-truth and model results on Sen2CHRIS data set with four high-resolution and three low-resolution multispectral channels. Similar to Fig. 16, seven images selected from test data in Chikusei (Column 1–3), Washington DC Mall (Column 4), and Xiong'an (Column 5–7), are displayed. From left to right, the wavelengths of reconstructed bands are 561, 641, 703, 755, 833, 905, and 997 nm.

PoNet can still recover better hyperspectral information with smooth spectral curve, which shows the stability of PoNet.

4.7. Discussion

4.7.1. Influence of sampling operator

In the proposed PoNet, up-sampling and downsampling operators play an important role in optimization stages. To explore the best up-down combination, several downsampling operators are discussed in this sub-section, including Avgpooling, Maxpooling, Convolution, Nearest, and Bilinear interpolation. Moreover, we also select four up-sampling operators into experiments, namely, Bilinear interpolation, Deconvolution, Pixelshuffle [63], and *Content-Aware ReAssembly of FEatures* (CARAFE) [64]. Among them, only CNN-based operators will increase the model parameter, namely, Convolution, Deconvolution, Pixelshuffle, and CARAFE.

Tables 8 and 9 list the quantitative results of different sampling combinations on Washington DC Mall in Sen2OHS data set. We display results of three indexes, including mSSIM for spatial fidelity, SAM for spectral maintaining, and ERGAS for global evaluation. It can be seen that the best sampling combination is Bilinear interpolation and Maxpooling, which show the best quantitative results among other complex sampling operators.

Specifically, analyzing the influence of up-sampling operators by row, Bilinear interpolation shows the lowest ERGAS with any downsampling operator. As for downsampling operators, although Maxpooling with Bilinear interpolation gets the best results, Maxpooling shows poor stability as changing into other up-sampling operators. On the contrary, Convolution could achieve a more stable operation, which is more likely due to the strong learning ability and the calculation process similar to the degradation in practice. All of the above is based on the comprehensive index, namely, ERGAS.

To discuss spatial fidelity only, from mSSIM, Deconvolution shows the second-best results after Bilinear interpolation. Moreover, Avgpooling gets the worst mSSIM among all downsampling operators, which is due to the sparse features in hyperspectral images [65]. For spectral maintaining, with pixel-by-pixel weights, CARAFE keeps more spectral information than Deconvolution and Pixelshuffle. As for downsampling operators, Convolution still shows better generalization. Visualization of three metrics is shown in Fig. 21, and the same conclusion could be drawn.

Finally, taking both performance and computation complexity into account, we selected Bilinear interpolation as up-sampling operator and Maxpooling as downsampling operator, which can achieve the best performance and require no parameters to be learned.

