Information Fusion 97 (2023) 101812

Contents lists available at ScienceDirect

Information Fusion

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

Spectral super-resolution meets deep learning: Achievements and challenges

Jiang He [a], Qiangqiang Yuan [a,*], Jie Li [a,*], Yi Xiao [a], Denghong Liu [a], Huanfeng Shen [b], Liangpei Zhang [c]

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

ARTICLE INFO

Keywords:
Spectral super-resolution
Deep learning
Multispectral imaging
Hyperspectral imaging
Review

ABSTRACT

Spectral super-resolution (sSR) is a very important technique to obtain hyperspectral images from only RGB images, which can effectively overcome the high acquisition cost and low spatial resolution of hyperspectral imaging. From linear interpolation to sparse recovery, spectral super-resolution have gained rapid development. In the past five years, as deep learning has taken off in various computer vision tasks, spectral super-resolution algorithms based on deep learning have also exploded. From residual learning to physical modeling, deep learning-based models used in spectral super-resolution is multifarious. This paper has collected almost all deep learning-based sSR algorithms and reviewed them according to their main contributions, involving network architecture, feature extraction, and physical modeling. This paper proposed a benchmark about deep learning-based spectral super-resolution algorithms: https://github.com/JiangHe96/DL4sSR, and besides spectral recovery, their potential in colorization and spectral compressive imaging is also systematically discussed. Furthermore, we presented our views about challenges and possible further trends of deep learning-based sSR. Light-weight model architecture with generalization is crucial to in-camera processing. Model robustness should be considered carefully to manage data with various degradation. Finally, multi-task sSR meets the multiple needs of humans and meanwhile achieves inter-task mutual improvement, including low-level with low-level, low-level with high-level, and data reconstruction with parameter inversion.

1. Introduction

Hyperspectral (HS) imaging is famous for its satisfactory spectral resolution, which captures more physical radiation characteristics of objects. Unlike multispectral (MS) or RGB images, each pixel of HS images contains continuous spectra, which has attracted much attention in several fields, such as food safe [1,2], agriculture monitoring [3–5], geological exploration [6–8], environmental monitoring [9–11], medical diagnosis [12,13], and remote sensing [14–16]. Involved both rich spatial and spectral information in one data cube, HS images are good data sources for many further applications, including image segmentation [17], object recognition [18,19], trajectory tracking [20], detection [21–23].

However, due to the unique imaging process and finer radiation capture, HS images always suffer from complex noises, high acquisition costs, low signal-to-noise ratios, and low spatial resolutions. This disadvantage significantly limits the further development of applications requiring high resolutions in spatial and spectral domains. In past years, many researchers utilized fusion-based methods to obtain high-resolution HS images to integrate the spatial details in MS images with finer spectra in HS images [24–29]. In practice, however, the corresponding MS–HS image pairs are hardly available. There are three typical difficulties in these methods. (1) High cost of HS images. Most HS images are for commercial use and not available for free. (2) Complex pre-preparation for HS images. Pre-processing HS images includes radiometric calibration, atmospheric correction, geometric correction, and denoising, where HS denoising is very complex because of the multi-type noise.

Generating HS images directly from MS or RGB images is an affordable way inspired by computational imaging techniques [30]. The image formation of RGB or MS sensors is defined as

$$M(w,h)=\int_{\Lambda}\phi(\lambda)G(w,h,\lambda)d\lambda \tag{1}$$

where $M(w,h)$ is the observed MS or RGB spectra and $G(w,h,\lambda)$ is the ground truth corresponding to the point (w,h) at the wavelength of λ.

* Corresponding authors.
E-mail addresses: hej96.work@gmail.com (J. He), yqiang86@gmail.com (Q. Yuan), jli89@sgg.whu.edu.cn (J. Li), xiao_yi@whu.edu.cn (Y. Xiao), denghong.liu@connect.polyu.hk (D. Liu), shenhf@whu.edu.cn (H. Shen), zlp62@whu.edu.cn (L. Zhang).

https://doi.org/10.1016/j.inffus.2023.101812
Received 31 October 2022; Received in revised form 12 April 2023; Accepted 14 April 2023
Available online 22 April 2023
1566-2535/

$\phi(\lambda)$ denotes the spectral response of sensors at the wavelength of λ. Λ represents the whole receptive spectrum of sensors. In other words, the observed channels are integrations over the receptive spectrum Λ. In discrete settings, it is equal to

$$M(w,h) = \sum_{i=1}^{C} \phi(\bar{\lambda}_i) G(w,h,\bar{\lambda}_i) \tag{2}$$

where $\bar{\lambda}_i$ is the sampled wavelengths. C denotes the hyperspectral channel number. Eq. (2) can be rewritten in tensor form as

$$\boldsymbol{M} = \boldsymbol{H\Phi} \tag{3}$$

where $\boldsymbol{M} \in \mathbb{R}^{W\times H\times c}$ is a c-band MS image with the size of $W \times H$. $\boldsymbol{H} \in \mathbb{R}^{W\times H\times C}$ is the corresponding HS image with C channels. $\boldsymbol{\Phi} \in \mathbb{R}^{C\times c}$ denotes the spectral response functions (SRFs) of MS or RGB cameras. The early attempts are optimizing linear models to achieve spectral super-resolution [31], such as principal component analysis [32] and Karhunen–Loève transformation [33]. These methods aim at searching for a perfect linear combination p_j that represents each hyperspectral pixel through k basis functions v_j

$$\boldsymbol{M} \approx v_0 + \sum_{j=1}^{k} v_j p_j \tag{4}$$

In 2008, Parmar et al. [34] first employ sparse recovery to enhance the band number of RGB images. They represent HS images with sparse representation:

$$\boldsymbol{H} = \boldsymbol{TD} \tag{5}$$

where $\boldsymbol{D} \in \mathbb{R}^{k\times C}$ is the complete sparse HS dictionary that could represent all HS pixels and $\boldsymbol{T} \in \mathbb{R}^{W\times H\times k}$ denotes the transform basis. The goal of sparse recovery-based super-resolution is to extract the perfect $\hat{\boldsymbol{D}}$ from a set of HS images:

$$\hat{\boldsymbol{D}} = \arg\min_{\boldsymbol{D}} \|\boldsymbol{M} - \boldsymbol{\Phi T D}\|_2 + \mu \|D\|_1 \tag{6}$$

where the ℓ_1-norm of $\boldsymbol{D}$ is the regularizer and μ is a parameter related to data fidelity.

Then, Nguyen et al. [35] proposed a hyperspectral data set to recover HS images by training strategy, and learning-based spectral super-resolution methods have sprung up since 2014 [36–40]. Robles-Kelly [36] leveraged material appearance information to better capture features in the training set. Arad et al. [37] used K-means Singular Value Decomposition to extract spectral dictionary. Wu et al. [39] transferred their spatial-super-resolution algorithm and enhanced dictionary learning.

Instead of considering the physical imaging process, another category of methods directly mapping MS images to the HS domain, especially after deep learning becomes more and more popular [41–88]. The goal of deep learning-based methods is

$$\boldsymbol{H} = f(\boldsymbol{M}, \theta) \tag{7}$$

$$\hat{\theta} = \arg\min_{\theta} \|f(\boldsymbol{M},\theta) - \boldsymbol{H}\|_r \tag{8}$$

where $\|\bullet\|_r$ denotes the ℓ_r-norm loss function, usually ℓ_1 and ℓ_2; θ represents the weight parameters of networks. Eq. (8) is the objective function of deep learning-based spectral super-resolution. The goal of deep learning is to find the optimal parameters that can recover the ideal hyperspectral images.

Deep learning can date back to 1958, Rosenblatt et al. [89] proposed perceptron algorithm, which started the first wave of neural network learning. In 1986, Rumelhart and Hinton [90] introduced back-propagating into multi-layer perceptron (MLP) to optimize weights. In 1989, LeCun et al. [91] designed a convolution-based neural network, which is the rudiment of convolution neural networks. Compared with MLP, CNN is famous for its focus on local window, which is more suitable for imaging processing. Thus, in spectral super-resolution, except CNN, there is few other deep learning-based methods. The early deep learning-based work is *DenseUnet* proposed by Galliani et al. [42]. Xiong et al. [44] improved a very deep spatial super-resolution residual CNN to enhance the spectral resolution of RGB images. Rangnekar et al. [43] trained an Unet architecture followed by a 31-kernel 1×1 convolution layer with Generative Adversarial Networks (GAN) to recover hyperspectral images. Han et al. [48] clustered the RGB images using K-means and employed different backpropagation neural networks (BPNNs) to recover different spectral classes. Fu et al. [46] combined spectral responses simulation and spectral recovery with spatial–spectral CNNs. Later, Koundinya et al. [49] compared 2D and 3D convolutions in spectral reconstruction. Wang et al. [56] proposed an optimization-inspired model with spatial–spectral convolutions to reconstruct hyperspectral images. Li et al. [60] utilized an adaptive weighted attention network with dual residual attention blocks to explicitly model interdependencies between channels in 2020. Martínez et al. [61] pretrained the model with the degradation loss using spectral responses and transferred the model into a new spectral data set. Deep learning-based methods have been proved to achieve great advances over traditional sparse recovery in spectral super-resolution. However, because of the rapid development of deep learning, the networks proposed by researchers are massive. A word cloud indicating the most ranked words appearing in the titles of 112 spectral super-resolution papers in the past twenty years is shown in Fig. 1. Therefore, it is appropriate to review the deep learning-based spectral super-resolution methods according to their core concepts.

In this paper, we reviewed 42 papers about spectral super-resolution based on deep learning from 2017 to 2022. Following the development of these methods, their main contributions can be divided into ten strategies as shown in Fig. 2. (i) *Unet-like*. Unet [92] is originally proposed in image segmentation. For its fast running and good performance, many early spectral super-resolution methods utilized Unet as the basic network. (ii) *Residual learning.* Model depth has been proved to be greatly important for performance in deep learning. However, vanishing or exploding gradients hamper the training convergence. Residual learning [93] is an efficient strategy to build deeper network, which attracted many works about spectral super-resolution. (iii) *Generative adversarial network (GAN)*. GAN [94] is proposed to improve the training process by alternative optimizing generator and discriminator. Some researchers regarded spectral super-resolution as image generation and employed GAN to solve it. (iv) *Multi-feature fusion*. Multiple features at shallow or deep stages contain different information that both make great sense to model performance. Densely connection [95] is a common way to reuse these features. Moreover, the bottlenecks keep the low computational complexity. (v) *2D–3D*. Classical convolutions in deep learning are usually 2D. Considering that spatial–spectral features in hyperspectral images are 3D, some works combined 2D convolutions with 3D convolutions in spectral super-resolution. (vi) *Attention*. Inspired by squeeze-and-excitation networks in classification, attentions are widely used in various types of CNNs to improve the spectral recovery. (vii) *Degradation simulation*. In 2018, Fu et al. [46] designed a spectral super-resolution framework containing spectral degradation and recovery though a convolution-based degradation simulation. Based on this work, there are many researches improving the degradation simulation with degradation loss. (viii) *Group recovery*. Realizing the difference among spectral channels or different objectives, many works grouped the input features in spatial or spectral domains and utilized different models to enhance their spectral resolutions. (ix) *Model-embedded learning*. Data-driven modeling for deep learning is a double-edged sword. On the one hand, users need not to learn how the model works and generate results directly by feeding data into the trained model. On the other hand, deep learning-based methods are always blamed for its lack of physical interpretability. To open the black box of deep learning, some works utilized physical models to help CNN-based modeling. (x) *Joint super-resolution*. Besides spectral super-resolution, some researchers jointed

Fig. 1. The word cloud about spectral super-resolution.

Fig. 2. Categorization of spectral super-resolution based on deep learning.

the spectral super-resolution with spatial super-resolution and solved it in an end-to-end network. It is observed that deep learning-based sSR has been developed rapidly in network architecture, feature extraction, and physical modeling.

The motivation of this work is as follows. Firstly, as far as we know, there is no work completely reviews the deep learning-based technologies in spectral super-resolution. Secondly, spectral super-resolution methods could be deployed into many scenarios, including image colorization, spectral recovery, and spectral compressive imaging, while existing works only address one specific task. [96] is a latest review about spectral recovery, while it only discussed spectral recovery and did not proposed a benchmark. In this paper, we propose a comprehensive review about deep learning-based methods and compare the model performance in three applications.

This article is structured as following:

- Section 2 reviews the deep learning-based spectral super-resolution methods.
- Section 3 compares typical methods in three applications.
- Trends and challenges of deep learning in spectral super-resolution are discussed in Section 4. Moreover, some thoughts on future directions are also given.
- A summary of this article is made in Section 5.

2. Method review

Researchers have been studying deep learning-based sSR methods for almost ten years. With the rapid development of deep learning and the more abundant hyperspectral data sets, there have been various works address the deep learning-based sSR. This article mainly surveys 42 works from 2017, as shown in Table 1.

As described in Section 1, The mentioned ten core ideas involve *Network architecture*, *Feature extraction*, and *Physical modeling*. Concretely, network architecture is significant for deep learning-based methods, a suitable architecture leads to a better performance, and the common architectures include Unet-like, Residual learning, and GAN. Secondly, the most important processing of spectral super-resolution is the spatial–spectral feature extraction, and fully mining the spatial–spectral information in the image is beneficial to the recovery of hyperspectral image. Multi-feature fusion, 2D–3D, and Attention are all effective means of spatial–spectral feature extraction. Finally, considering physical imaging model in algorithms is also a new trend in spectral super-resolution. Many works have tried to using physical modeling in their CNNs, and the main ideas include Degradation simulation, Group recovery, and Model-embedded learning. Note that, some works may consist of multiple strategies. To keep their original contributions, this review does not divide them into separate categories.

Table 1
An overview of spectral super-resolution methods based on deep learning.

Methods	Introduction	Strategies
2017alvarez [41]	Unet without batch normalizations trained by PatchGAN	Unet-like; GAN
2017galliani [42]	densely connected Unet	Unet-like; Multi-feature fusion
2017rangnekar [43]	Unet followed by a 1 × 1 convolution trained by GAN	Unet-like; GAN
2017xiong [44]	very deep residual CNN with one global skip connection	Residual learning
2018can [45]	moderately deep residual CNN with 9 layers	Residual learning
2018fu [46]	spatial–spectral CNN with SRF simulation	Degradation simulation
2018han [47]	deep CNN with middle global skip connection	Residual learning
2018han2 [48]	clustering using K-means before multiple BPNNs	Group recovery
2018koundinya [49]	5-layer CNN with 2D convolutions and 3D convolutions	2D–3D
2018shi [50]	densely connected CNN with path-widening fusion	Multi-feature fusion
2018stiebel [51]	Unet considered about non-ideal imaging	Unet-like
2018yan [52]	densely connected Unet with multi-scale convolutions	Unet-like; Multi-feature fusion
2019gewali [53]	pixel-by-pixel reconstruction using a deep residual CNN	Residual learning; Group recovery
2019kaya [54]	moderately deep residual CNN with SRF and class prior	Residual learning; Degradation simulation
2019lore [55]	GAN-trained Unet to refine spectral super-resolution	Unet-like; GAN
2019wang [56]	optimization-inspired model with spatial–spectral convolutions	Model-embedded learning
2020banerjee [57]	Unet with XResnet	Unet-like
2020fubara [58]	learn SRF to achieve supervised and unsupervised spectral recovery	Degradation simulation
2020li [59]	2D and 3D residual attentions with structure tensor constraints	2D–3D; Attention
2020li2 [60]	adaptive weighted attention network with dual residual attention blocks	Attention
2020martinez [61]	transfer the pretrained model with degradation loss	Degradation simulation
2020mei [62]	joint a spatial super-resolution network with a spectral super-resolution network	Joint super-resolution
2020nathan [63]	attention-based network with a dense branch and a Unet-like branch in parallel	Unet-like; Multi-feature fusion; Attention
2020peng [64]	no-pooling residual channel attention and multi-depth features fusion	Multi-feature fusion; Attention
2020stiebel [65]	combine signal formation and CNN with metameric loss	Model-embedded learning
2020wang [66]	deep non-local unrolling network	Model-embedded learning
2020wei [67]	one-shot CNN with the help of spectral unmixing	Model-embedded learning
2020yan [68]	Unet with category and coordinate information	Unet-like
2020zhang [69]	deep CNN with multi-path function-mixture blocks	Multi-feature fusion
2020zhao [70]	densely connected CNNs with PixelUnShuffle to extract multi-scale features	Residual learning; Multi-feature fusion
2021he [71]	end-to-end deep-unrolling CNN with the SRF guide	Group recovery; Model-embedded learning
2021hang [72]	group spectral bands with correlation and reconstruct using multiple residual CNNs	Residual learning; Group recovery
2021li [73]	2D–3D convolution-based CNN with second-order channel attention and structure tensor attention	2D–3D; Attention
2021li2 [74]	dual channel attention networks with degradation loss	Attention; Degradation simulation
2021sun [75]	RGB synthesis with learnable Infrared-cut filter	Degradation simulation
2021zheng [76]	CNN with spatial and spectral residual attentions	Attention
2021zhu [77]	end-to-end CNN by unfolding amended gradient descent progress	Model-embedded learning
2021zhu2 [78]	degradation model-aware network trained by GAN to achieve unsupervised spectral super-resolution	GAN; Degradation simulation
2022chen [79]	semi-supervised spectral attention-based network with spectral degradation	Attention; Degradation simulation
2022han [80]	superpixel clustering before multi-branch BPNNs	Group recovery
2022he [81]	universal optimization-driven CNN combining spatial–spectral fusion with spectral super-resolution	Joint super-resolution; Model-embedded learning
2022li [82]	CNN with hybrid residual attentions and RGB structure information	Attention
2022ma [83]	deep spatial–spectral feature interaction network	Joint super-resolution
2022mei [84]	CNN with spatial–spectral convolutions and band-by-band reconstruction	Group recovery
2022ma2[85]	deep unrolling-based CNN joint spectral and spatial super-resolutions	Joint super-resolution; Model-embedded learning