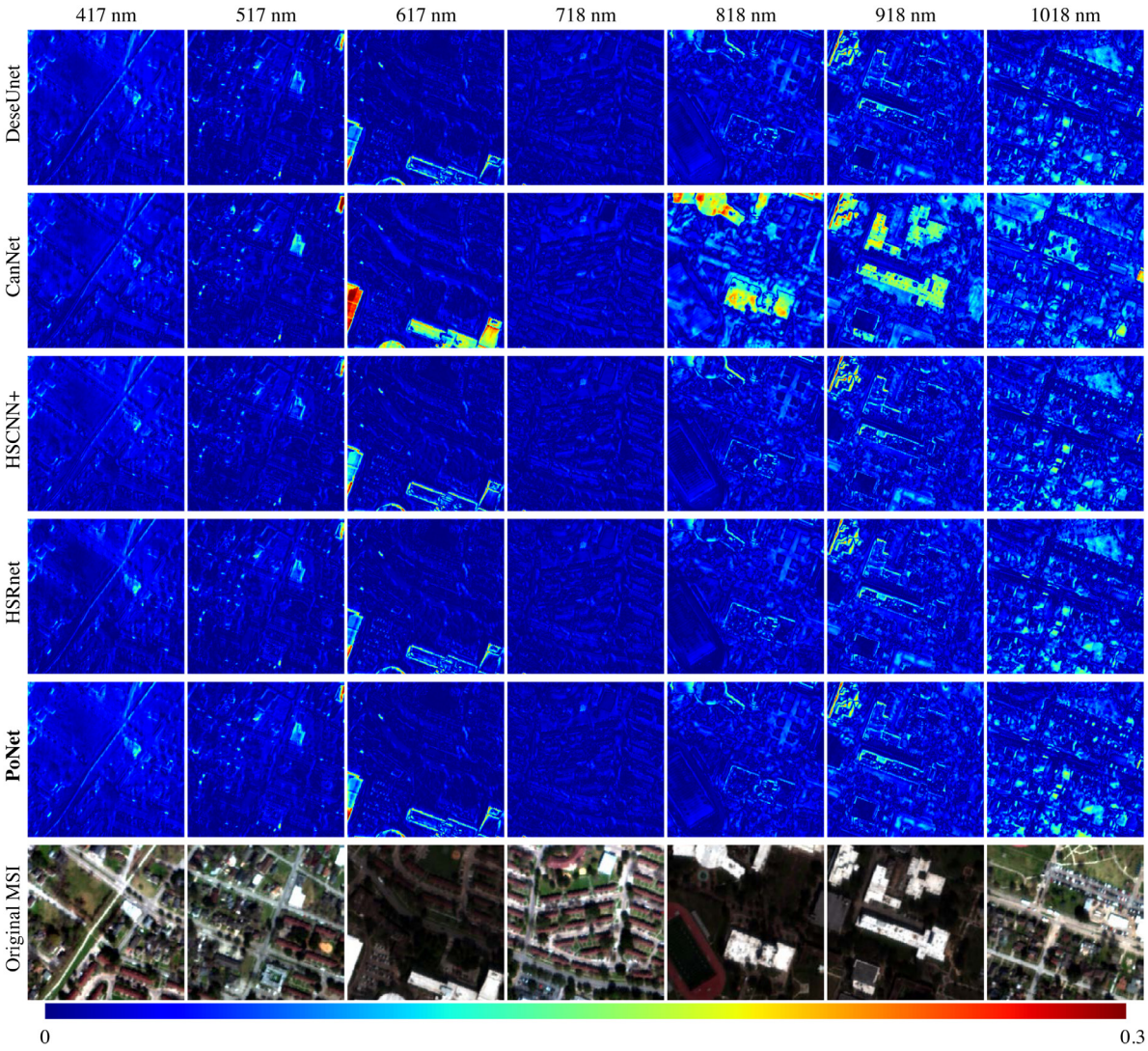


Fig. 18. Visualization of the error maps between ground-truth and model results on RgB2CASI data set with only Red and Blue channels, namely plain SSR. From left to right, reconstructed bands of wavelength 417, 517, 617, 718, 818, 918, and 1018 nm, are displayed.

Table 8
SAM and mSSIM of different combinations between downsampling and up-sampling operators on Washington DC Mall in Sen2OHS data set. In each column, the best is in bold and the second best is marked with an underline.

	SAM					mSSIM				
	Avg	Bili	Conv	Max	Near	Avg	Bili	Conv	Max	Near
Bili	**0.8883**	**0.8459**	**0.8241**	**0.8169**	**0.8315**	**0.99808**	**0.99824**	**0.99832**	**0.99837**	0.99827
CARAFE	1.0264	0.9274	0.9177	0.9056	0.9265	0.99742	0.99793	0.99807	0.99808	0.99794
Deconv	1.5033	0.9364	0.8624	0.9260	0.8674	0.99667	0.99808	0.99827	0.99813	0.99828
PixelS	1.0258	1.1251	0.8649	0.9196	0.8396	0.99726	0.99700	0.99821	0.99813	**0.99832**

Table 9
ERGAS of different combinations between downsampling and up-sampling operators on Washington DC Mall in Sen2OHS data set. In each column, the best is in bold and the second best is marked with an underline.

	Avg	Bili	Conv	Max	Near
Bili	**0.9828**	**0.9456**	**0.9307**	**0.9201**	**0.9328**
CARAFE	1.1534	1.0392	0.9971	1.0022	1.0342
Deconv	1.6843	1.0227	0.9897	1.0281	0.9782
PixelS	1.2234	1.2494	0.9734	1.0144	0.9518

4.7.2. Influence of input channel number

In spectral super-resolution, input channel number is also of great importance. In this paper, we discussed the relationship between spectral recovery results and the input multispectral bands number on Chikusei data subset of Sen2CHRIS data set. As can be seen in Fig. 22, channels for test are of two spatial resolution. Thus, we first test channels, from short center wavelength to long center wavelength, with 10 m resolution and then channels with 20 m resolution and results are shown in Fig. 23.

As input bands number increases, no matter PSNR or SAM gets better, extremely after the fourth band is used. The fourth band has a center wavelength of 842 nm and covers wavelength from 776 to