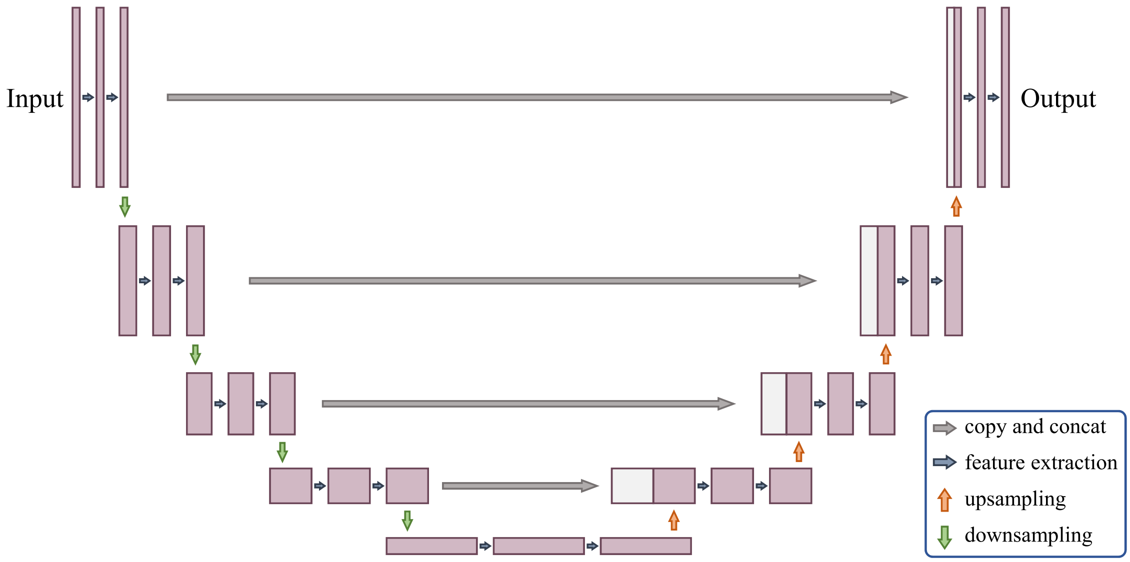


Fig. 3. Unet-like architecture.

2.1. Network architecture

2.1.1. Unet-like

Unet [92] is firstly proposed in image segmentation and attracted wide attentions in various tasks. Due to the fast runtime and multi-scale feature reuse, Unet is the main framework in many early deep learning-based sSR methods, as shown in Fig. 3. In the early work on RGB images, Rangnekar [43] added a 31 band 1 × 1 convolution layer after Unet to increase the channel number. And Alvarez-Gila et al. [41] replaced the Batch Normalization layer in Unet. To improve the feature extraction in Unet, Galliani et al. [42] employed a Tiramisu-like network consists of densely connected blocks. Moreover, Yan et al. [52] introduced multi-scale convolutions into the Tiramisu-like network.

Considering about real-world imaging, Stiebel et al. [51] used Unet to address non-ideal imaging-based sSR. In 2019, Lore et al. [55] firstly refined the concept of spectral super-resolution, including RGB to hyperspectral recovery and selected bands to hyperspectral reconstruction. Moreover, they utilized Unet to discuss different cases. As more and more efficient CNN modules have been brought up, various feature extractions in Unet are proposed. Banerjee et al. [57] improved the Unet with XResnet [97]. Yan et al. [68] injected the category and coordinate information into Unet to enhance the feature extraction.

There are three main reasons why researchers used Unet-like architecture in sSR. Firstly, with multiple down-and-up samplings, Unet could exploit the multi-scale features in images. Secondly, running time on down-sampled features is significantly reduced. So, the computational speed is very high. Lastly, shallow features contain great texture and structure, and reusing shallow features relieves the information loss caused by the downsampling.

However, the weakness of Unet is obvious. On the one hand, downsampling truly boost the spectral recovery, but on the other, the information loss is inevitable. Thus, images generated by Unet-like methods usually lose some spatial details [71].

2.1.2. Residual learning

Network depth makes great sense for model performance in deep learning. Unfortunately, gradients in deep CNNs would vanish during the backpropagation. Residual learning [11] is a strategy to build an identity mapping from shallow to deep modules without neither extra parameter nor computation complexity, which is able to address the vanishing gradients in deep CNNs.

Inspired by these motivations, residual learning is a widely used strategy in spectral super-resolution as shown in Fig. 4. As early as

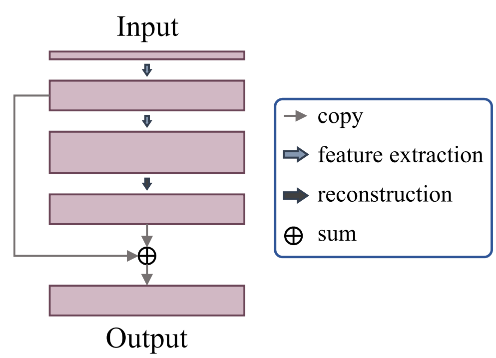


Fig. 4. Residual learning.

2017, Xiong et al. [44] firstly improved a very deep spatial super-resolution residual CNN to enhance the spectral resolution of RGB images. Han et al. [47] utilized a global skip connection to enhance the recovery of high-frequency spectral contents. With global residual connections, Can et al. [45] did not build a very deep CNN while a moderately deep CNN with two residual blocks between two 5×5 convolutions. In 2019, Gewali et al. [53] used a deep residual CNN with 1D convolutions to reconstruct hyperspectral images pixel by pixel. Combining with dense blocks, Zhao et al. [70] extracted multi-scale features with residual dense blocks to explicitly consider spectral context information. Having realized the difference among spectral bands, Hang et al. [72] grouped spectral bands with correlation matrixes and reconstructed them using residual networks respectively.

Although residual learning requires no more parameters or computational complexity, it cannot be directly used in spectral super-resolution. Because the input channel number and the output channel number in sSR are different, the element-wise summation is inapplicable. Thus, more convolutions should be employed before summation, which leads to that the added features are not physically explicable as residuals in other image enhancements, such as denoising, dehazing, and deraining.

2.1.3. GAN

Among early works about spectral super-resolution, some methods regarded sSR as image generation and employed GAN to solve it. GAN [98] is proposed to improve the training process by alternative

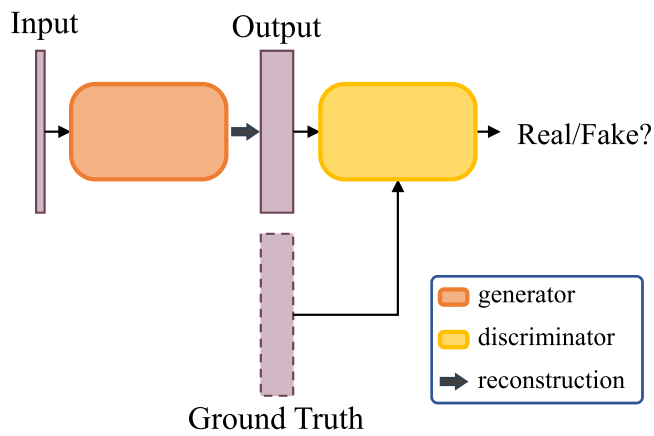


Fig. 5. Generative adversarial network.

optimizing generator and discriminator. Given a generator $G(M;\theta_g)$, a discriminator $D(x;\theta_d)$, the input distribution p_{ms}, and the target hyperspectral distribution p_{hs}, the objective of GAN-based sSR is to play a two-player minimax game, as shown in Fig. 5, with value function $V(G, D)$:

$$\min_G \max_D V(G, D) = E_{x \sim p_{hs}}[log D(x)] + E_{M \sim p_{ms}}[1 - D(G(M))] \tag{9}$$

As early as in 2019, Alvarez-Gila et al. [41] employed PatchGAN [99] as the discriminator to train an Unet-like architecture without Batch Normalization layers to improve the spectral resolution of RGB images. In the same year, Rangnekar et al. [43] trained an Unet architecture followed by a 31 band 1 × 1 convolution layer with GAN to recover aerial hyperspectral images. Lore et al. [55]used conditional GAN to train CNN and discussed their refined concept of spectral super-resolution. After this, there is few sSR methods employing GAN as training. In 2021, Zhu et al. [78] proposed a degradation model-aware hyperspectral image generation and utilized GAN to achieve unsupervised spectral reconstruction.

From the development of GAN-based spectral super-resolution methods, we can find that only the works that regarding sSR as image generation would employ GAN. Obviously, the disadvantage of GAN is the unavoidable fake information in the reconstructed images. So, if the ground truth is unavailable in a task, GAN is a good choice to acquire natural visual performance. However, the spectra recovered by spectral super-resolution should be as similar as possible to the observation of the real hyperspectral sensor. Because, there is physical significance in spectra, which can reflect the radiation characteristics of the observed object.

2.2. Feature extraction

2.2.1. Multi-feature fusion

The memory and features of convolutional neural networks (CNNs) are affected as they deepen, compared with shallow CNNs. Although residual learning is an acceptable approach to bring shallow features into deep layers, the combined features depend on shallow features through identity mapping. In many cases, shallow and deep features are both greatly important in spectral super-resolution. Thus, dense connections [95] for multi-feature fusion is proposed to reuse these features equally, as shown in Fig. 6. Galliani [42] and Yan [52] employed a Tiramisu-like network embedding dense blocks into Unet to address spectral super-resolution. Shi et al. [50] designed and compared two types of CNNs with residual skips, and then found that joining dense connections with path-widening fusion could improve the model performance. Nathan et al. [63] proposed a network with a dense branch and a Unet-like branch in parallel. Peng et al. [64] fused multi-depth features from stacked residual channel attention blocks to improve spectral recovery. In 2020, Zhang et al. [69] proposed multi-path function-mixture blocks to adaptively adjust the learned pixel-wise spectral mapping.

Multi-feature fusion with dense connections always requires an extra bottleneck layer to reduce the concatenated features to significantly boost the computation. If the use of dense blocks is limited, then the parameters introduced by bottlenecks are negligible. However, when dense blocks are used repeatedly in the model, the calculated amount is increased.

2.2.2. 2D–3D

Some researchers suggest that classical 2D convolutions can learn a nonlinear function, considering the spatial–spectral features in hyperspectral images. However, they neglect the deeper inter-channel or spatial feature correlations among middle layers. Human action recognition includes a type of convolution that can capture spatiotemporal features from videos [100]. Differences between 2D and 3D convolutions are shown in Fig. 7. Thus, to exploit the spatial–spectral features in hyperspectral images, 3D convolutions are also used in spectral super-resolution.

In 2018, Koundinya et al. [49] compared the performance of 2D and 3D convolutions for spectral reconstructions from RGB images and presented that 3D CNN is more suitable. Li et al. [59] utilized 2D and 3D residual attentions with structure tensor constraints to recover hyperspectral images. Subsequently, Li et al. [73] further explored the second-order channel attention and structure tensor attention with 2D and 3D convolutions to improve spectral recovery.

From a complete 3D convolution-based CNN to 2D–3D CNN, 3D convolutions in spectral super-resolution do not gain remarkable attention. The latest works only employed 3D convolutions as a branch in parallel. A serious problem with 3D CNN is the great number of parameters in 3D kernels. With 3D convolutions, the models truly reconstruct more ideal hyperspectral images, but require a considerably longer time [73]. Embedding 2D CNN with 3D convolutions can alleviate this problem.

2.2.3. Attention

Since attention mechanisms [101] are proposed in 2017, many works have realized the difference between features. Especially, Hu et al. [102] designed squeeze-extraction networks to capture the inter-band attentions. In spectral super-resolution, attention mechanisms were introduced in the last three years. The main attentions are divided into spatial and spectral attentions, as shown in Fig. 8. In addition, some researchers utilized hybrid attentions by combining the two attentions.

In 2020, Li et al. [60] utilized an adaptive weighted attention network with dual residual attention blocks to explicitly model interdependencies between channels. In the same year, they further improved their residual attentions with structure tensor constraints and 2D–3D convolutions to recover hyperspectral images [59]. Peng et al. [64] removed the pooling of residual channel attention blocks and fused multi-depth features to improve the spectral recovery. Nathan et al. [63] proposed an attention-based network with a dense branch and a Unet-like branch in parallel. In 2021, Li et al. [73] explored a second-order channel attention based on their previous 2D–3D structure tensor attention. Gu et al. [74] combined 2D and 3D branches with different channel attentions and used spectral responses to constrain model optimization. Moreover, Zheng et al. [76] proposed a CNN consisting of different spatial–spectral branches with spatial and spectral residual attentions. In 2022, Li et al. [82] improved hybrid residual attentions again and injected the structure information of RGB input into intermediate features. Chen et al. [79] utilized spectral attention-based networks to achieve semi-supervised spectral super-resolution.

The attention mechanism is an effective strategy to explore the relationships within images. It guides the model to focus on interesting features. Spatial attentions capture the spatial difference, which

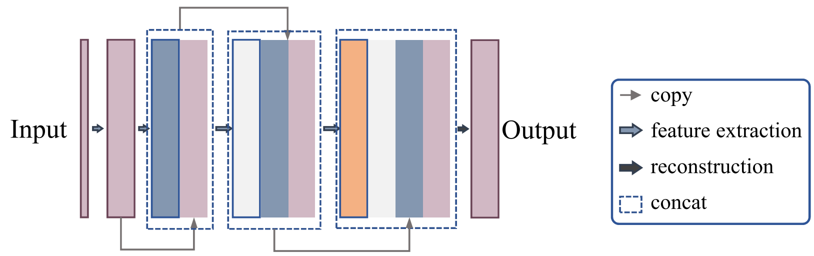
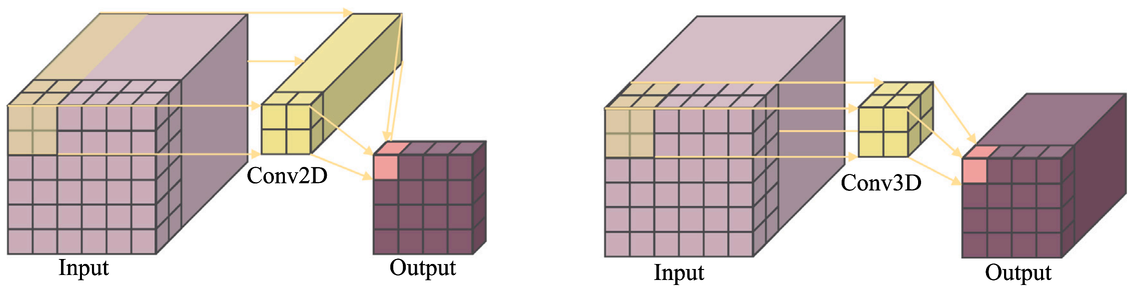
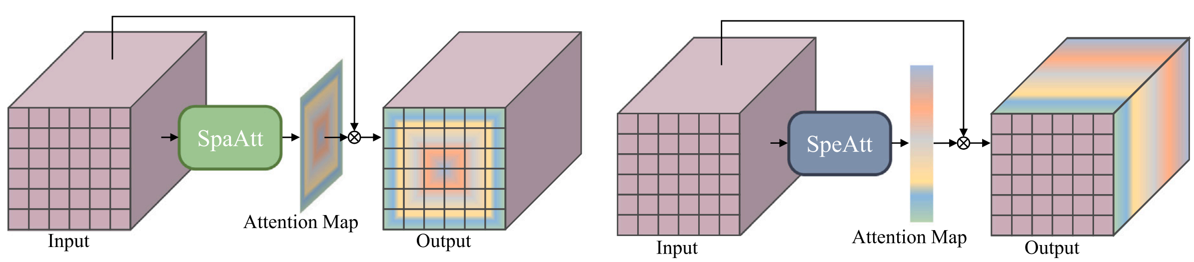


Fig. 6. Multi-feature fusion.