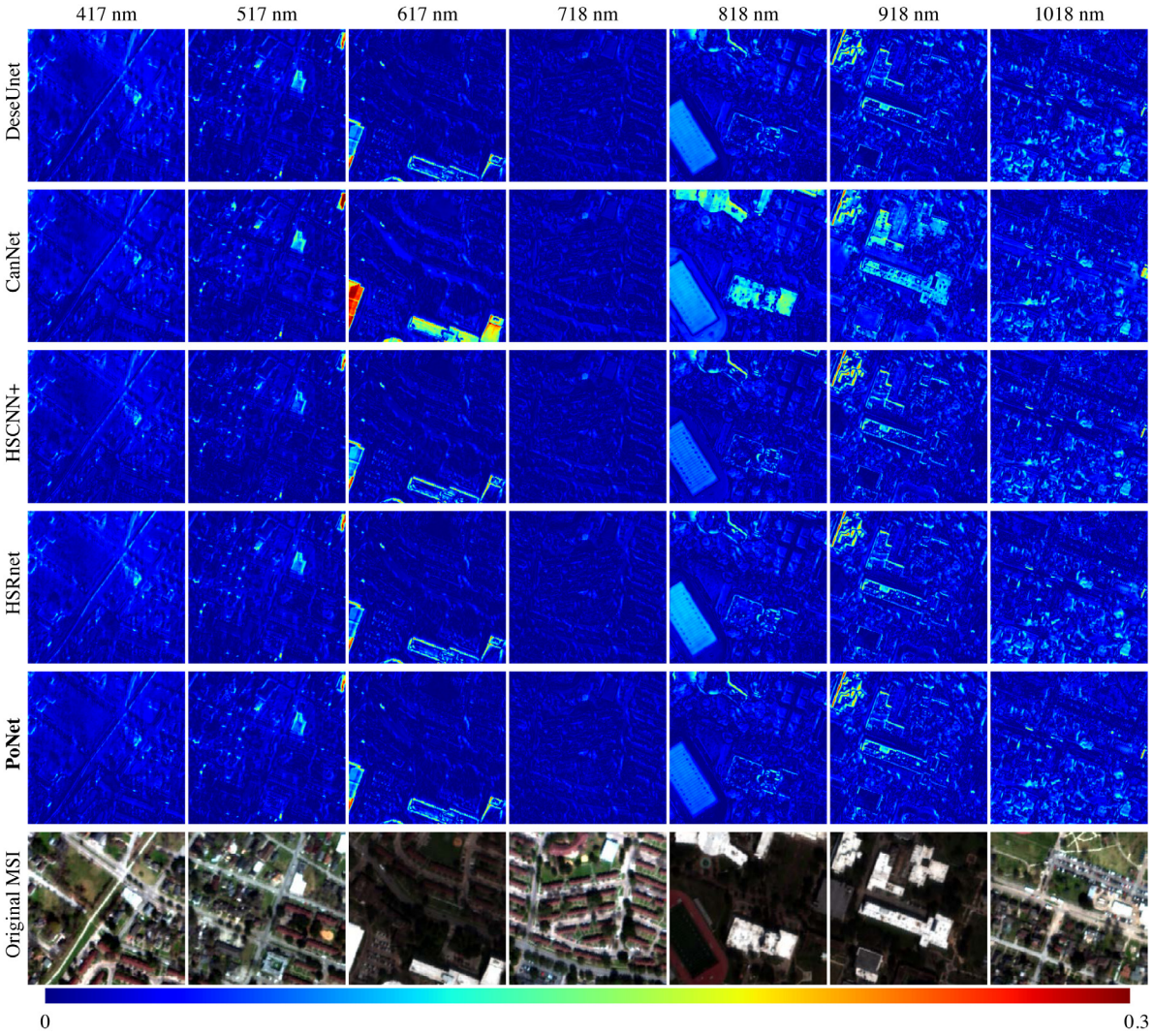


Fig. 19. Visualization of the error maps between ground-truth and model results on RgB2CASI data set with all multispectral channels, including the LR Green channel. From left to right, reconstructed bands of wavelength 440, 490, 540, 590, 640, and 690 nm, are displayed.

912 nm, which is the widest channel of Sentinel-2 sensor. The wider wavelength channels cover, the more spectral information channels contain. This is the reason results are improved dramatically after the fourth band is used. Furthermore, when we add more 20m-resolution channels, SAM and PSNR are also getting better.

4.7.3. Ablation study

To improve our PoNet, several modules and strategies are proposed in this paper and we discussed their effectiveness in this subsection using RgB2CAVE data set. Four modules and strategies are presented in this work, including physical optimization-based stages, a new combination of down- and up-sampling operators, the Cross-Dimensional Channel Attention module, and the Cross-Depth Feature Fusion strategy.

As reported in Table 10, HSCNN+ is selected as baseline for its good performance among traditional deep learning-based algorithms. HSRnet is our previous work which is also based on an optimization algorithm. PoNet w/o CDCA is the model with optimization stages and sampling operators. As for PoNet w/o CDFF, CDCA is employed in this model but CDFF is not used. The final PoNet is the complete model proposed in this paper.

Compared with HSCNN+, other models with optimization stages are with great improvement whether in spatial or spectral domain, which states that physical interpretability plays a very important role in model construction. Furthermore, traditional deep learning-based models even require more parameters.

Seeing the difference between HSRnet and PoNet w/o CDCA, we employ Bilinear interpolation and Maxpooling as sampling operators, but HSRnet does not consider the up-sampling and down-sampling in the model. As listed in the table, indexes to evaluate spatial fidelity, such as CC, mPSNR, and mSSIM, get a great improvement. It indicates that adding sampling into the model can increase the learning ability to spatial degradation. Nevertheless, SAM gets an increase, which means the model has a slight loss on spectral maintaining. Certainly, it still can reach better quantitative evaluation than traditional deep learning-based models.