(a) 2D convolutions   **(b)** 3D convolutions

Fig. 7. 2D convolutions and 3D convolutions.



(a) Spatial attentions   **(b)** Spectral attentions

Fig. 8. Spatial attentions and spectral attentions.

may represent the class of the observed objects. Moreover, spectral attentions ensure that the models can distinguish between bands with different radiation characteristics. The classical attentions require few extra parameters, but require more floating-point calculations because of the element-wise product. Especially, more works prefer using hybrid attentions, thereby further increasing computational time.

2.3. Physical modeling

2.3.1. Degradation simulation

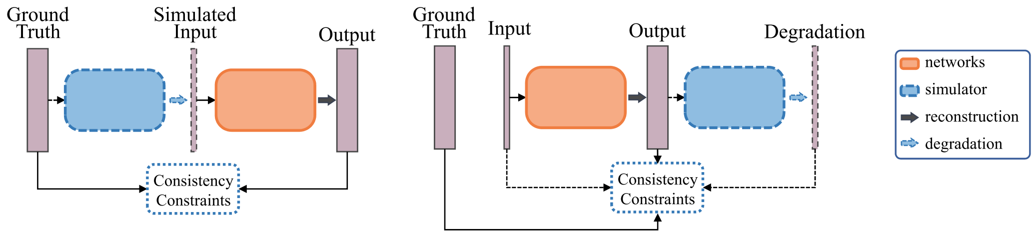
Similar to the classical image reconstruction, spectral super-resolution maps information from RGB domains to hyperspectral domains, exhibiting an ill-posed inverse problem. Thus, RGB images degraded from the recovered hyperspectral images should be consistent with the original RGB images because only one fixed type of degradation occurs during the imaging when using a fixed sensor. Many researchers have focused on appropriately simulating the ideal degradation in their works to introduce the consistency constraints into spectral super-resolution. They can be divided into two groups, including simulation for better input and simulation for better constraints, as shown in Fig. 9. Works using simulation for better input suggest that the quality of the reconstructed images is sensitive to the spectral response functions used to generate the input image [103]. Other studies focused more on enhancing the constraints of the output image, such as spectral degradation loss functions.

As early as 2018, Fu et al. [46] simulated several spectral responses and achieved spectral recovery with spatial–spectral CNNs. Kaya et al. [54] designed two networks to learn spectral responses and classes and discussed the advantages and disadvantages of different types of spectral super-resolution frameworks. Martínez et al. [61] pretrained the model with the degradation loss using spectral responses. Sun et al. [75] simulated an infrared-cut filter to achieve learnable RGB synthesis. Li et al. [74] used spectral responses to optimize their 2D–3D attention-based model. With degradation simulation, some works attempted to achieve unsupervised spectral super-resolution. In 2020, Fubara et al. [58] used the learned spectral basis functions as a guide to achieve supervised and unsupervised spectral recovery. Zhu et al. [78] proposed a degradation model-aware hyperspectral image generation and utilized generative adversarial training to achieve unsupervised spectral reconstruction. Chen et al. [79] utilized physical spectral degradation to achieve semi-supervised spectral super-resolution.

Degradation simulation can effectively improve the constrains of the recovered hyperspectral images, especially when the specific spectral degradation matrices are known. The problem is that errors of degradation simulation further affect the accuracy of spectral super-resolution. Moreover, improving the efficiency and accuracy of degradation simulation is important.

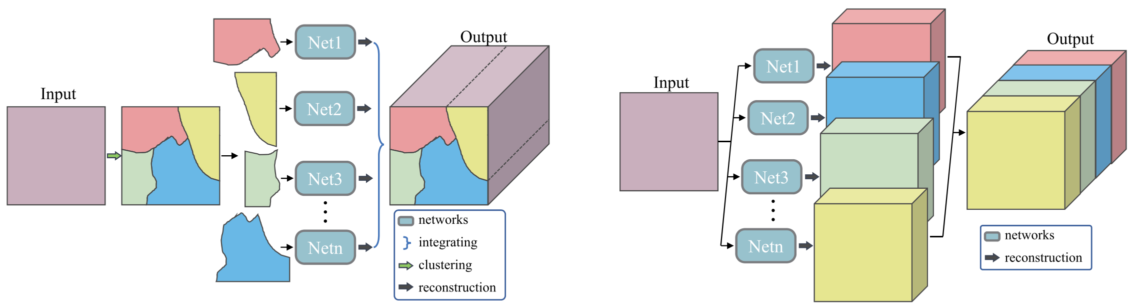
2.3.2. Group recovery

Local and nonlocal similarity in hyperspectral images is very useful to improve the reconstruction in several imaging processes, such as denoising [104], spatial super-resolution [105], and image fusion [106].



(a) Simulation for better input **(b)** Simulation for better constraints

Fig. 9. Degradation simulation.



(a) Spatial group recovery **(b)** Spectral group recovery

Fig. 10. Two group recovery strategies dealing with spatial and spectral domains.

Inspired by these works, the two types of strategies in introducing local and nonlocal similarities into spectral super-resolution are spatial group recovery and spectral group recovery, as shown in Fig. 10.

In the early work utilizing self-similarity in spectral super-resolution, Han et al. [48] clustered the input RGB image using K-means and employed different backpropagation neural networks to recover different spectral classes. In 2022, they [80] further clustered RGB images and used multi-branch backpropagation neural networks to generate hyperspectral images. These works are based on spatial group recovery. For spectral similarity, He et al. [71] grouped spectral bands using spectral responses and applied deep-unrolling networks to recover them. Moreover, Hang et al. [72] grouped spectral bands with correlation matrixes and reconstructed them using residual networks.

However, some works are contrary to this idea. They believe that information in different bands or pixels would interfere with each other. Therefore, they achieved the spectral super-resolution of each pixel or band separately. Gewali et al. [53] used a deep residual CNN with 1D convolutions and reconstructed hyperspectral images pixel-by-pixel. Mei et al. [84] reconstructed hyperspectral images band by band and then optimized them with spatial–spectral convolutions.

Group recovery, whether in spatial domains or spectral domains, could truly inject more physical prior for networks. However, the following question is raised: how many groups do we need? The two extremes in group recovery include the common sSR methods, which do not group images whether in spatial or spectral domains, and the group number is the minimum 1; the other method proposed by Gewali [53] and Mei [84] grouped images pixel by pixel or band by band, where the group number is the maximum. Some works suggested that the improvement of group recovery depends remarkably on the group numbers [72]. They also stated that the most suitable group numbers for different sensors are different. Thus, enabling deep learning-based models to determine group number adaptively is an urgent problem to be solved for group recovery-based algorithms.

2.3.3. Model-embedded learning

Deep learning has truly gained considerable attention given its strong ability to learn complex nonlinear mapping implicitly. However, the black box of deep learning lacks sufficient physical interpretability and causes difficulty to provide a unified framework to deal with various tasks. Physical model-based algorithms, such as total variation, Bayesian maximum a posterior, and other optimization algorithm, are easily deployed into different tasks because they are modeled based on specific physical processes. Since 2017, Sun et al. [107] Sun built a bridge that combines the optimization algorithm with deep learning in image restoration. Many works that use the physical model to build a network structure have been conducted [108,109]. Fig. 11 shows the idea of combining the physical model with deep learning in spectral super-resolution. The common approach might be following the data flow in optimization algorithms and building deep learning-based networks; this method is also called deep unrolling [110–116].

In 2019, Wang et al. [56] proposed an optimization-inspired model with spatial–spectral convolutions to reconstruct hyperspectral images. In the second year, they designed a non-local unrolling network for computational spectral imaging [66]. In 2021, He et al. [71] unrolled a half-quadratic splitting-based method and employed parametric self-learning to achieve spectral recovery. At the same year, Zhu et al. [77] proposed an end-to-end CNN by unfolding amended gradient descent progress. Recently, Ma et al. [85] unfolded an alternative direction multiplier method to achieve joint spectral and spatial super-resolution. Wei et al. [67] combined spectral unmixing with CNN to increase the spectral channels of RGB images. Stiebel et al. [65] assumed that the reconstructed spectra should involve two solutions based on physical model derivation and used CNN to learn one of them.

Model-embedded learning brings the known physical prior into deep learning, which can guide the model to accurately learn the corresponding process to the physical model. Meanwhile, model-embedded learning maintains the end-to-end manner and the data-driven training of deep learning. This advantage alleviates the problem that the traditional physical model requires considerable manual adjustment. However, existing model-embedded learning-based methods are diverse, and no work has proven which type of these methods is the most suitable to spectral super-resolution. Moreover, frequent arithmetical calculations lead to more floating-point operations and longer runtime of model-embedded learning-based deep learning sSR, even with few parameters.

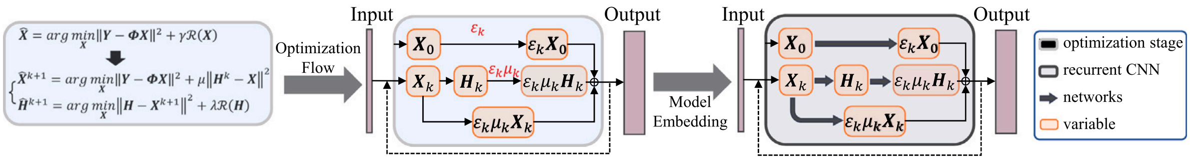
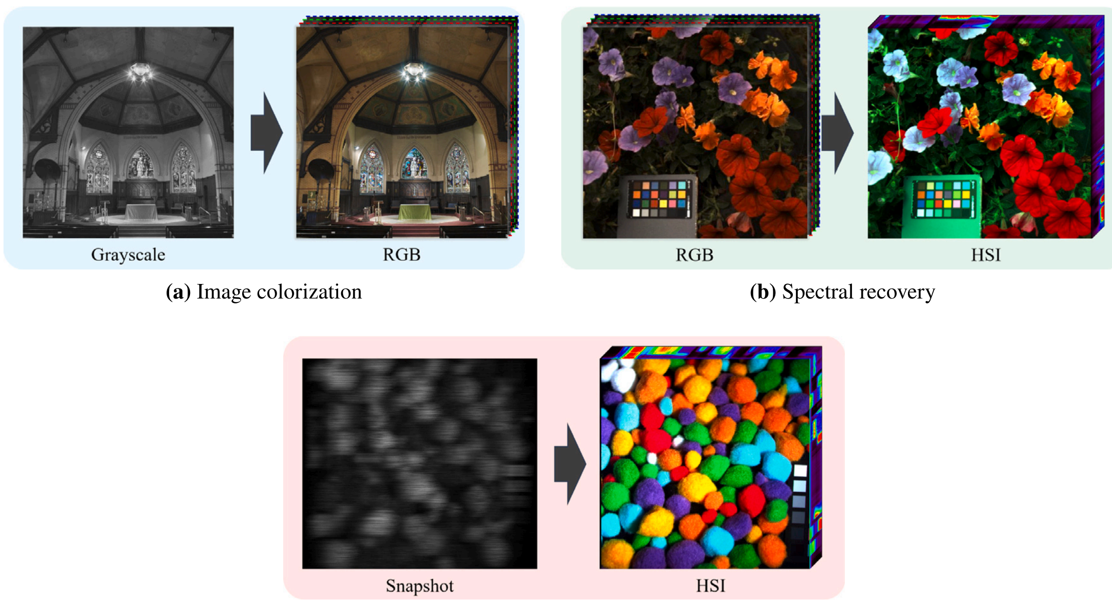


Fig. 11. Model-embedded learning.



(a) Image colorization **(b)** Spectral recovery

(c) Spectral compressive imaging

Fig. 12. schematic for three applications of spectral super-resolution. HSI denotes hyperspectral images.

2.4. Joint super-resolution

In recent years, many works not only achieve spectral super-resolution but also deal with spatial super-resolution in one model, namely, joint super-resolution. The difference between spectral super-resolution and joint super-resolution is whether it increases the spatial resolution simultaneously. The frameworks available to address joint super-resolution include two types. One approach is solving these problems with different models successively, and the other is considering spatial degradation into spectral super-resolution.

The earliest work in joint super-resolution is proposed by Mei et al. [62], where spectral super-resolution networks were stacked with spatial super-resolution networks to simultaneously enhance spatial and spectral resolution of RGB images. In 2022, He et al. [81] proposed a universal CNN by unfolding an optimization-based algorithm considering spectral degradation and spatial degradation. Ma et al. [83] proposed a deep spatial–spectral feature interaction network to generate high-resolution hyperspectral images. In the same year, Ma et al. [85] further unfolded an alternative direction multiplier method to improve the spatial resolution as well as the spectral resolution, which considers the observation models of the spatial super-resolution and spectral super-resolution.

3. Applications and comparisons

Spectral super-resolution can effectively enhance the spectral information of the observed objects to allow individuals to explore the world from another perspective. The sSR-derived applications are denoted to provide more information about the radiation characteristics of captured images. The normal application is spectral recovery, which recovers hyperspectral information from low spectral-resolution images. In this review, we further discuss the potential of deep learning-based sSR algorithms in the two other applications, *i.e.*, image colorization and spectral compressive imaging, which can intuitively demonstrate the importance of spectral super-resolution. Fig. 12 illustrates the difference between three applications.

Spectral recovery builds a bridge between RGB images and hyperspectral images [117]. Hyperspectral images that involve fine spectral radiance properties with numerous bands always suffer from low spatial resolution and high cost of acquisition. RGB and MS images are highly favored for their rapid imaging, high spatial resolution, and ubiquitous data sources. The difference between MS and RGB images are their wavelength range. RGB images only cover visible spectrum with three bands, whereas MS images usually cover a wider spectral range with more spectral bands. Meanwhile, compared with hyperspectral images, no matter RGB and MS images still suffer from low spectral resolution. Spectral recovery is an application of spectral super-resolution that can directly reconstruct hyperspectral images from available RGB images. As shown in Fig. 12(b), we would discuss spectral recovery based on RGB-to-hyperspectral mapping of spectral super-resolution, which is a widely accepted benchmark proposed by *New Trends in Image Restoration and Enhancement* (NTIRE) [118–120].

Image colorization is performed to reconstruct a plausible color version of the observed grayscale photograph to transform black-and-white images into RGB images [121–129]. The given grayscale images only consist of one band in the spectral domain, whereas the reconstructed RGB images have three bands, as shown in Fig. 12(a). Thus, image colorization is also a one-to-three mapping of spectral super-resolution.

Spectral compressive imaging aims at acquiring large volumes of spatial–spectral information of the photographed subject from a set of under-sampled observations, which may result from the restrictions on acquisition time and power consumption in the actual imaging procedure [130–139]. The coded aperture snapshot spectral imaging

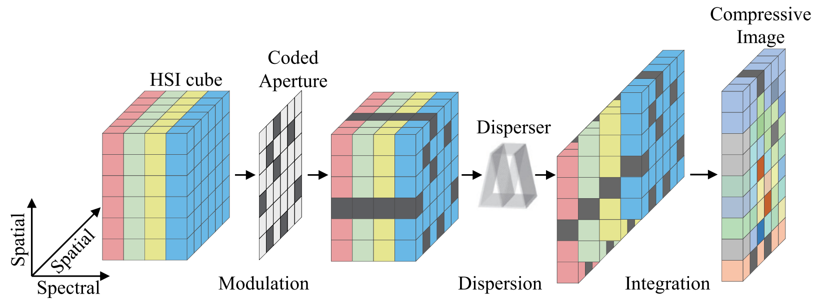


Fig. 13. A schematic of CASSI.

Table 2
Implementation details of the comparison methods in spectral recovery.

Methods	Paper ID	Learning rate	Hyperparameter	Parameter	FLOPs/G	Runtime/s
DenseUnet [42]	2017galliani	0.002	DenseB4; L4; F16	1360.1K	30.158	0.0249
CanNet [45]	2018can	0.0005	ResB2	163.0K	39.675	0.0136
HSCNN+ [50]	2018shi	0.0002	B38; F16	1642.0K	399.672	0.0450
sRCNN [53]	2019gewali	0.0001	L16; F128; K3	789.3K	5955.898	0.0891
AWAN [60]	2020li2	0.0001	M8; DRAB8; F200	34893.9K	4254.316	0.0400
FMNet [69]	2020zhang	0.0001	FMB4; P3; N3; M2; F64	12603.9K	2878.594	0.0333
HRNet [70]	2020zhao	0.0001	F64	63213.9K	610.188	0.0271
HSRnet [71]	2021he	0.0001	OS8	713.4K	158.698	0.0297
HSACS [73]	2021li	0.00005	2DResB16; 3DResB4	19707.7K	4894.209	1.5290
GDNet [77]	2021zhu	0.001	S9	785.664K	191.540	0.0247
SSDCN [79]	2022chen	0.01	B4; F128	415.231K	97.085	0.0157

(CASSI) system is most frequently used. Fig. 13 shows a schematic of CASSI [132]. In CASSI, a coded aperture creates spatial modulation by its transmission function, and a dispersive prism creates spectral shear along a spatial dimension according to the wavelength-dependent dispersion function. Finally, the 2D compressive image is generated by integrating the coded information. As shown in Fig. 12(c), spectral compressive imaging attempts to reconstruct the ideal hyperspectral images from the coded 2D compressive image; it is a one-to-many mapping of spectral super-resolution.

Most importantly, an end-to-end benchmark under the same training and testing frameworks is proposed in this paper.[1] All methods are compared without any other help. To the best of our knowledge, image colorization was initially presented as a one-to-three mapping for coloring cartoon pictures. Therefore, most image colorization methods require weak supervision or class information as guide. As for spectral compression imaging, a one-to-many mapping, the coded aperture is always considered as a prior. However, the selected spectral super-resolution methods do not depend on any other assistance in an end-to-end manner. They can directly increase the spectral band number given a low-spectral-resolution image as input. In this case, we did not compare with a special task method, which considers an extra prior.

3.1. Quantitative metrics

The commonly used quantitative metrics include correlation coefficient, mean relative absolute errors, root mean square error, mean peak signal-to-noise ratio, mean structural similarity, and spectral angle mapper.

Correlation coefficient (CC) ranges 0 to 1, and the larger CC presents the better performance. The formula is as follows:

$$CC = \frac{\sum_{i=1}^{N}\left(X_i-\mu\right)\left(\hat{X}_i-\hat{\mu}\right)}{\sqrt{\sum_{i=1}^{N}\left(X_i-\mu\right)^2 \sum_{i=1}^{N}\left(\hat{X}_i-\hat{\mu}\right)^2}} \tag{10}$$

where N denotes the pixel number; X and $\hat{X}$ denote the reconstructed images and the ground truth; μ and $\hat{\mu}$ represent the mean of X and $\hat{X}$.

Mean relative absolute errors (MRAE) is non-negative, and the closer to 0 indicates the higher fidelity. The formula is shown in below:

$$MRAE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\left|X_i-\hat{X}_i\right|}{\hat{X}_i}\right) \tag{11}$$

Root mean square error (RMSE) is also non-negative, and the closer to 0 indicates the higher fidelity. RMSE can be defined as follows:

$$RMSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(X_i-\hat{X}_i\right)} \tag{12}$$

Mean peak signal-to-noise ratio (mPSNR) is positive, and the higher indicates the less distortion, which is calculated as:

$$mPSNR = \frac{1}{C}\sum_{j=1}^{C} 20\cdot\log_{10}\left(\frac{MAX_I}{RMSE_j}\right) \tag{13}$$

where C denotes the channel number; MAX_I is the maximum of the image. In this paper, because all images are normalized, MAX_I is set to be 1.

Structural similarity (SSIM) [140] ranges 0 to 1, and the larger mSSIM presents the higher fidelity. The formula is shown in below:

$$SSIM = \frac{\left(2\mu\hat{\mu}+C_1\right)\left(2\sigma_{X\hat{X}}+C_2\right)}{\left(\mu^2+\hat{\mu}^2+C_1\right)\left(\sigma^2+\hat{\sigma}^2+C_2\right)} \tag{14}$$

where μ and $\hat{\mu}$ denote the mean of X and $\hat{X}$; σ and $\hat{\sigma}$ are the standard deviation of X and $\hat{X}$; $\sigma_{X\hat{X}}$ is the covariance; C_1 and C_2 are two constants and equal to $0.01\cdot MAX_I{}^2$ and $0.03\cdot MAX_I{}^2$, respectively. Noted that, SSIM is calculated on local windows of the image. We usually use mean SSIM (mSSIM) to evaluate the overall image quality:

$$mSSIM = \frac{1}{K}\sum_{j=1}^{K} SSIM\left(x_j,\hat{x}_j\right) \tag{15}$$

where K is the local window number of the image; x_j and $\hat{x}_j$ are the image blocks at the jth local window.

[1] Benchmark can be found at: https://github.com/JiangHe96/DL4sSR.

Table 3
Properties of five public hyperspectral data sets including indoor and outdoor scenes.

	Sensor	Wavelength (nm)	Channels	Size	Amount
ARAD_1K	Specim IQ	400–700	31	512 × 512	1000
CAVE	Apogee Alta U260	400–700	31	512 × 512	32
ICVL	Specim PS Kappa DX4	400–700	31	1392 × 1300	203
Harvard	Nuance FX	420–720	31	1392 × 1040	50
NUS	PFD-CL-65-V10E	400–700	31	1312 × 1924	64

Table 4
Properties of ten public aerial or satellite hyperspectral data sets.

	Sensor	Wavelength(nm)	Channels	Size
DC Mall	Hydice	400–2400	191	1208 × 307
Urban	Hydice	400–700	210	307 × 307
Pavia U	ROSIS	430–860	103	610 × 340
Pavia C	ROSIS	430–860	102	1096 × 715
Indian Pines	AVIRIS	400–2500	220	145 × 145
Cuprite	AVIRIS	370–2480	224	512 × 614
DFC2018	CASI	380–1050	48	2384 × 601
KSC	AVIRIS	400–2500	224	512 × 614
Chikusei	HH-VNIR-C	363–1018	128	2517 × 2335
Xiong'an	Unknown	400–1000	250	3750 × 1580

Spectral angle mapper (SAM) [141] is non-negative, and the closer to 0 indicates the higher spectral consistency. SAM is defined as:

$$SAM = \arccos\left(\frac{< z, \hat{z} >}{\|z\|_2 \cdot \|\hat{z}\|_2}\right) \tag{16}$$

where z and $\hat{z}$ denote the spectral vectors of X and $\hat{X}$ on the same location; $\langle \bullet \rangle$ denotes the dot product; and $\|\bullet\|_2$ represents the L_2 norm.

3.2. Public hyperspectral data sets

For learning-based methods, training data with ground truth are greatly important. In spectral super-resolution, hyperspectral images are always unavailable. Thus, many works have attempted to propose reliable hyperspectral data sets. Table 3 reports the indoor/outdoor hyperspectral data sets. The indoor/outdoor hyperspectral data sets always cover the wavelength range from 400 nm to 700 nm. The hyperspectral images all have 31 channels corresponding to a RGB image. Among them, ARAD_1K contains the most scenes involving various objectives proposed in NTIRE 2022 and has been the most authoritative data set. CAVE has been a widely used hyperspectral data set and consists of only 32 scenes. ICVL, Harvard, and NUS are three data sets involving large views.

Some aerial/satellite hyperspectral data sets are listed in Table 4. Aerial/satellite hyperspectral images cover wider wavelength with more channels to better reflect the radiation properties of ground objects. These data sets always consist of one scene but with classification labels. Some works also use these hyperspectral images to evaluate the spectral super-resolution algorithms.

We collect public hyperspectral data sets given that high-quality hyperspectral data sets are essential for learning-based algorithms. The online links of these data sets are included in our benchmark.

3.3. Experimental setting

3.3.1. Comparison methods

In this work, we select 11 methods, including DenseUnet [42], CanNet [45], HSCNN+ [50], sRCNN [53], AWAN [60], FMNet [69], HRNet [70], HSRnet [71], HSACS [73], GDNet [77], and SSDCN [79]. These methods are relatively representative in the network architecture, feature extraction, and physical modeling. Furthermore, we have collected as many codes as possible.

3.3.2. Assessment details

We select different quantitative quality metrics for three applications. For spectral recovery, the spectral fidelity is more important. Thus, in addition to MRAE and RMSE in the NTIRE 2022 [120], we used mPSNR in decibel units and SAM as the quantitative indexes. For image colorization, we use CC, mPSNR, and mSSIM. In spectral compressive imaging, mPSNR, mSSIM, and SAM are used to evaluate the reconstruction quality of the comparison methods.

3.3.3. Implementation details

All codes are implemented with their recommended parameter settings. Table 2 lists details in spectral recovery, including the model super-parameters and learning rates. The models are trained by Pytorch framework running in the Linux environment with 64 GB RAM and one Nvidia RTX A5000 GPU. We employ the adaptive moment estimation (Adam) [142] optimizer (β_1=0.9 and β_2=0.999) for 200 epochs to train models with Cosine Annealing scheme [143].

3.4. Computational complexity and speed

Table 2 reports the parameter numbers, floating-point operations (FLOPs), and runtime in the last three columns. HRNet employed many densely connected blocks to extract multiscale and multi-depth features, thereby requiring the most parameters. CaNet is famous as a moderately deep residual CNN with the least parameter and with fast-running speed. With multiple downsampling and dense bottlenecks, although DenseUnet requires 10 times as many parameters as CanNet, its FLOP is low. Moreover, the FLOP of sRCNN is the highest due to the pixel-by-pixel reconstruction, but its parameter number is small. The parameter in AWAN is only half as many as in HRNet; it requires seven times more FLOPs for more residual attention blocks. HSRnet requires less parameters than DensUnet but five times more FLOPs due to the extra addition and subtraction operations in deep unrolling. In terms of running time, CanNet and SSDCN are two most rapid small models. On the contrary, HSACS requires a hundred times more SSDCN for 3D convolutions.

3.5. Spectral recovery

NTIRE workshop aims to provide an overview of the new trends and advances in image restoration, enhancement, and manipulation; it is the largest and the most famous competition including spectral recovery track. Therefore, we utilize the available ARAD1K data set provided by NTIRE 2022 as a benchmark to evaluate the spectral recovery performance of the comparison methods in this work.

Table 5
Quantitative results of eleven spectral super-resolution methods in three applications. The best is highlighted in bold and the second-best is underlined.

Methods	Spectral recovery				Image colorization			Spectral compressive imaging		
	MRAE	RMSE	mPSNR	SAM	CC	mPSNR	mSSIM	mSSIM	mPSNR	SAM
DenseUnet [42]	0.3710	0.0518	26.2478	8.1449	0.9629	26.1123	0.9306	0.7038	33.5929	14.3444
CanNet [45]	0.4254	0.0601	24.7973	6.1849	0.9642	26.8944	0.9440	0.7003	33.3752	14.5810
HSCNN+ [50]	0.3627	0.0553	25.4941	5.6498	0.9644	26.9500	0.9445	0.6918	32.7866	15.3633
sRCNN [53]	0.4768	0.0660	24.1108	7.5244	0.9643	26.8817	0.9441	0.5483	29.7971	19.3639
AWAN [60]	0.2133	0.0379	30.3096	**5.4097**	0.9633	25.3571	0.9317	**0.7610**	**35.1526**	**12.6573**
FMNet [69]	0.3728	0.0486	26.7271	5.7330	0.9632	26.5658	0.9412	0.6729	32.9031	16.7854
HRNet [70]	0.3624	0.0544	25.7755	5.5840	**0.9646**	26.8922	0.9440	0.7209	33.6897	13.9708
HSRnet [71]	0.2462	0.0436	29.2922	6.0272	0.9636	26.5587	0.9403	0.7096	33.5001	14.0956
HSACS [73]	**0.2015**	**0.0320**	**31.4928**	6.0432	0.9645	26.8579	0.9436	0.7095	33.4033	14.3408
GDNet [77]	0.3300	0.0461	27.1448	6.5008	0.9638	26.8986	0.9406	0.6107	29.8844	22.1558
SSDCN [79]	0.3000	0.0585	26.8972	6.2957	0.9644	**26.9535**	**0.9448**	0.6378	31.7991	15.8913

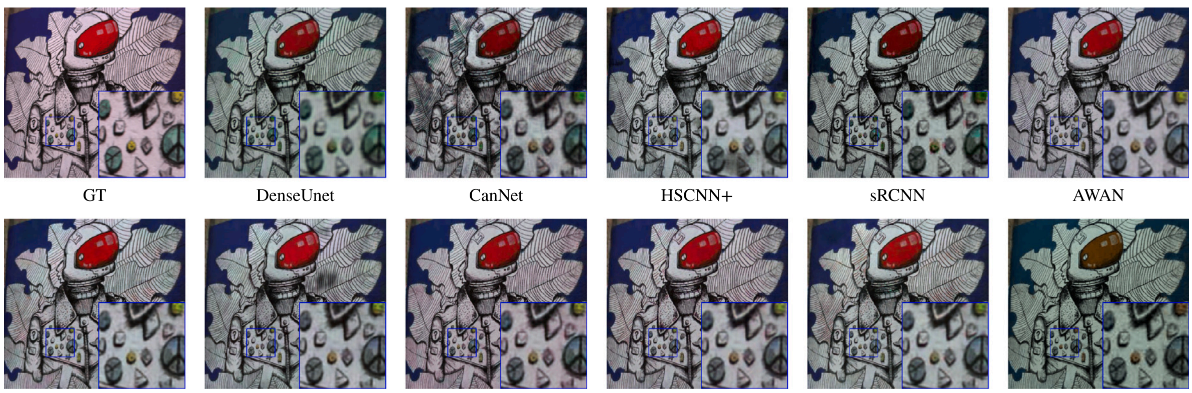


Fig. 14. Spectral recovery results of eleven deep learning-based methods, the chosen true-color images is "ARAD_1K_912".

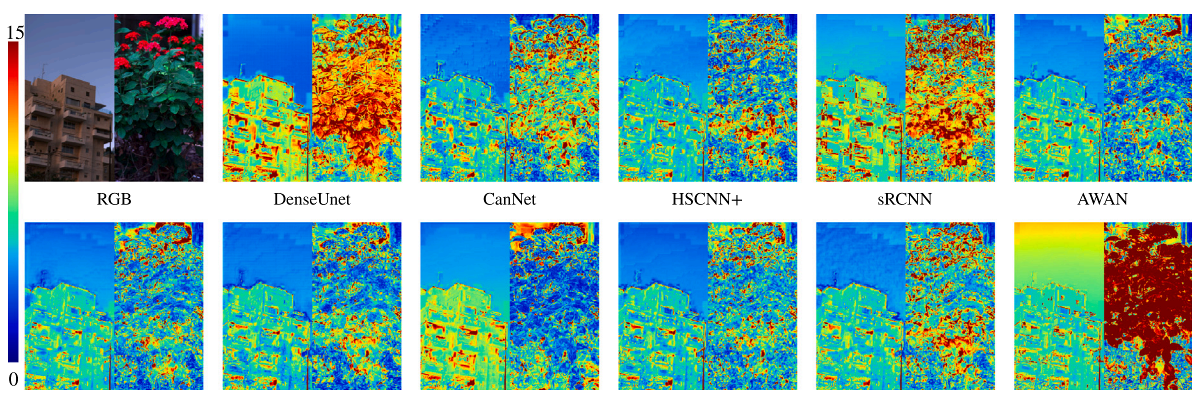
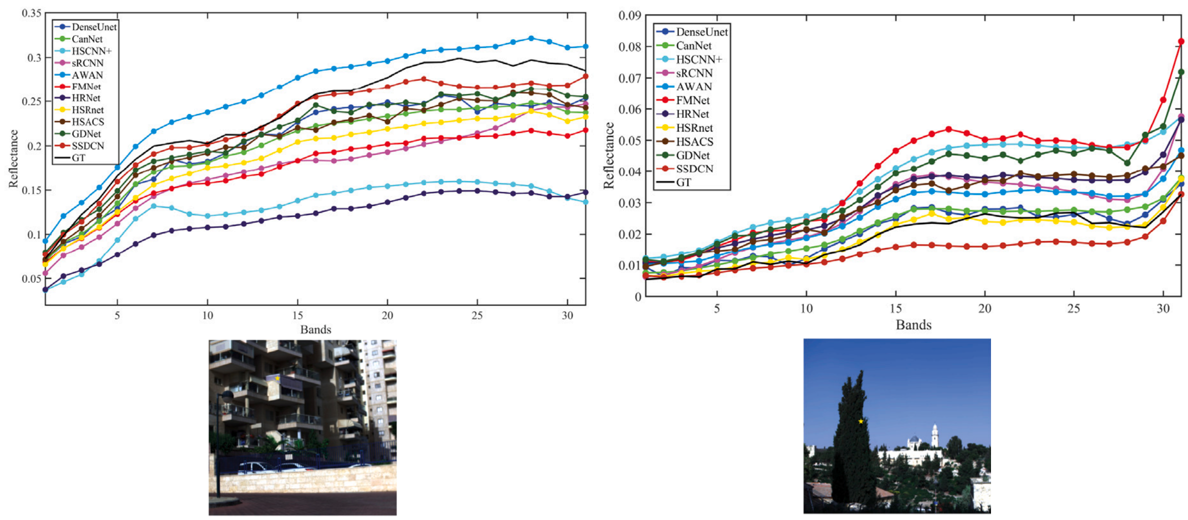


Fig. 15. Spectral recovery results of eleven deep learning-based methods. Different color shows different SAM values, as defined in color bar.