With the help of CDCA, PoNet w/o CDFF can hold the balance between spectral information and spatial details, showing lower SAM and higher mPSNR in the table. It states that better parametric self-learning pixel by pixel can truly improve the spatial fidelity and spectral enhancement of models.

Seen from the last two rows in Table 10, the complete PoNet can get further improvement in all indexes, no matter SAM or CC, mPSNR,

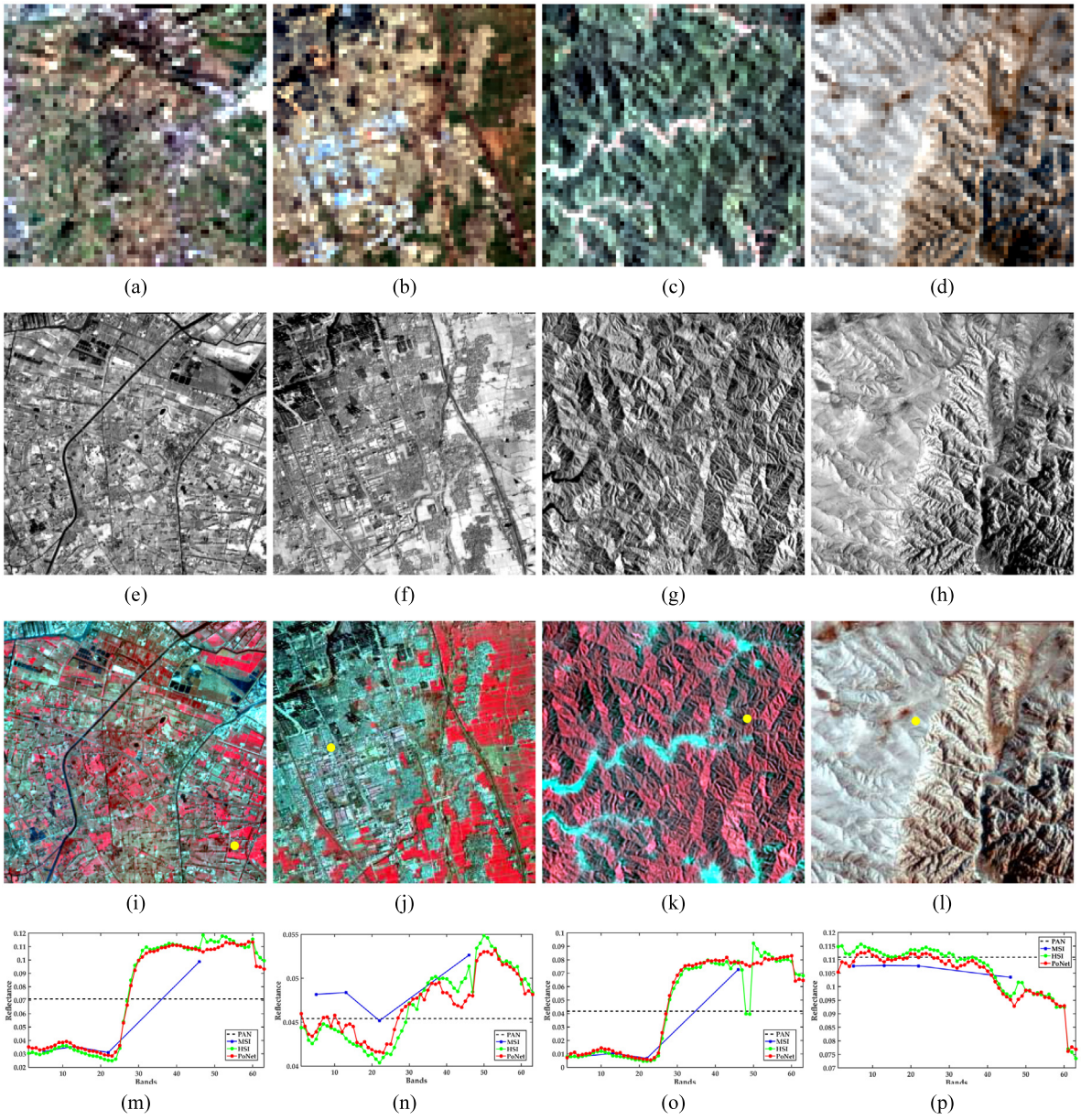


Fig. 20. PansSR results of four randomly selected images. (a)–(d) Up-sampled MS images. (e)–(h) PAN images. (i)–(l) False-color images of spectral reconstruction results. (m)–(p) Reflectance of PAN, MSI, reference HSI, and PoNet results at the randomly selected location.

Table 10
Quantitative results in ablation study of the proposed modules and strategies. The best results are highlighted in bold and the second is underlined.