ARAD1K data set consists of 1000 images with the size of 482×512; it is divided into training images (900 RGB-HS pairs), validation images (50 pairs), and test images (50 pairs). We use 900 images as training data and 50 validation images to perform quantitative evaluation because NTIRE 2022 has only released the training and validation images. Limited by the CodaLab platform, NTIRE 2022 only requires participants to upload the central 226×256 region of results for quantitative evaluation. Moreover, MRAE, RMSE, mPSNR, and SAM are calculated over the entire region.

Table 5 lists the quantitative results of 11 deep learning-based spectral super-resolution algorithms. HSACS achieves the best performance among all indexes, and AWAN obtains the second-best results. The two models are considerably ahead of other algorithms in all indicators, especially in mPSNR. Nevertheless, SAM of HSACS is even higher than HSRnet. In addition, the good performance of HSACS comes with the high computational cost. HSACS runs 50 times longer than AWAN, 70 times longer than DenseUnet, and 150 times longer than CanNet. sRCNN achieves the worst MRAE, RMSE, and mPSNR, indicating that pixel-by-pixel recovery harms the spatial structure. DenseUnet obtains the highest SAM and shows poor spectral maintenance. CanNet, which shows poor spatial fidelity, obtains relatively low SAM among these methods.

Fig. 14 display the true-color-synthesis image of data named "ARAD_1K_912" with bands 27, 17, and 10. sRCNN, DenseUnet, and SSDCN show severe spectral distortion as the emerald in the upper right corner of zoom-in box turn into other colors. DenseUnet suffers blurring especially in edges. In general, AWAN, HSRnet, and HSACS perform effectively in spatial details and spectral fidelity. To discuss the spectral maintenance of these deep learning-based methods on different targets, Fig. 15 presents the SAM maps of the two images. Their main objectives are buildings and a bushy tree. Except SSDCN, all methods obtain low SAM on the background. FMNet shows better spectral fidelity around



(a) ARAD_1K_946

(b) ARAD_1K_948

Fig. 16. Spectral recovery results of eleven deep learning-based methods. We display the recovered spectra at different locations.

Fig. 17. Colorization results of eleven deep learning-based methods on two selected gray images.

building edges. Moreover, HSCNN+ and FMNet show superiority on the buildings with repetitive regular geometry. HSRnet and AWAN recover better spectra of leaves with lower SAM. To further discuss this phenomenon, we select two pixels on buildings and trees and display the recovered results generated by all methods, as shown in Fig. 16. Similar conclusions could be found, that is, HSRnet can reconstruct more similar reflectance as real spectra of vegetation. Moreover, the spectra recovered by FMNet show high consistency with real buildings.

3.6. Image colorization

In this review, we selected SUN attribute database [144] to compare the performance of the selected sSR methods in image colorization. SUN attribute database includes 14 340 RGB images in “jpg” format with more than 700 categories, involving vehicles, buildings, sights, and indoor scenes. We only used the 872 images starting with the letter “a” to build our training and test data in this work. We randomly

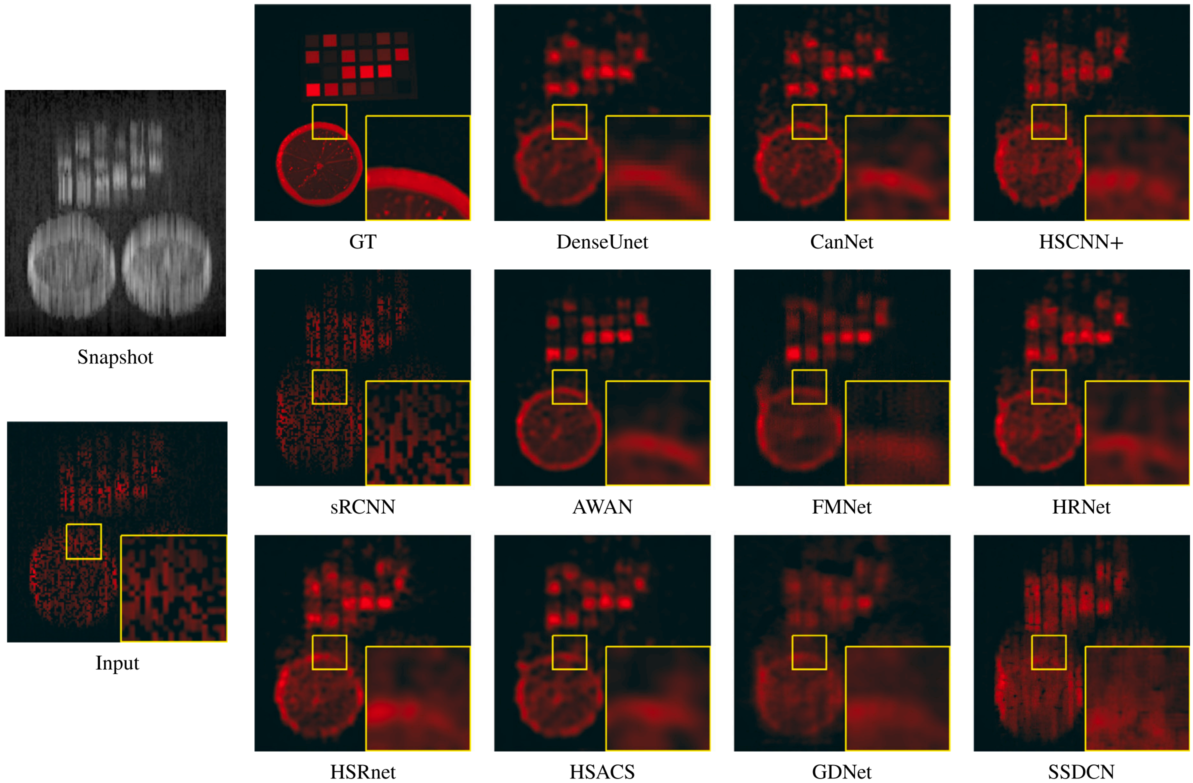


Fig. 18. Spectral compressive imaging results of eleven deep learning-based methods. "Snapshot" denotes the observed snapshot single-band data. "Input" means the shift-back data cube following [133]. Except "Snapshot", we displayed the band 27 for all hyperspectral images.

take 42 images with the size over 450 × 450 for test, where only the 450 × 450 central areas are used. Then, we cut the remaining 830 images into 256 × 256 patches and obtain 3524 training samples. Following [145], we transformed RGB images into $YCbCr$ color space and used the luminance Y as the input grayscale image.

Quantitative results are listed in Table 5, where all the methods show similar colorization performance. SSDCN and HSCNN+ achieve relatively better results. SSDCN obtains the highest mPSNR and mSSIM, whereas HSCNN+ achieves the second best in the two indexes. This phenomenon reveals that image colorization is a one-to-three mapping, which can be appropriately dealt with by either shallow or deep networks. Furthermore, the gap between the best and the worst results is not evident. Fig. 17 displays the colorization results on two selected gray images. Results on the first image truly obtain natural colors and are in better accordance with human vision. Although the gap of colorization performance is tiny, DenseUnet suffers some spatial degradation. Compared with the results on the second image, image colorization using spectral super-resolution algorithms performs well on the images where the overall tone is more consistent. They can hardly achieve good performance when dealing with images with various colors.

Results in this application have proven the potential of deep learning-based spectral super-resolution algorithms in image colorization, and they also revealed that the existing sSR algorithm could hardly mine the spectral diversity when the input information is insufficient because RGB can be regarded as a relatively low-dimensional spectral information. Thus, designing a robust model that could extract more accurate spectral features and handle with various spectral mapping is a good route to further improve the sSR performance.

3.7. Spectral compressive imaging

We followed the same spectral compressive imaging experimental procedure as in [133], including the simulated CASSI system and the preprocessing prior to reconstruction, to maintain consistency with mainstream works. The training data and test images used are obtained from CAVE database [146]. CAVE database consists of 32 hyperspectral images with the size of 512 × 512. We selected 26 images randomly as training data and the remaining images for test. All samples were cropped into 256 × 256 patches. After data augmentation, we obtained 832 training samples.

As listed in Table 5, AWAN performs great potential for achieving spectral compressive imaging. All methods obtained a high SAM, indicating that the generalization of all spectral super-resolution-based algorithms in this domain is limited. Moreover, the highest mSSIM is 0.7610, although mPSNR is up to 35, implying that the spatial structure of the recovered images may be insufficient. As shown in Fig. 18, only few methods could recover the original shape under the special degradation of CASSI. Among the 11 deep learning-based algorithms, AWAN reconstructs the completed lemons and color squares with the highest similarity the same as the real hyperspectral image. FMNet could obtain rough shapes; however, it also produces some artifacts around the upper left corner of the lemon. sRCNN suffers from the coded aperture mask for the pixel-by-pixel processing.

Results in spectral compressive imaging indicate the limitations of deep learning-based sSR algorithms. Under a special degradation, such as CASSI, CNN-based algorithms can hardly learn the accurate one-to-many mapping. In CASSI, the original purpose of the coded aperture mask is to maintain as much spectral information as possible. Moreover, algorithms designed for spectral compressive imaging always consider the mask through their complete framework. However, spectral super-resolution algorithms only consider the spectral mapping from low-spectral-resolution images to high-spectral-resolution images. For spectral super-resolution algorithms, the coded aperture mask even obstructs them from preserving spatial fidelity. According to this assumption, we build one more experiment. We simplify the degradation and only maintain the integration in CASSI. In other words, we calculated the mean image in spectral dimension and obtained the

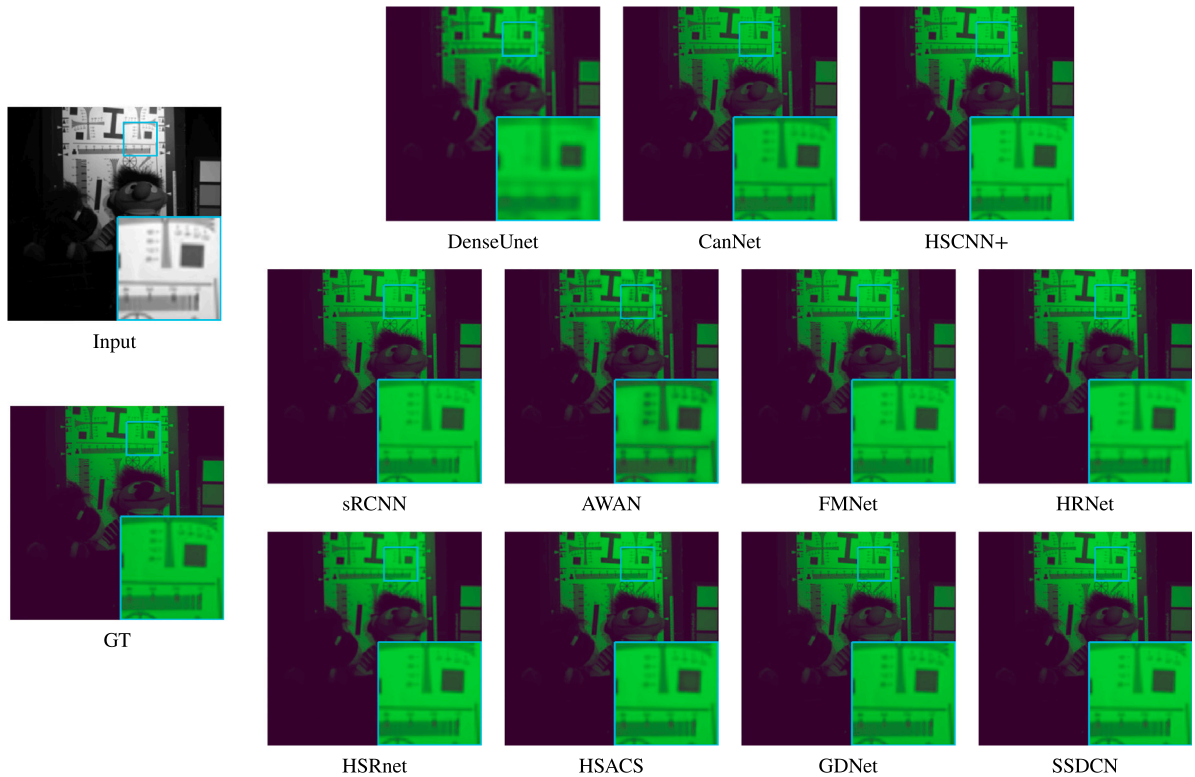
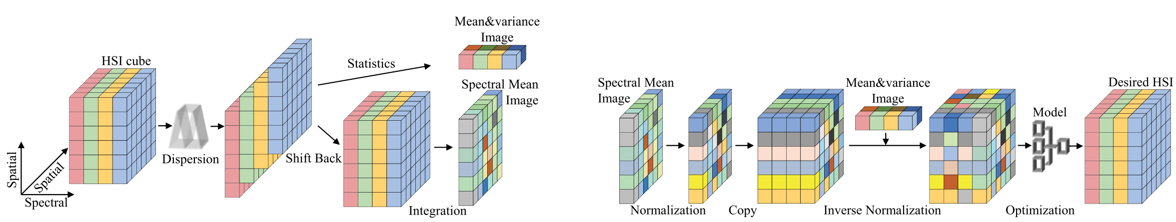


Fig. 19. Spectral compressive imaging results with mean images as input. "Input" denotes the mean images of the ground truth hyperspectral image. We displayed the band 17 for all hyperspectral images.



(a) Observation **(b)** Reconstruction

Fig. 20. A schematic of a new assumptive spectral imaging.

Table 6
Quantitative results of eleven spectral super-resolution methods in spectral compressive imaging under the new assumption. The best is highlighted in bold and the second-best is underlined.

	DenseUnet	CanNet	HSCNN+	sRCNN	AWAN	FMNet	HRNet	HSRnet	HSACS	GDNet	SSDCN
mSSIM	0.8727	0.9158	0.9316	<u>0.9340</u>	0.8737	0.9325	0.9246	0.9113	0.9302	0.9237	**0.9362**
mPSNR	36.3706	34.9508	37.2827	37.3718	33.4847	**37.4866**	37.0962	36.2611	<u>37.4726</u>	35.6169	37.3850
SAM	14.4575	17.5100	14.5470	14.4883	21.0537	14.5010	14.5325	15.2748	<u>14.2508</u>	16.3637	**14.0658**

degraded single-band observations. Quantitative results are listed in Table 6. The table shows that, on the average, mSSIM increased by 35% and mPSNR increases by 4 dB. However, SAM also increases by 3%. The increase in spectral errors proved that the coded aperture masks truly helped in maintaining spectral fidelity. In our new protocol, SSDCN achieves the best performance. HSACS, sRCNN, and FMNet also obtain good recovery effects. Visual results are shown in Fig. 19. Compared with Fig. 18, the reconstructed images are improved in the spatial domain. DenseUnet exhibited some blurring, and AWAN showed some shadows around the numbers.

In conclusion, although deep learning-based sSR algorithms show weakness in handling spectral compressive imaging, they also exhibit their potential to achieve one-to-many mapping. On the one hand, the experimental results indicate that the universality of spectral super-resolution algorithms requires further study. On the other hand, the form of efficient spectral imaging is unnecessarily limited to spectral compressive imaging. Here, we provide a new assumption about spectral imaging, as shown in Fig. 20. The observed images are a single-band spectral mean image with a mean and variance image. When hyperspectral images are required, people could normalize the spectral mean image and copy it to build an image cube with the same size as the hyperspectral images. Then, the means and variances of different bands are used to achieve the inverse normalization. Finally, a CNN-based model is built to optimize the generated hyperspectral images. In this approach, various algorithms can be utilized to achieve spectral imaging with low cost as well as spatial–spectral fidelity.

4. Trends and challenges

Spectral super-resolution is greatly import to enhance the physical radiation characteristic of targets. Experiments proved that spectral super-resolution helps in computational synthesis of hyperspectral images of available low-spectral-resolution data, better perception of the world of gray-scale images, and recovery of hyperspectral data from the low-cost snapshot. Thus, spectral super-resolution covers the shortage of sensors in the form of computational imaging, which yields twice results with half the effort of improving the sensors.

Since the early 1990s, spectral super-resolution has been developed over decades with the efforts of many researchers and satisfied the wave of deep learning after 2015. From Unet, ResNet, and GAN to attention, joint data- and model-driven approach, deep learning is growing fast and certainly boost spectral super-resolution. While deep learning brings benefits, it also comes with challenges.

First, existing deep learning-based methods have achieved satisfactory visual effects consistent with human eyes, for further better results, lightweight parameters, fast running speed, and strong generalization ability; the development of model architecture still has a long way to go. Second, current studies are based on clean images, whereas images captured in the real world usually suffer various quality degradation, including image noise, rains, fogs, clouds, overexposure, and underexposure. Improving the algorithm robustness is also necessary in deep learning-based spectral super-resolution. Finally, the future trend of image processing is cooperative multitasking. Humans are always greedy, and we particularly feel the desire for fulfilling multiple tasks through a single model. Exploring the commonality among different tasks and leveraging useful information among them is the key to realize multi-task learning.