Models	Optimization stage	Down and up	CDCA	CDFF	CC	mPSNR	mSSIM	SAM	ERGAS
HSCNN+	✗	✗	✗	✗	0.9920	33.0453	0.9698	8.9057	8.7036
HSRnet	✓	✗	✗	✗	0.9929	33.6890	0.9720	8.0627	8.0939
PoNet w/o CDCA	✓	✓	✗	✗	0.9930	33.8864	0.9734	8.1364	8.0202
PoNet w/o CDFF	✓	✓	✓	✗	<u>0.9931</u>	<u>34.0456</u>	<u>0.9743</u>	<u>7.9463</u>	<u>7.7565</u>
PoNet	✓	✓	✓	✓	**0.9933**	**34.2215**	**0.9744**	**7.9154**	**7.6652**

mSSIM, and ERGAS. Although the improvement seems tiny, CDFF, as a strategy which is easy to achieve, also makes great sense to improve convergence speed in training.

4.7.4. Computational speed analysis

For deep learning-based algorithms, not only the model effect, but model complexity and computational speed are also important. Table 11 lists the parameter numbers, floating-point operations (FLOPs), and training and test time of deep learning methods. Training and test time are all counted on the CAVE data set in FusSR. DenseUnet requires the most parameters but shorter test time benefited from numerous dense blocks to speed calculation. CanNet requires the least parameters and fastest running time because of its shallow structure, while PoNet can achieve the earliest convergence due to the physical optimization-based structure. To sum up, PoNet with acceptable parameter numbers and computation complexity can get the best results. In addition, PoNet can also realize early convergence with not long running time, which are all conducive to daily application.

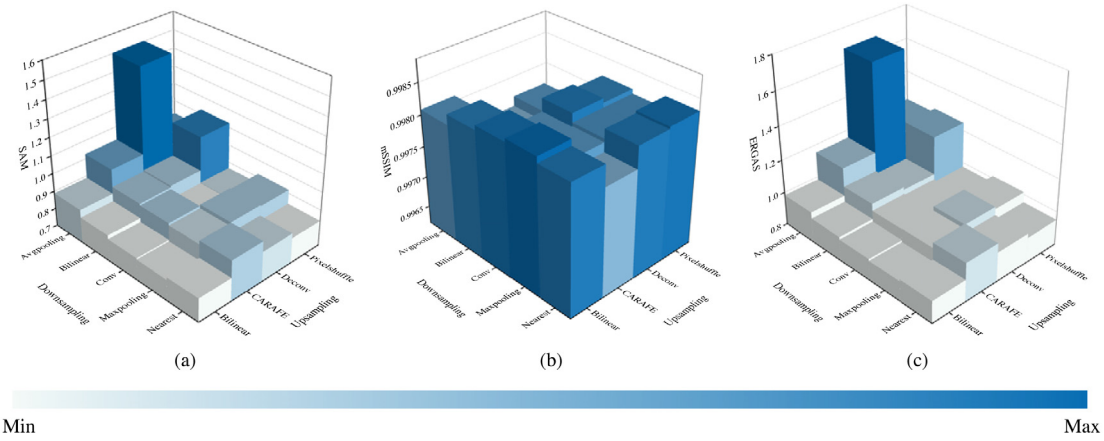


Fig. 21. Visualization of quantitative indicators in Tables 8 and 9, which are calculated on the results using different sampling operators. (a) SAM. (b) mSSIM. (c) ERGAS.

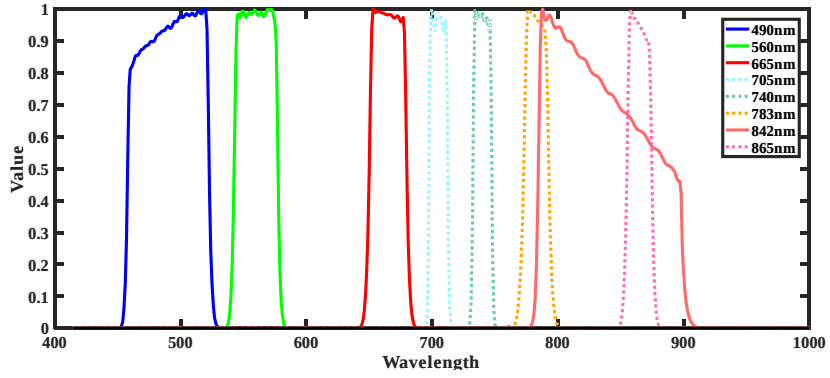


Fig. 22. Spectral response functions of Sentinel-2 satellite. Solid lines indicate channels with 10 m resolution, while dotted lines represent channels with 20 m resolution. Moreover, legend presents channel with center wavelength.

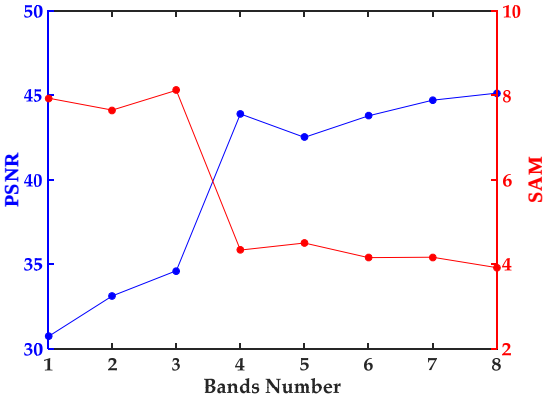


Fig. 23. PSNR and SAM of results with different input bands number. Black line represents PSNR, while red line represents SAM.