In accordance with these concepts, we propose further trends and challenges in spectral super-resolution, where three aspects are involved, namely, model architecture, robustness, and multi-task learning.

4.1. Model architecture

Spatial–spectral feature extraction. Deep learning-based spectral super-resolution aims to extract spatial–spectral features and learn the mapping from low-spectral-resolution domain to high-spectral-resolution domain through big data set. The key of these algorithms is their ability to extract spatial–spectral features. In the development of deep learning-based spectral super-resolution, studies utilized U-net, very deep networks, GAN, or attention mechanism and explored more relationships between spatial–spectral features. Other studies injected physical knowledge into the model to enhance its ability of spatial–spectral feature extraction. Compared with traditional algorithms or early networks, existing models have made great progress in performance. However, they currently face two problems. First, using the same convolution kernel to reconstruct different ground objects may not be the best option [147]. Second, convolutions can only learn the local relationship but hardly consider the influence of long-range objects. As mention in Section 2.3.2, some works have attempted to solve the first problem using group recovery. However, the grouping-then-reconstruction processing breaks the end-to-end approach of deep learning and increases the propagation error. The main method to solve this problem is building a network that could extract various spatial–spectral features differentially. Moreover, for the second problem, we can determine some inspiration in the latest achievements of deep learning. Transformer is famous for its strong ability to capture global information and long-range interaction on similar objects [148,149]. As transformer has shown promising performance in hyperspectral image processing [150–153], some works attempted to introduce transformer into spectral super-resolution and were successful [154–156]. However, lack of focus on local information hinders the more general development of transformer in spectral super-resolution. Thus, integrating CNN with architecture that can capture global long-range relationship is the better solution of the second problem.

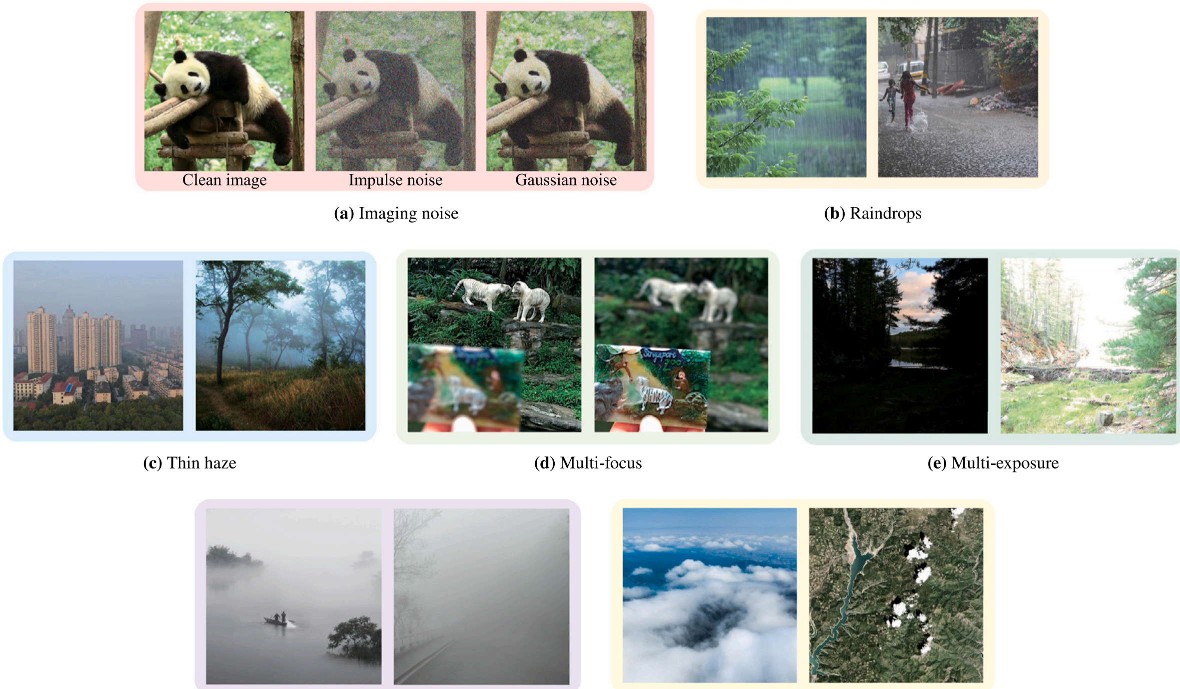
Algorithm complexity and running speed. Spectral super-resolution is always employed as an image enhancement after sensor imaging, providing more spectral information for the subsequent applications that involve high-level image semantic. Existing state-of-the-art algorithms usually depend on a large number of parameters and extremely frequent floating-point operations, improving the reconstruction performance as well as increasing the computational burden on the processor. The future trend of image enhancement would gradually move towards computational imaging, *i.e.*, in-camera processing. To realize in-camera processing, algorithm complexity should be reduced, and the runtime speed should be increased. More efficient convolutional operator [157–160], better optimization algorithm in training [161], and model compression [162–165] might be some possible solutions for this problem. However, when transferring these methods into this task, we should also consider the physical characteristics of spectral super-resolution. Significantly, the improvement should not affect the model ability to learn the spectral degradation; otherwise, it would reduce the performance.

Generalization. Different sensors capture images in various spectral resolution due to the different spectral response functions. The major stumbling block in model generalization of spectral super-resolution is that algorithms are trained in the single specific degradation. It results in a fixed mapping that the networks could learn. Further study is required for the development of models to handle diverse types of spectral degradation that are applicable in real-world scenarios. Furthermore, the fact that CNNs cannot adaptively adjust the number of filters according to input and output also limits the generalization of deep learning in spectral super-resolution. One feasible solution seems to be the head-and-tail replaceable network. The idea is pre-training a backbone network in the high-dimensional feature domain using an overcomplete hyperspectral data set involving various sensors, and then multiple head and tail networks are trained for different RGB input and hyperspectral output data. In this manner, we are only required to select and stack the most suitable head and tail for addressing different spectral super-resolution scenarios, greatly improving the algorithm generalization.

4.2. Robustness

Low-quality data. As shown in Figs. 21(a) to 21(e), images in real world always suffer from various quality degradation due to uncontrolled imaging environment, including imaging noise [166], raindrops [167], haze [168], multi-focus [169,170], and illumination changes [171]. Therefore, improving the algorithm stability of spectral super-resolution is necessary to deal with complex low-quality data. Fu et al. [172] noticed the importance of interference factors and used external–internal learning to obtain more robust algorithms. Unfortunately, the internal learning utilized is similar to the self-supervised mechanism, which is highly dependent on the comprehensiveness of training data. Existing spectral super-resolution data set usually involves one type of interference factor at most. A data set that contains clean hyperspectral and real-world low-spectral-resolution image pairs is necessary to increase the algorithm robustness of spectral super-resolution. This data set should consider as many interference factors as possible, including digital photograph and satellite data. Furthermore, the images should have better covered diverse objects and involve various climates and light conditions.

Information missing. In addition to these quality degradation factors shown in Figs. 21(f) and 21(g), an extreme condition attributed to the thick fog or clouds includes missing information. Covered by thick fog, the captured photographs can only show the details of close-range objects, whereas targets in the distant view lose their own radiation characteristics. Clouds always appear in the satellite images, directly



(a) Imaging noise **(b)** Raindrops

(c) Thin haze **(d)** Multi-focus **(e)** Multi-exposure

(f) Thick haze **(g)** Clouds

Fig. 21. Different types of disturbance suffered by real-world images that affect the algorithm robustness.[2]

cover the original ground targets, and seriously influence the subsequent remote-sensing image interpretation. Existing data sets ignore the information missing in spectral super-resolution and obtaining clean hyperspectral images may be relatively difficult. We can build a spectral super-resolution data set that is even more challenging. Low spectral-resolution images in this data set are covered by different levels of clouds and fogs, which are synthesized by the algorithm. Introducing synthesized prior into original spectral super-resolution data set to overcome information missing would further strengthen the algorithm's robustness.

Lack of real spectra. Existing spectral super-resolution algorithms couldachieve great effect because they are based on a hyperspectral data set. Supervised by the prior knowledge of data sets, algorithms can learn the relationship between high and low spectral-resolution domains effectively. In practical application, however, we can hardly obtain satisfactory real spectra all the time, especially when we have a specific research subject. In remote sensing, most satellites revisit the same area periodically, except the geostationary satellite. Thus, when researchers are interested in a particular area at a specific time, they can hardly obtain the real spectra. Semi-supervised [175] and unsupervised learning [176] allows us to train the model with few (or even without) samples. Some researchers have focused on semi-supervised or unsupervised spectral super-resolution algorithms and achieved good performance [58,77,79]. In the future, exploiting hyperspectral prior in pre-existing data sets and employing transfer learning with the spectral response functions of the target sensor can further overcome the lack of real spectra in spectral super-resolution algorithms.

Spectral interpolation. The current spectral super-resolution method is discussed based on the assumption that the input image is degraded from the corresponding hyperspectral wavelength range, for example, using spectral response functions of RGB in the visual range. With the increasing demand for high-quality hyperspectral data, recovering spectral channels out of the original range is also an interesting work and a further direction of spectral super-resolution. In 2019, Lore et al. [55] selected some spectral bands in the hyperspectral image as input, and the missing bands were filled with interpolated values between the two neighboring available channels. They found that as the number of discarded bands increases, the performance of the model improves. In the same year, Gewali [53] found that increasing the number of input bands from 4 to 8 significantly reduces the reconstruction error in remote sensing imagery. They also believed that a larger number of input bands can capture more information about the target spectra. Thus, with a few available channels, recovering high-resolution hyperspectral images with numerous channels even covering a wider spectrum, such as infrared, ultraviolet, short wave, microwave, is extremely urgent.

Spectral imager. The spectral super-resolution methods reviewed in this survey are all algorithms addressing the captured images. No matter how good the performance is, there still is spectral information missing during computational imaging. Thus, some works tried to achieve finer-resolution spectral imaging in terms of the hardware, as listed in Table 7. Zhang et al. [177] developed a deeply learned broadband encoding stochastic hyperspectral camera and improved the spectral resolution up to 1 nm. Mu et al. [178] integrated spectrally-modulated polarimetry into the *Optically Replicating and Remapping Imaging Spectrometer* (ORRIS). Ji et al. [179] proposed a learning-powered snapshot hyperspectral imaging and used it to measure hemodynamic parameters with a mobile phone. Yako et al. [180] developed a video-rate HS camera with an array of 64 *Complementary Metal–Oxide–Semiconductor* (COMS)-compatible Fabry–Pérot filters. Wang et al. [181] designed an on-chip spectrometer using arrays of photodetectors. Xiong et al. [182] presented a silicon real-time ultraspectral imaging chip based on reconfigurable metasurfaces and used it in imaging brain hemodynamics. Finer resolution in spectral imager brings fast imaging speed and accurate spectral information with only one-time production cost, which is more beneficial to industrial applications. Thus, spectral super-resolution in hardware, including the finer-resolution spectral imagers and even the ultraspectral imagers, is also a work direction with bright future.

[2] The shown multi-focus image pair comes from *MEF* data set [173], and the multi-exposure image pair comes from *Lytro* data set [174]. The others are downloaded from the website: https://image.baidu.com.

Table 7
Some works about deploying finer-resolution spectral imaging on the spectral imager.

Works	Introduction	Range (nm)	Resolution (nm)
Zhang et al. [177]	a deeply learned broadband encoding stochastic hyperspectral camera is developed.	400–700	1
Mu et al. [178]	spectrally-modulated polarimetry is integrated into the ORRIS.	450–750	≈6
Ji et al. [179]	learning-powered snapshot hyperspectral imaging is introduced to measure hemodynamic parameters with a mobile phone.	380–720	0.5
Yako et al. [180]	a video-rate HS camera is developed with an array of 64 CMOS-compatible Fabry–Pérot filters.	450–650	10
Wang et al. [181]	an on-chip spectrometer is designed using arrays of photodetectors.	550–750	1
Xiong et al. [182]	a silicon real-time ultraspectral imaging chip based on reconfigurable metasurfaces is proposed and then used in imaging brain hemodynamics.	450–750	0.8

4.3. Multi-task learning

Low level & Low level. Spectral super-resolution, as a low-level image processing, aims to enhance the quality of the acquired images through increasing spectral channels. Although the reconstructed images represent the radiation characteristics of the target better, they still suffer from other quality degradations, as described in Section 4.2. Low quality and missing information seriously affect the wide deployment of the reconstructed images in subsequent applications. With the ever-growing requirements for high-quality hyperspectral data, combining multiple low-level image processing tasks to solve multiple quality degradation problems simultaneously would be a new direction of spectral super-resolution in the future. The low-level tasks that can be solved jointly include denoising [60], cloud removal, dehazing, inpainting, spatial super-resolution [62], spatial–spectral fusion [81], multi-focus fusion, and multi-exposure fusion.

Low level & High level. As a computer vision task, spectral super-resolution is only used as an assistive technique to generate higher-quality hyperspectral data. Acquiring high-quality data is only the first step in perceiving the world from image. Extracting semantic information and combining human priors to interpret images into knowledge that people can easily understand would satisfy human needs better. Common techniques include semantic segmentation, classification, object detection and extraction, and tracking. In some early works [48, 54,68,80], researchers have proven that results in high-level computer vision tasks, such as class, category, or clustering information, can improve the spectral super-resolution performance. On the contrary, better spectral information leads to better segmentation, classification, or detection. Combining spectral super-resolution with different high-level tasks is performed not only to obtain data and class results simultaneously, but also to share the feature information extracted from different-level tasks, which could benefit each other.

Data reconstruction & Parameter inversion. Hyperspectral images, as a type of data merging spatial geometric information with spectral radiation characteristic, have been widely used in estimating biological or environmental parameters. Schlerf et al. [183] employed a physical model to estimate structural canopy variables from hyperspectral remote sensing data. Wei et al. [184] used a combined spectral index model to retrieve the soil organic matter content from hyperspectral data. Wang et al. [185] discussed the feasibility and challenges in obtaining heavy metal content in soil and vegetation using hyperspectral sensing. Zarco-Tejada et al. [186] calculated chlorophyll content in closed forest canopies with hyperspectral data. Edelman et al. [187] estimated the age of blood stains using hyperspectral imaging. Combining spectral super-resolution with parameter inversion reduces the acquisition cost and meanwhile improves the accuracy of parameter estimation. Some works have focused on this problem. For example, Park [188] improved the spectral resolution of RGB inner eyelid images and estimated the blood hemoglobin. With the continuous popularization and rapid development of smart phones, people can easily obtain high-quality RGB images. The algorithm that combines spectral super-resolution with parameter inversion can help people obtain real-time information closely related to human life and health, such as biochemical parameters, food freshness, pesticide residue, and air quality.

5. Conclusions

Deep learning-based spectral super-resolution works as human's third eye and helps people better perceive the world from a new perspective with spectral radiation information. This study reviews the development of deep leaning-based spectral super-resolution algorithms from the network architecture, feature extraction, and physical modeling. We have collected almost all spectral super-resolution algorithms based on deep learning. To the best of our knowledge, studies that comprehensively review spectral super-resolution algorithms are limited. In addition, the normal application of spectral recovery, we have discussed the potential of deep learning-based sSR algorithms in image colorization and spectral compressive imaging. Moreover, a comparable benchmark involving three applications is proposed. A toolbox including unified data sets and training codes is provided in https://github.com/JiangHe96/DL4sSR. Researchers are only required to upload their models and easily obtain their results. At the end of this review, we also present our views on the trends and challenges of deep learning-based spectral super-resolution. For example, improvements in model generalization and lightweight model are promising works. Another potential direction is combining multiple tasks, including low-level tasks, high-level tasks, or parameter inversion. Considering multiple degradation into algorithm modeling may be a good solution to increase robustness. Furthermore, we would make more comprehensive reviews about image colorization and spectral compressive imaging based on deep learning and discuss about their difference from spectral super-resolution.

CRediT authorship contribution statement

Jiang He: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Preparation. **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. **Yi Xiao:** Methodology, Resources, Data curation, Writing – original draft, Writing – review & editing. **Denghong Liu:** Methodology, Writing – original draft, Writing – review & editing. **Huanfeng Shen:** Conceptualization, Supervision, Project administration. **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.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61971319, Grant 42230108, and Grant 62071341.