Table 11
Comparisons on model complexity and computational speed between five deep learning-based models.

	DenseUnet	CanNet	HSCNN+	HSRnet	PoNet
Params	1360.1K	163.0K	915.1K	769.7K	363.2K
FLOPs	3.02×10^{10}	3.97×10^{10}	2.23×10^{11}	1.79×10^{11}	8.51×10^{10}
Training	68 655 s	49 285 s	57 805 s	30 831 s	15 568 s
Test	1.2598 s	1.2387 s	1.7996 s	1.5364 s	1.4381 s

5. Conclusions

This paper presents a universal physical optimization-based CNN named PoNet to address spectral super-resolution for arbitrary multispectral images including data with multiple spatial resolutions, namely, SSR, FusSR, and PansSR. Unfolding optimization algorithm considering physical degradation to deep learning gives CNN the important physical interpretability, which provides great help to recover hyperspectral information. Besides, to learn parameters channel-to-channel adaptively, as well as boost computation, cross-dimensional channel attention is proposed. We also use both deep and shallow features to perform better spectral enhancement and spatial fidelity. To propose a baseline for evaluating model performance in three multispectral acquisitions, five data sets involving natural images and cross-scale remote sensing images, namely, RgB2CAVE, Sen2OHS, Sen2CHRIS, RgB2CASI, and GF2Hyper, are built in this paper. Quantitative and visual comparisons on these data sets illustrates that physical degradation model can help deep learning-based model recover spectral information better. Furthermore, we also discussed the influence of sampling operators and ablation study to verify the effectiveness of proposed strategies.

CRediT authorship contribution statement

Jiang He: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Qiangqiang Yuan:** Conceptualization, Resources Writing – original draft, Supervision, Project administration, Funding acquisition. **Jie Li:** Conceptualization, Methodology, Resources, Data curation, Writing – original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Liangpei Zhang:** Conceptualization, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 41922008, Grant 62071341, and Grant 61971319; in part by the Hubei Science Foundation for Distinguished Young Scholars under Grant 2020CFA051; and in part by the Fundamental Research Funds for the Central Universities under Grant 531118010209.

References

[1] Y. Tarabalka, J. Chanussot, J. Benediktsson, Segmentation and classification of hyperspectral images using watershed transformation, Pattern Recognit. 43 (7) (2010) 2367–2379.
[2] J. Li, J.M. Bioucas-Dias, A. Plaza, Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning, IEEE Trans. Geosci. Remote Sens. 48 (11) (2010) 4085–4098.
[3] J.C. Harsanyi, C.. Chang, Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach, IEEE Trans. Geosci. Remote Sens. 32 (4) (1994) 779–785.
[4] G. Camps-Valls, L. Bruzzone, Kernel-based methods for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 43 (6) (2005) 1351–1362.
[5] G. Camps-Valls, D. Tuia, L. Bruzzone, J.A. Benediktsson, Advances in hyperspectral image classification: Earth monitoring with statistical learning methods, IEEE Signal Process. Mag. 31 (1) (2014) 45–54.
[6] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, IEEE Trans. Geosci. Remote Sens. 42 (8) (2004) 1778–1790.
[7] M. Fauvel, J.A. Benediktsson, J. Chanussot, J.R. Sveinsson, Spectral and spatial classification of hyperspectral data using svms and morphological profiles, IEEE Trans. Geosci. Remote Sens. 46 (11) (2008) 3804–3814.
[8] H. Ren, C.-I. Chang, Automatic spectral target recognition in hyperspectral imagery, IEEE Trans. Aerosp. Electron. Syst. 39 (4) (2003) 1232–1249.
[9] D. Manolakis, G. Shaw, Detection algorithms for hyperspectral imaging applications, IEEE Signal Process. Mag. 19 (1) (2002) 29–43.
[10] H.V. Nguyen, A. Banerjee, R. Chellappa, Tracking via object reflectance using a hyperspectral video camera, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2010, pp. 44–51.
[11] T. Wang, Z. Zhu, E. Blasch, Bio-inspired adaptive hyperspectral imaging for real-time target tracking, IEEE Sens. J. 10 (3) (2010) 647–654.
[12] A. Gowen, C. O'Donnell, P. Cullen, G. Downey, J. Frias, Hyperspectral imaging-an emerging process analytical tool for food quality and safety control, Trends Food Sci. Technol. 18 (12) (2007) 590–598.
[13] G.K. Naganathan, L.M. Grimes, J. Subbiah, C.R. Calkins, A. Samal, G.E. Meyer, Visible/near-infrared hyperspectral imaging for beef tenderness prediction, Comput. Electron. Agric. 64 (2) (2008) 225–233.
[14] M.J. Barnsley, J.J. Settle, M.A. Cutter, D.R. Lobb, F. Teston, The PROBA/CHRIS mission: a low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere, IEEE Trans. Geosci. Remote Sens. 42 (7) (2004) 1512–1520.
[15] W. Verhoef, H. Bach, Coupled soil-leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data, Remote Sens. Environ. 109 (2) (2007) 166–182.
[16] C.J. Keith, K.S. Repasky, R.L. Lawrence, S.C. Jay, J.L. Carlsten, Monitoring effects of a controlled subsurface carbon dioxide release on vegetation using a hyperspectral imager, Int. J. Greenh. Gas Control 3 (5) (2009) 626–632.
[17] L. Zhi, D. Zhang, J. qi Yan, Q.-L. Li, Q. lin Tang, Classification of hyperspectral medical tongue images for tongue diagnosis, Comput. Med. Imaging Graph. 31 (8) (2007) 672–678.
[18] G. Lu, B. Fei, Medical hyperspectral imaging: a review, J. Biomed. Opt. 19 (1) (2014) 1–24.
[19] R. Dian, S. Li, A. Guo, L. Fang, Deep hyperspectral image sharpening, IEEE Trans. Neural Netw. Learn. Syst. 29 (11) (2018) 5345–5355.
[20] R.M.H. Nguyen, D.K. Prasad, M.S. Brown, Training-based spectral reconstruction from a single RGB image, in: Computer Vision – ECCV 2014, Springer International Publishing, 2014, pp. 186–201.
[21] A. Robles-Kelly, Single image spectral reconstruction for multimedia applications, in: Proceedings of the 23rd ACM International Conference on Multimedia, ACM, 2015.
[22] B. Arad, O. Ben-Shahar, Sparse recovery of hyperspectral signal from natural RGB images, in: Computer Vision – ECCV 2016, Springer International Publishing, 2016, pp. 19–34.
[23] Y. Jia, Y. Zheng, L. Gu, A. Subpa-Asa, A. Lam, Y. Sato, I. Sato, From RGB to spectrum for natural scenes via manifold-based mapping, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, 2017.
[24] J. Wu, J. Aeschbacher, R. Timofte, In defense of shallow learned spectral reconstruction from RGB images, in: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), IEEE, 2017.
[25] N. Akhtar, A. Mian, Hyperspectral recovery from RGB images using Gaussian processes, IEEE Trans. Pattern Anal. Mach. Intell. 42 (1) (2020) 100–113.
[26] J. Li, C. Wu, R. Song, W. Xie, C. Ge, B. Li, Y. Li, Hybrid 2-d-3-d deep residual attentional network with structure tensor constraints for spectral super-resolution of RGB images, IEEE Trans. Geosci. Remote Sens. (2020) 1–15.
[27] J. Jiang, H. Sun, X. Liu, J. Ma, Learning spatial-spectral prior for super-resolution of hyperspectral imagery, IEEE Trans. Comput. Imaging 6 (2020) 1082–1096.
[28] X. Wang, J. Ma, J. Jiang, Hyperspectral image super-resolution via recurrent feedback embedding and spatial-spectral consistency regularization, IEEE Trans. Geosci. Remote Sens. (2021) 1–13.
[29] S. Mei, J. Ji, J. Hou, X. Li, Q. Du, Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks, IEEE Trans. Geosci. Remote Sens. 55 (8) (2017) 4520–4533, http://dx.doi.org/10.1109/TGRS.2017.2693346.
[30] L. Zhang, J. Nie, W. Wei, Y. Li, Y. Zhang, Deep blind hyperspectral image super-resolution, IEEE Trans. Neural Netw. Learn. Syst. 32 (6) (2021) 2388–2400, http://dx.doi.org/10.1109/TNNLS.2020.3005234.
[31] J. Ma, H. Xu, J. Jiang, X. Mei, X.-P. Zhang, Ddcgan: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion, IEEE Trans. Image Process. 29 (2020) 4980–4995, http://dx.doi.org/10.1109/TIP.2020.2977573.
[32] R. Dian, S. Li, X. Kang, Regularizing hyperspectral and multispectral image fusion by cnn denoiser, IEEE Trans. Neural Netw. Learn. Syst. 32 (3) (2020) 1124–1135.
[33] A. Alvarez-Gila, J.V.D. Weijer, E. Garrote, Adversarial networks for spatial context-aware spectral image reconstruction from RGB, in: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), IEEE, 2017.
[34] Y. Yan, L. Zhang, J. Li, W. Wei, Y. Zhang, Accurate spectral super-resolution from single RGB image using multi-scale CNN, in: Pattern Recognition and Computer Vision, Springer International Publishing, 2018, pp. 206–217.
[35] T. Stiebei, S. Koppers, P. Seltsam, D. Merhof, Reconstructing spectral images from RGB-images using a convolutional neural network, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2018.
[36] S. Koundinya, H. Sharma, M. Sharma, A. Upadhyay, R. Manekar, R. Mukhopadhyay, A. Karmakar, S. Chaudhury, 2D-3D CNN based architectures for spectral reconstruction from RGB images, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2018.
[37] C. Yi, Y.-Q. Zhao, J.C.-W. Chan, Spectral super-resolution for multispectral image based on spectral improvement strategy and spatial preservation strategy, IEEE Trans. Geosci. Remote Sens. 57 (11) (2019) 9010–9024.
[38] B. Kaya, Y.B. Can, R. Timofte, Towards spectral estimation from a single RGB image in the wild, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), IEEE, 2019.
[39] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, Y. Bengio, The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2017.
[40] S. Galliani, C. Lanaras, D. Marmanis, E. Baltsavias, K. Schindler, Learned spectral super-resolution, 2017, ArXiv E-Prints arXiv:1703.09470.
[41] A. Rangnekar, N. Mokashi, E. Ientilucci, C. Kanan, M. Hoffman, Aerial spectral super-resolution using conditional adversarial networks, 2017, ArXiv E-Prints arXiv:1712.08690.
[42] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, F. Wu, HSCNN: CNN-based hyperspectral image recovery from spectrally undersampled projections, in: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), IEEE, 2017.
[43] Z. Shi, C. Chen, Z. Xiong, D. Liu, F. Wu, Hscnn+: Advanced CNN-based hyperspectral recovery from RGB images, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2018.