References

[1] H. Pu, L. Lin, D.-W. Sun, Principles of hyperspectral microscope imaging techniques and their applications in food quality and safety detection: a review, Compr. Rev. Food Sci. Food Saf. 18 (4) (2019) 853–866.
[2] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, Hyperspectral imaging–an emerging process analytical tool for food quality and safety control, Trends Food Sci. Technol. 18 (12) (2007) 590–598.
[3] C. Wang, B. Liu, L. Liu, Y. Zhu, J. Hou, P. Liu, X. Li, A review of deep learning used in the hyperspectral image analysis for agriculture, Artif. Intell. Rev. 54 (7) (2021) 5205–5253.
[4] P.S. Thenkabail, R.B. Smith, E. De Pauw, Hyperspectral vegetation indices and their relationships with agricultural crop characteristics, Remote Sens. Environ. 71 (2) (2000) 158–182.
[5] Y. Nan, J. Del Ser, S. Walsh, C. Schönlieb, M. Roberts, I. Selby, K. Howard, J. Owen, J. Neville, J. Guiot, et al., Data harmonisation for information fusion in digital healthcare: A state-of-the-art systematic review, meta-analysis and future research directions, Inf. Fusion (2022).
[6] E.A. Cloutis, Review article hyperspectral geological remote sensing: evaluation of analytical techniques, Int. J. Remote Sens. 17 (12) (1996) 2215–2242.
[7] F.D. Van der Meer, H.M. Van der Werff, F.J. Van Ruitenbeek, C.A. Hecker, W.H. Bakker, M.F. Noomen, M. Van Der Meijde, J.B. Carranza, T. Woldai, Multi-and hyperspectral geologic remote sensing: A review, Int. J. Appl. Earth Obs. Geoinf. 14 (1) (2012) 112–128.
[8] Y. Wang, Q. Yuan, T. Li, L. Zhu, L. Zhang, Estimating daily full-coverage near surface O3, CO, and NO2 concentrations at a high spatial resolution over China based on S5P-TROPOMI and GEOS-FP, ISPRS J. Photogramm. Remote Sens. 175 (2021) 311–325.
[9] Y. Wang, Q. Yuan, T. Li, L. Zhu, Global spatiotemporal estimation of daily high-resolution surface carbon monoxide concentrations using Deep forest, J. Clean. Prod. 350 (2022) 131500.
[10] Y. Wang, Q. Yuan, S. Zhou, L. Zhang, Global spatiotemporal completion of daily high-resolution TCCO from TROPOMI over land using a swath-based local ensemble learning method, ISPRS J. Photogramm. Remote Sens. 194 (2022) 167–180.
[11] S. Salcedo-Sanz, P. Ghamisi, M. Piles, M. Werner, L. Cuadra, A. Moreno-Martínez, E. Izquierdo-Verdiguier, J. Muñoz-Marí, A. Mosavi, G. Camps-Valls, Machine learning information fusion in earth observation: A comprehensive review of methods, applications and data sources, Inf. Fusion 63 (2020) 256–272.
[12] G. Lu, B. Fei, Medical hyperspectral imaging: a review, J. Biomed. Opt. 19 (1) (2014) 010901.
[13] L. Khaodhiar, T. Dinh, K.T. Schomacker, S.V. Panasyuk, J.E. Freeman, R. Lew, T. Vo, A.A. Panasyuk, C. Lima, J.M. Giurini, et al., The use of medical hyperspectral technology to evaluate microcirculatory changes in diabetic foot ulcers and to predict clinical outcomes, Diabetes Care 30 (4) (2007) 903–910.
[14] J.M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, J. Chanussot, Hyperspectral remote sensing data analysis and future challenges, IEEE Geosci. Remote Sens. Mag. 1 (2) (2013) 6–36.
[15] R. Dian, S. Li, B. Sun, A. Guo, Recent advances and new guidelines on hyperspectral and multispectral image fusion, Inf. Fusion 69 (2021) 40–51.
[16] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, L. Zhang, Satellite video super-resolution via multiscale deformable convolution alignment and temporal grouping projection, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–19.
[17] J. Wang, C. Tang, Z. Li, X. Liu, W. Zhang, E. Zhu, L. Wang, Hyperspectral band selection via region-aware latent features fusion based clustering, Inf. Fusion 79 (2022) 162–173.
[18] K. Jiang, Z. Wang, P. Yi, T. Lu, J. Jiang, Z. Xiong, Dual-path deep fusion network for face image hallucination, IEEE Trans. Neural Netw. Learn. Syst. 33 (1) (2020) 378–391.
[19] X. Zhang, H. Zhao, Hyperspectral-cube-based mobile face recognition: A comprehensive review, Inf. Fusion 74 (2021) 132–150.
[20] B. Uzkent, M.J. Hoffman, A. Vodacek, Real-time vehicle tracking in aerial video using hyperspectral features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 36–44.
[21] D. Manolakis, G. Shaw, Detection algorithms for hyperspectral imaging applications, IEEE Signal Process. Mag. 19 (1) (2002) 29–43.
[22] V. Ferraris, N. Dobigeon, M. Chabert, Robust fusion algorithms for unsupervised change detection between multi-band optical images-a comprehensive case study, Inf. Fusion 64 (2020) 293–317.
[23] S. Wang, Y. Tang, X. Liao, J. He, H. Feng, H. Jiao, X. Su, Q. Yuan, An ensemble learning approach with multi-depth attention mechanism for road damage detection, in: 2022 IEEE International Conference on Big Data, Big Data, IEEE, 2022, pp. 6439–6444.
[24] R. Dian, S. Li, L. Fang, Learning a low tensor-train rank representation for hyperspectral image super-resolution, IEEE Trans. Neural Netw. Learn. Syst. 30 (9) (2019) 2672–2683.
[25] R. Dian, S. Li, L. Fang, Q. Wei, Multispectral and hyperspectral image fusion with spatial-spectral sparse representation, Inf. Fusion 49 (2019) 262–270.
[26] Q. Xie, M. Zhou, Q. Zhao, Z. Xu, D. Meng, MHF-Net: An interpretable deep network for multispectral and hyperspectral image fusion, IEEE Trans. Pattern Anal. Mach. Intell. (2020).
[27] L. Zhang, J. He, Q. Yang, Y. Xiao, Q. Yuan, Data-driven multi-source remote sensing data fusion: Progress and challenges, Acta Geod. Cartogr. Sinica 51 (7) (2022) 1317–1337.
[28] L.-J. Deng, G. Vivone, M.E. Paoletti, G. Scarpa, J. He, Y. Zhang, J. Chanussot, A. Plaza, Machine learning in pansharpening: A benchmark, from shallow to deep networks, IEEE Geosci. Remote Sens. Mag. (2022) 2–38.
[29] J. He, J. Li, Q. Yuan, H. Li, H. Shen, Spatial-spectral fusion in different swath widths by a recurrent expanding residual convolutional neural network, Remote Sens. 11 (19) (2019) 2203.
[30] C.-H. Yeh, C.-H. Lin, M.-H. Lin, L.-W. Kang, C.-H. Huang, M.-J. Chen, Deep learning-based compressed image artifacts reduction based on multi-scale image fusion, Inf. Fusion 67 (2021) 195–207.
[31] T. Okamoto, I. Yamaguchi, Simultaneous acquisition of spectral image information, Opt. Lett. 16 (16) (1991) 1277–1279.
[32] F. Agahian, S.A. Amirshahi, S.H. Amirshahi, Reconstruction of reflectance spectra using weighted principal component analysis, Color Res. Appl. 33 (5) (2008) 360–371.
[33] N. Eslahi, S.H. Amirshahi, F. Agahian, Recovery of spectral data using weighted canonical correlation regression, Opt. Rev. 16 (3) (2009) 296–303.
[34] M. Parmar, S. Lansel, B.A. Wandell, Spatio-spectral reconstruction of the multispectral datacube using sparse recovery, in: 2008 15th IEEE International Conference on Image Processing, IEEE, 2008, pp. 473–476.
[35] R.M. Nguyen, D.K. Prasad, M.S. Brown, Training-based spectral reconstruction from a single RGB image, in: European Conference on Computer Vision, Springer, 2014, pp. 186–201.
[36] A. Robles-Kelly, Single image spectral reconstruction for multimedia applications, in: Proceedings of the 23rd ACM International Conference on Multimedia, 2015, pp. 251–260.
[37] B. Arad, O. Ben-Shahar, Sparse recovery of hyperspectral signal from natural RGB images, in: European Conference on Computer Vision, Springer, 2016, pp. 19–34.
[38] 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: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4705–4713.
[39] J. Aeschbacher, J. Wu, R. Timofte, In defense of shallow learned spectral reconstruction from RGB images, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 471–479.
[40] N. Akhtar, A. Mian, Hyperspectral recovery from RGB images using gaussian processes, IEEE Trans. Pattern Anal. Mach. Intell. 42 (1) (2018) 100–113.
[41] A. Alvarez-Gila, J. Van De Weijer, E. Garrote, Adversarial networks for spatial context-aware spectral image reconstruction from RGB, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 480–490.
[42] S. Galliani, C. Lanaras, D. Marmanis, E. Baltsavias, K. Schindler, Learned spectral super-resolution, 2017, arXiv preprint arXiv:1703.09470.
[43] A. Rangnekar, N. Mokashi, E. Ientilucci, C. Kanan, M. Hoffman, Aerial spectral super-resolution using conditional adversarial networks, 2017, arXiv preprint arXiv:1712.08690.
[44] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, F. Wu, Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 518–525.
[45] Y.B. Can, R. Timofte, An efficient CNN for spectral reconstruction from RGB images, 2018, arXiv preprint arXiv:1804.04647.
[46] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, H. Huang, Joint camera spectral sensitivity selection and hyperspectral image recovery, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 788–804.

[47] X.-H. Han, B. Shi, Y. Zheng, Residual hsrcnn: Residual hyper-spectral reconstruction cnn from an RGB image, in: 2018 24th International Conference on Pattern Recognition, ICPR, IEEE, 2018, pp. 2664–2669.

[48] X. Han, J. Yu, J.-H. Xue, W. Sun, Spectral super-resolution for RGB images using class-based BP neural networks, in: 2018 Digital Image Computing: Techniques and Applications, DICTA, IEEE, 2018, pp. 1–7.

[49] 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: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 844–851.

[50] Z. Shi, C. Chen, Z. Xiong, D. Liu, F. Wu, HSCNN+: Advanced cnn-based hyperspectral recovery from RGB images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 939–947.

[51] T. Stiebel, S. Koppers, P. Seltsam, D. Merhof, Reconstructing spectral images from RGB-images using a convolutional neural network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 948–953.

[52] Y. Yan, L. Zhang, J. Li, W. Wei, Y. Zhang, Accurate spectral super-resolution from single RGB image using multi-scale CNN, in: Chinese Conference on Pattern Recognition and Computer Vision, PRCV, Springer, 2018, pp. 206–217.

[53] U.B. Gewali, S.T. Monteiro, E. Saber, Spectral super-resolution with optimized bands, Remote Sens. 11 (14) (2019) 1648.

[54] 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, pp. 3546–3555.

[55] K.G. Lore, K.K. Reddy, M. Giering, E.A. Bernal, Generative adversarial networks for spectral super-resolution and bidirectional RGB-to-multispectral mapping, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, IEEE, 2019, pp. 926–933.

[56] L. Wang, C. Sun, Y. Fu, M.H. Kim, H. Huang, Hyperspectral image reconstruction using a deep spatial-spectral prior, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 8032–8041.

[57] A. Banerjee, A. Palrecha, Mxr-u-nets for real time hyperspectral reconstruction, 2020, arXiv preprint arXiv:2004.07003.

[58] B.J. Fubara, M. Sedky, D. Dyke, RGB to spectral reconstruction via learned basis functions and weights, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 480–481.

[59] 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. 59 (3) (2020) 2321–2335.

[60] J. Li, C. Wu, R. Song, Y. Li, F. Liu, Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 462–463.

[61] E. Martínez, S. Castro, J. Bacca, H. Arguello, Efficient transfer learning for spectral image reconstruction from RGB images, in: 2020 IEEE Colombian Conference on Applications of Computational Intelligence, IEEE ColCACI 2020, IEEE, 2020, pp. 1–6.

[62] 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.

[63] D.S. Nathan, K. Uma, D.S. Vinothini, B.S. Bama, S. Roomi, Light weight residual dense attention net for spectral reconstruction from RGB images, 2020, arXiv preprint arXiv:2004.06930.

[64] H. Peng, X. Chen, J. Zhao, Residual pixel attention network for spectral reconstruction from RGB images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 486–487.

[65] T. Stiebel, P. Seltsam, D. Merhof, Enhancing deep spectral super-resolution from RGB images by enforcing the metameric constraint, in: VISIGRAPP, Vol. 4, VISAPP, 2020, pp. 57–66.

[66] L. Wang, C. Sun, M. Zhang, Y. Fu, H. Huang, Dnu: Deep non-local unrolling for computational spectral imaging, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1661–1671.

[67] W. Wei, Y. Sun, L. Zhang, J. Nie, Y. Zhang, Boosting one-shot spectral super-resolution using transfer learning, IEEE Trans. Comput. Imaging 6 (2020) 1459–1470.

[68] L. Yan, X. Wang, M. Zhao, M. Kaloorazi, J. Chen, S. Rahardja, Reconstruction of hyperspectral data from RGB images with prior category information, IEEE Trans. Comput. Imaging 6 (2020) 1070–1081.

[69] L. Zhang, Z. Lang, P. Wang, W. Wei, S. Liao, L. Shao, Y. Zhang, Pixel-aware deep function-mixture network for spectral super-resolution, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 07, 2020, pp. 12821–12828.

[70] Y. Zhao, L.-M. Po, Q. Yan, W. Liu, T. Lin, Hierarchical regression network for spectral reconstruction from RGB images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 422–423.

[71] 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).

[72] R. Hang, Q. Liu, Z. Li, Spectral super-resolution network guided by intrinsic properties of hyperspectral imagery, IEEE Trans. Image Process. 30 (2021) 7256–7265.

[73] J. Li, C. Wu, R. Song, Y. Li, W. Xie, L. He, X. Gao, Deep hybrid 2-D-3-D CNN based on dual second-order attention with camera spectral sensitivity prior for spectral super-resolution, IEEE Trans. Neural Netw. Learn. Syst. (2021).

[74] T. Li, Y. Gu, Progressive spatial–spectral joint network for hyperspectral image reconstruction, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–14.

[75] B. Sun, J. Yan, X. Zhou, Y. Zheng, Tuning IR-cut filter for illumination-aware spectral reconstruction from RGB, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 84–93.

[76] X. Zheng, W. Chen, X. Lu, Spectral super-resolution of multispectral images using spatial-spectral residual attention network, IEEE Trans. Geosci. Remote Sens. (2021).

[77] Z. Zhu, H. Liu, J. Hou, S. Jia, Q. Zhang, Deep amended gradient descent for efficient spectral reconstruction from single RGB images, IEEE Trans. Comput. Imaging 7 (2021) 1176–1188.

[78] Z. Zhu, H. Liu, J. Hou, H. Zeng, Q. Zhang, Semantic-embedded unsupervised spectral reconstruction from single RGB images in the wild, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 2279–2288.

[79] W. Chen, X. Zheng, X. Lu, Semisupervised spectral degradation constrained network for spectral super-resolution, IEEE Geosci. Remote Sens. Lett. 19 (2022) 1–5.

[80] X. Han, H. Zhang, J.-H. Xue, W. Sun, A spectral–spatial jointed spectral super-resolution and its application to HJ-1A satellite images, IEEE Geosci. Remote Sens. Lett. 19 (2022) 1–5.

[81] J. He, Q. Yuan, J. Li, L. Zhang, PoNet: A universal physical optimization-based spectral super-resolution network for arbitrary multispectral images, Inf. Fusion 80 (2022) 205–225.

[82] J. Li, S. Du, R. Song, C. Wu, Y. Li, Q. Du, HASIC-net: Hybrid attentional convolutional neural network with structure information consistency for spectral super-resolution of RGB images, IEEE Trans. Geosci. Remote Sens. (2022).

[83] Q. Ma, J. Jiang, X. Liu, J. Ma, Multi-task interaction learning for spatiospectral image super-resolution, IEEE Trans. Image Process. 31 (2022) 2950–2961.

[84] S. Mei, Y. Geng, J. Hou, Q. Du, Learning hyperspectral images from RGB images via a coarse-to-fine CNN, Sci. China Inf. Sci. 65 (5) (2022) 1–14.

[85] Q. Ma, J. Jiang, X. Liu, J. Ma, Deep unfolding network for spatiospectral image super-resolution, IEEE Trans. Comput. Imaging 8 (2022) 28–40.

[86] K. Jiang, Z. Wang, P. Yi, J. Jiang, Hierarchical dense recursive network for image super-resolution, Pattern Recognit. 107 (2020) 107475.

[87] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, J. Jiang, Edge-enhanced GAN for remote sensing image superresolution, IEEE Trans. Geosci. Remote Sens. 57 (8) (2019) 5799–5812.

[88] Y. Xiao, Y. Wang, Q. Yuan, J. He, L. Zhang, Generating a long-term (2003–2020) hourly 0.25∘ global PM2.5 dataset via spatiotemporal downscaling of CAMS with deep learning (DeepCAMS), Sci. Total Environ. 848 (2022) 157747.

[89] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, Psychol. Rev. 65 (6) (1958) 386.

[90] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536.

[91] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, L. Jackel, Handwritten digit recognition with a back-propagation network, Adv. Neural Inf. Process. Syst. 2 (1989).

[92] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[93] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[94] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).

[95] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[96] J. Zhang, R. Su, Q. Fu, W. Ren, F. Heide, Y. Nie, A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging, Sci. Rep. 12 (1) (2022) 1–17.

[97] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, M. Li, Bag of tricks for image classification with convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 558–567.

[98] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.

[99] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.

[100] S. Ji, W. Xu, M. Yang, K. Yu, 3D convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2012) 221–231.

[101] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[102] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[103] B. Arad, O. Ben-Shahar, Filter selection for hyperspectral estimation, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 3153–3161.

[104] Q. Yuan, Q. Zhang, J. Li, H. Shen, L. Zhang, Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network, IEEE Trans. Geosci. Remote Sens. 57 (2) (2018) 1205–1218.

[105] D. Liu, J. Li, Q. Yuan, A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution, IEEE Trans. Geosci. Remote Sens. 59 (9) (2021) 7711–7725.

[106] S. Li, B. Yang, J. Hu, Performance comparison of different multi-resolution transforms for image fusion, Inf. Fusion 12 (2) (2011) 74–84.

[107] J. Sun, H. Li, Z. Xu, et al., Deep ADMM-Net for compressive sensing MRI, Adv. Neural Inf. Process. Syst. 29 (2016).

[108] H. Shen, M. Jiang, J. Li, C. Zhou, Q. Yuan, L. Zhang, Coupling model- and data-driven methods for remote sensing image restoration and fusion: Improving physical interpretability, IEEE Geosci. Remote Sens. Mag. 10 (2) (2022) 231–249.

[109] K. Zhang, W. Zuo, S. Gu, L. Zhang, Learning deep CNN denoiser prior for image restoration, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3929–3938.

[110] V. Monga, Y. Li, Y.C. Eldar, Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing, IEEE Signal Process. Mag. 38 (2) (2021) 18–44.

[111] R. Liu, S. Cheng, L. Ma, X. Fan, Z. Luo, Deep proximal unrolling: Algorithmic framework, convergence analysis and applications, IEEE Trans. Image Process. 28 (10) (2019) 5013–5026.

[112] K. Wei, A. Aviles-Rivero, J. Liang, Y. Fu, C.-B. Schönlieb, H. Huang, Tuning-free plug-and-play proximal algorithm for inverse imaging problems, in: International Conference on Machine Learning, PMLR, 2020, pp. 10158–10169.

[113] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, X. Lu, Denoising prior driven deep neural network for image restoration, IEEE Trans. Pattern Anal. Mach. Intell. 41 (10) (2018) 2305–2318.

[114] X. Xie, J. Wu, G. Liu, Z. Zhong, Z. Lin, Differentiable linearized ADMM, in: International Conference on Machine Learning, PMLR, 2019, pp. 6902–6911.

[115] J. He, Q. Yuan, J. Li, L. Zhang, A knowledge optimization-driven network with normalizer-free group ResNet prior for remote sensing image pan-sharpening, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–16.

[116] D. Liu, J. Li, Q. Yuan, L. Zheng, J. He, S. Zhao, Y. Xiao, An efficient unfolding network with disentangled spatial-spectral representation for hyperspectral image super-resolution, Inf. Fusion 94 (2023) 92–111.

[117] J.C. White, N. Saarinen, V. Kankare, M.A. Wulder, T. Hermosilla, N.C. Coops, P.D. Pickell, M. Holopainen, J. Hyyppä, M. Vastaranta, Confirmation of post-harvest spectral recovery from landsat time series using measures of forest cover and height derived from airborne laser scanning data, Remote Sens. Environ. 216 (2018) 262–275.

[118] B. Arad, O. Ben-Shahar, R.N. Timofte, L. Van Gool, L. Zhang, M.N. Yang, Challenge on spectral reconstruction from RGB images, in: Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 2018, pp. 18–22.

[119] B. Arad, R. Timofte, O. Ben-Shahar, Y.-T. Lin, G.D. Finlayson, Ntire 2020 challenge on spectral reconstruction from an RGB image, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 446–447.

[120] B. Arad, R. Timofte, R. Yahel, N. Morag, A. Bernat, Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, et al., NTIRE 2022 spectral recovery challenge and data set, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 863–881.

[121] R. Zhang, P. Isola, A.A. Efros, Colorful image colorization, in: European Conference on Computer Vision, Springer, 2016, pp. 649–666.

[122] Q. Luan, F. Wen, D. Cohen-Or, L. Liang, Y.-Q. Xu, H.-Y. Shum, Natural image colorization, in: Proceedings of the 18th Eurographics Conference on Rendering Techniques, 2007, pp. 309–320.

[123] A. Deshpande, J. Lu, M.-C. Yeh, M. Jin Chong, D. Forsyth, Learning diverse image colorization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6837–6845.

[124] K. Nazeri, E. Ng, M. Ebrahimi, Image colorization using generative adversarial networks, in: International Conference on Articulated Motion and Deformable Objects, Springer, 2018, pp. 85–94.

[125] S. Huang, X. Jin, Q. Jiang, L. Liu, Deep learning for image colorization: Current and future prospects, Eng. Appl. Artif. Intell. 114 (2022) 105006.

[126] D. Wu, J. Gan, J. Zhou, J. Wang, W. Gao, Fine-grained semantic ethnic costume high-resolution image colorization with conditional GAN, Int. J. Intell. Syst. 37 (5) (2022) 2952–2968.

[127] X. Pan, X. Zhan, B. Dai, D. Lin, C.C. Loy, P. Luo, Exploiting deep generative prior for versatile image restoration and manipulation, IEEE Trans. Pattern Anal. Mach. Intell. (2021).

[128] G. Larsson, M. Maire, G. Shakhnarovich, Learning representations for automatic colorization, in: European Conference on Computer Vision, Springer, 2016, pp. 577–593.

[129] J. Gu, Y. Shen, B. Zhou, Image processing using multi-code gan prior, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3012–3021.

[130] Y. Cai, J. Lin, X. Hu, H. Wang, X. Yuan, Y. Zhang, R. Timofte, L. Van Gool, Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 17502–17511.

[131] M.E. Gehm, R. John, D.J. Brady, R.M. Willett, T.J. Schulz, Single-shot compressive spectral imaging with a dual-disperser architecture, Opt. Express 15 (21) (2007) 14013–14027.

[132] G.R. Arce, D.J. Brady, L. Carin, H. Arguello, D.S. Kittle, Compressive coded aperture spectral imaging: An introduction, IEEE Signal Process. Mag. 31 (1) (2013) 105–115.

[133] Z. Meng, J. Ma, X. Yuan, End-to-end low cost compressive spectral imaging with spatial-spectral self-attention, in: European Conference on Computer Vision, Springer, 2020, pp. 187–204.

[134] I. Choi, M. Kim, D. Gutierrez, D. Jeon, G. Nam, High-Quality Hyperspectral Reconstruction Using a Spectral Prior, Technical Report, 2017.

[135] J.F. Florez-Ospina, A.K. Alrushud, D.L. Lau, G.R. Arce, Block-based spectral image reconstruction for compressive spectral imaging using smoothness on graphs, Opt. Express 30 (5) (2022) 7187–7209.

[136] T. Huang, W. Dong, X. Yuan, J. Wu, G. Shi, Deep gaussian scale mixture prior for spectral compressive imaging, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 16216–16225.

[137] Z. Cheng, B. Chen, R. Lu, Z. Wang, H. Zhang, Z. Meng, X. Yuan, Recurrent neural networks for snapshot compressive imaging, IEEE Trans. Pattern Anal. Mach. Intell. (2022).

[138] H. Rueda, D. Lau, G.R. Arce, Multi-spectral compressive snapshot imaging using RGB image sensors, Opt. Express 23 (9) (2015) 12207–12221.

[139] H. Yuan, P. Zhang, F. Gao, Compressive hyperspectral Raman imaging via randomly interleaved scattering projection, Optica 8 (11) (2021) 1462–1470.

[140] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[141] F.A. 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.

[142] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.

[143] I. Loshchilov, F. Hutter, Sgdr: Stochastic gradient descent with warm restarts, 2016, arXiv preprint arXiv:1608.03983.

[144] G. Patterson, J. Hays, Sun attribute database: Discovering, annotating, and recognizing scene attributes, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 2751–2758.

[145] S. Iizuka, E. Simo-Serra, H. Ishikawa, Let there be color! joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification, ACM Trans. Graph. 35 (4) (2016) 1–11.

[146] F. Yasuma, T. Mitsunaga, D. Iso, S.K. Nayar, Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum, IEEE Trans. Image Process. 19 (9) (2010) 2241–2253.

[147] Y. Xiao, Q. Yuan, K. Jiang, J. He, Y. Wang, L. Zhang, From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution, Inf. Fusion (2023).

[148] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[149] X. Jin, J. He, Y. Xiao, Q. Yuan, Learning a local-global alignment network for satellite video super-resolution, IEEE Geosci. Remote Sens. Lett. (2023) 1.

[150] X. He, Y. Chen, Optimized input for CNN-based hyperspectral image classification using spatial transformer network, IEEE Geosci. Remote Sens. Lett. 16 (12) (2019) 1884–1888.

[151] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, J. Chanussot, SpectralFormer: Rethinking hyperspectral image classification with transformers, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–15.

[152] L. Sun, G. Zhao, Y. Zheng, Z. Wu, Spectral–spatial feature tokenization transformer for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–14.

[153] Y. Xiao, Q. Yuan, J. He, Q. Zhang, J. Sun, X. Su, J. Wu, L. Zhang, Space-time super-resolution for satellite video: A joint framework based on multi-scale spatial-temporal transformer, Int. J. Appl. Earth Obs. Geoinf. 108 (2022) 102731.

[154] J. He, Q. Yuan, J. Li, Y. Xiao, X. Liu, Y. Zou, DsTer: A dense spectral transformer for remote sensing spectral super-resolution, Int. J. Appl. Earth Obs. Geoinf. 109 (2022) 102773.

[155] Y. Cai, J. Lin, H. Wang, X. Yuan, H. Ding, Y. Zhang, R. Timofte, L. Van Gool, Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging, 2022, arXiv preprint arXiv:2205.10102.

[156] Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, H. Pfister, R. Timofte, L. Van Gool, Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2022, pp. 745–755.

[157] M. Danelljan, G. Bhat, F. Shahbaz Khan, M. Felsberg, Eco: Efficient convolution operators for tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6638–6646.

[158] S. Gong, L. Chen, M. Bronstein, S. Zafeiriou, Spiralnet++: A fast and highly efficient mesh convolution operator, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.

[159] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, Ghostnet: More features from cheap operations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1580–1589.

[160] E.J. Crowley, G. Gray, A.J. Storkey, Moonshine: Distilling with cheap convolutions, Adv. Neural Inf. Process. Syst. 31 (2018).

[161] G. Li, X. Ma, X. Wang, H. Yue, J. Li, L. Liu, X. Feng, J. Xue, Optimizing deep neural networks on intelligent edge accelerators via flexible-rate filter pruning, J. Syst. Archit. 124 (2022) 102431.

[162] Z. Liu, J. Li, Z. Shen, G. Huang, S. Yan, C. Zhang, Learning efficient convolutional networks through network slimming, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2736–2744.

[163] Y. Chen, X. Dai, D. Chen, M. Liu, X. Dong, L. Yuan, Z. Liu, Mobile-former: Bridging mobilenet and transformer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5270–5279.

[164] D. Blalock, J.J. Gonzalez Ortiz, J. Frankle, J. Guttag, What is the state of neural network pruning? Proc. Mach. Learn. Syst. 2 (2020) 129–146.

[165] J. Gou, B. Yu, S.J. Maybank, D. Tao, Knowledge distillation: A survey, Int. J. Comput. Vis. 129 (6) (2021) 1789–1819.

[166] T. Zhang, Y. Fu, C. Li, Hyperspectral image denoising with realistic data, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 2248–2257.

[167] H. Wang, Q. Xie, Q. Zhao, D. Meng, A model-driven deep neural network for single image rain removal, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.

[168] X. Ma, Q. Wang, X. Tong, A spectral grouping-based deep learning model for haze removal of hyperspectral images, ISPRS J. Photogramm. Remote Sens. 188 (2022) 177–189.

[169] Y. Liu, L. Wang, H. Li, X. Chen, Multi-focus image fusion with deep residual learning and focus property detection, Inf. Fusion (2022).

[170] H. Zhang, Z. Le, Z. Shao, H. Xu, J. Ma, MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion, Inf. Fusion 66 (2021) 40–53.

[171] D. Han, L. Li, X. Guo, J. Ma, Multi-exposure image fusion via deep perceptual enhancement, Inf. Fusion 79 (2022) 248–262.

[172] Y. Fu, T. Zhang, L. Wang, H. Huang, Coded hyperspectral image reconstruction using deep external and internal learning, IEEE Trans. Pattern Anal. Mach. Intell. (2021).

[173] M. Nejati, S. Samavi, S. Shirani, Multi-focus image fusion using dictionary-based sparse representation, Inf. Fusion 25 (2015) 72–84.

[174] J. Cai, S. Gu, L. Zhang, Learning a deep single image contrast enhancer from multi-exposure images, IEEE Trans. Image Process. 27 (4) (2018) 2049–2062.

[175] J.E. Van Engelen, H.H. Hoos, A survey on semi-supervised learning, Mach. Learn. 109 (2) (2020) 373–440.

[176] W. Kim, A. Kanezaki, M. Tanaka, Unsupervised learning of image segmentation based on differentiable feature clustering, IEEE Trans. Image Process. 29 (2020) 8055–8068.

[177] W. Zhang, H. Song, X. He, L. Huang, X. Zhang, J. Zheng, W. Shen, X. Hao, X. Liu, Deeply learned broadband encoding stochastic hyperspectral imaging, Light Sci. Appl. 10 (1) (2021) 108.

[178] T. Mu, F. Han, H. Li, A. Tuniyazi, Q. Li, H. Gong, W. Wang, R. Liang, Snapshot hyperspectral imaging polarimetry with full spectropolarimetric resolution, Opt. Lasers Eng. 148 (2022) 106767.

[179] Y. Ji, S.M. Park, S. Kwon, J.W. Leem, V.V. Nair, Y. Tong, Y.L. Kim, mHealth hyperspectral learning for instantaneous spatiospectral imaging of hemodynamics, 2023, arXiv preprint arXiv:2303.16205.

[180] M. Yako, Y. Yamaoka, T. Kiyohara, C. Hosokawa, A. Noda, K. Tack, N. Spooren, T. Hirasawa, A. Ishikawa, Video-rate hyperspectral camera based on a CMOS-compatible random array of Fabry–Pérot filters, Nat. Photonics (2023) 1–6.

[181] Z. Wang, S. Yi, A. Chen, M. Zhou, T.S. Luk, A. James, J. Nogan, W. Ross, G. Joe, A. Shahsafi, et al., Single-shot on-chip spectral sensors based on photonic crystal slabs, Nature Commun. 10 (1) (2019) 1020.

[182] J. Xiong, X. Cai, K. Cui, Y. Huang, J. Yang, H. Zhu, W. Li, B. Hong, S. Rao, Z. Zheng, et al., Dynamic brain spectrum acquired by a real-time ultraspectral imaging chip with reconfigurable metasurfaces, Optica 9 (5) (2022) 461–468.

[183] M. Schlerf, C. Atzberger, Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data, Remote Sens. Environ. 100 (3) (2006) 281–294.

[184] L. Wei, Z. Yuan, Z. Wang, L. Zhao, Y. Zhang, X. Lu, L. Cao, Hyperspectral inversion of soil organic matter content based on a combined spectral index model, Sensors 20 (10) (2020) 2777.

[185] F. Wang, J. Gao, Y. Zha, Hyperspectral sensing of heavy metals in soil and vegetation: Feasibility and challenges, ISPRS J. Photogramm. Remote Sens. 136 (2018) 73–84.

[186] P.J. Zarco-Tejada, J.R. Miller, T.L. Noland, G.H. Mohammed, P.H. Sampson, Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data, IEEE Trans. Geosci. Remote Sens. 39 (7) (2001) 1491–1507.

[187] G. Edelman, T.G. Van Leeuwen, M.C. Aalders, Hyperspectral imaging for the age estimation of blood stains at the crime scene, Forensic Sci. Int. 223 (1–3) (2012) 72–77.

[188] S.M. Park, M.A. Visbal-Onufrak, M.M. Haque, M.C. Were, V. Naanyu, M.K. Hasan, Y.L. Kim, mHealth spectroscopy of blood hemoglobin with spectral super-resolution, Optica 7 (6) (2020) 563–573.