[44] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, H. Huang, Joint camera spectral sensitivity selection and hyperspectral image recovery, in: Computer Vision – ECCV 2018, Springer International Publishing, 2018, pp. 812–828.
[45] S. Nie, L. Gu, Y. Zheng, A. Lam, N. Ono, I. Sato, Deeply learned filter response functions for hyperspectral reconstruction, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018.
[46] Y.B. Can, R. Timofte, An efficient CNN for spectral reconstruction from RGB images, 2018, ArXiv E-Prints arXiv:1804.04647.
[47] L. Zhang, Z. Lang, P. Wang, W. Wei, S. Liao, L. Shao, Y. Zhang, Pixel-aware deep function-mixture network for spectral super-resolution, Proc. AAAI Conf. Artif. Intell. 34 (07) (2020) 12821–12828.
[48] S. Mei, R. Jiang, X. Li, Q. Du, Spatial and spectral joint super-resolution using convolutional neural network, IEEE Trans. Geosci. Remote Sens. 58 (7) (2020) 4590–4603.
[49] K. Wei, A. Aviles-Rivero, J. Liang, Y. Fu, C.-B. Sch?nlieb, H. Huang, Tuning-free plug-and-play proximal algorithm for inverse imaging problems, 2020, ArXiv E-Prints arXiv:2002.09611.
[50] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018.
[51] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019.
[52] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder–decoder for statistical machine translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, 2014.
[53] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[54] G. Huang, Z. Liu, L.V.D. Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017.
[55] A.S. Charles, C.J. Rozell, Spectral superresolution of hyperspectral imagery using reweighted ℓ_1 spatial filtering, IEEE Geosci. Remote Sens. Lett. 11 (3) (2014) 602–606.
[56] Y. Cen, L. Zhang, X. Zhang, Y. Wang, W. Qi, S. Tang, P. Zhang, Aerial hyperspectral remote sensing classification dataset of xiongan new area (matiwan village), J. Remote Sens. 24 (11) (2020) 1299, http://dx.doi.org/10.11834/jrs.20209065.
[57] L. Biehl, D. Landgrebe, Multispec-a tool for multispectral-hyperspectral image data analysis, Comput. Geosci. 28 (10) (2002) 1153–1159.
[58] N. Yokoya, A. Iwasaki, Airborne unmixing-based hyperspectral super-resolution using RGB imagery, in: 2014 IEEE Geoscience and Remote Sensing Symposium, IEEE, 2014.
[59] L. Wald, T. Ranchin, Fusion of images and raster-maps of different spatial resolutions by encrustation: An improved approach, Comput. Environ. Urban Syst. 19 (2) (1995) 77–87.
[60] J. He, J. Li, Q. Yuan, H. Shen, L. Zhang, Spectral response function-guided deep optimization-driven network for spectral super-resolution, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–15, http://dx.doi.org/10.1109/TNNLS.2021.3056181.
[61] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[62] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, A. Goetz, The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data, Remote Sens. Environ. 44 (2–3) (1993) 145–163.
[63] W. Shi, J. Caballero, F. Husz, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, 2016, ArXiv E-Prints arXiv:1609.05158.
[64] J. Wang, K. Chen, R. Xu, Z. Liu, C.C. Loy, D. Lin, Carafe: Content-aware ReAssembly of features, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 3007–3016.
[65] Y.-L. Boureau, N.L. Roux, F. Bach, J. Ponce, Y. LeCun, Ask the locals: Multiway local pooling for image recognition, in: 2011 International Conference on Computer Vision, IEEE, 2011.