Spectrum-Aware and Transferable Architecture Search for Hyperspectral Image Restoration

Wei He[1⋆], Quanming Yao[2⋆], Naoto Yokoya[3 4], Tatsumi Uezato[5], Hongyan Zhang[1§], and Liangpei Zhang[1]

[1] LIESMARS, Wuhan University, Wuhan, China
[2] Department of Electronic Engineering, Tsinghua University, Beijing, China
[3] The University of Tokyo, Tokyo, Japan
[4] RIKEN AIP, Tokyo, Japan
[5] Hitachi, Ltd, Tokyo, Japan
{weihe1990,zhanghongyan,zlp62}@whu.edu.cn qyaoaa@tsinghua.edu.cn
yokoya@k.u-tokyo.ac.jp tatsumi.uezato.ay@hitachi.com

Abstract. Convolutional neural networks have been widely developed for hyperspectral image (HSI) restoration. However, making full use of the spatial-spectral information of HSIs still remains a challenge. In this work, we disentangle the 3D convolution into lightweight 2D spatial and spectral convolutions, and build a spectrum-aware search space for HSI restoration. Subsequently, we utilize neural architecture search strategy to automatically learn the most efficient architecture with proper convolutions and connections in order to fully exploit the spatial-spectral information. We also determine that the super-net with global and local skip connections can further boost HSI restoration performance. The proposed STAS is optimized in a noise independent mode to increase transferability. The searched architecture on the CAVE dataset has been adopted for various reconstruction tasks, and achieves remarkable performance. On the basis of fruitful experiments, we conclude that the transferability of searched architecture is dependent on the spectral information and independent of the noise levels.

Keywords: hyperspectral restoration, spatial-spectral, convolutional neural networks, denoising

1 Introduction

As three-dimensional (3D) cubes, hyperspectral images (HSIs) have the ability to distinguish precise details between similar materials [21], and therefore have been widely utilized in various applications [15, 36, 19, 45, 32]. However, due to the sensitive imaging process and complex external imaging environment, HSIs tend to suffer from various degradations, *i.e.*, noise [4, 13, 5, 16], missing [44, 20], and undersampling [28, 39]. Thus, HSI restoration is an important preprocessing step, essential to improving image quality and the subsequent applications.

Unlike RGB, *the main challenge of HSI processing is the exploration of high spectral information*. To that end, low-rank regularization has been successfully

⋆ indicates equal contribution, § indicates corresponding author.

utilized to explore the spectral correlation and achieves remarkable restoration results [50, 33, 7, 53]. The further spatial-spectral regularizations [41, 11, 5, 28, 20] boost restoration performance to the state-of-the-art (SOTA). However, the complexity of these methods is unacceptable for real-time processing [43, 28].

Recently, convolutional neural networks (CNNs) have been introduced for HSI restoration [48]. The CNN-based methods utilize numerous simulated degradation datasets to train the model, and then apply the trained model to denoise the HSIs. Generally, the training stage is time consuming, but the test stage is fast. Initially, separable 3DUnet [14] and recurrent architecture [42] were proposed to denoise the HSI. Following, attention module [31, 35, 30] and multi-scale architecture [48] were also embedded into the CNN architecture.

In contrast to traditional RGB restoration tasks where CNN models tend to be promising, training CNN models for HSI restoration is difficult due to the large number of spectral bands. First, it has been pointed out that simultaneously processing the whole spectrum can improve restoration accuracy [20]. However, because of the huge computation burden, previous networks [3, 14] fail to fully exploit the spatial-spectral information of HSIs. Second, the hand-crafted CNN models, such as 3DUnet [14] and TSA [30] are efficient for specific dataset, but may be suboptimal for different HSI dataset with spectra/noise diversity. In summary, the main challenge of CNN model for HSI restoration is the exploration of spatial-spectral information, including full-spectrum processing and the proper choice of convolution operations with proper connections.

The past few years have witnessed the development of automatic architecture design. Compared to hand-crafted architectures, the learned architectures have achieved excellent performance in various applications, *e.g.,* classification [56], objection detection [37], and super-resolution [25, 22]. Inspired by these works, we propose the Spectrum-aware and Transferable Architecture Search (STAS) to build a lightweight and efficient network for HSI restoration tasks. The proposed super-net STAS is composed of repeated super-cell modules with global and local skip connections. Inspired by the previous works [14, 24], we design a spectrum-aware search space that takes into consideration spatial convolution, spectral convolution, and spatial-spectral separable convolution. Specifically, *we utilize STAS to automatically choose the efficient convolutions with proper connections for the adaptive spatial-spectral information exploration of HSI.*

For different HSI restoration tasks, the noise type is different, *i.e.*, the Gaussian/stripe noise in denoising [14], and model errors in HSI imaging [31]. Without unified noise analysis framework, existing CNN-based methods [14, 30] are mostly designed for specific dataset and tasks. Therein, we introduce the noise-independent analysis for the proposed STAS, to increase the transferability of the searched architectures across different datasets and different tasks. The contributions are summarized as follows:

– We disentangle the 3D convolution into lightweight 2D spatial and spectral convolutions, and introduce a spectrum-aware search space. The proposed STAS, which includes local and global skip connections, is utilized to automatically learn an adaptive spatial-spectral network for efficient HSI restoration.

The searched architectures can be successfully applied to various dataset restoration tasks.
- We conclude that the transferability of searched architecture is dependent on the spectral information and independent of the noise levels. Specifically, for CAVE with fewer spectral bands, spatial convolution is effective. for Pavia with larger spectral bands, separable convolution is effective.

2 Related Work

CNN-based HSI restoration. For HSI restoration, *The main principles of deep network are the spatial-spectral exploration and transferability across tasks.* Unfortunately, the pioneering methods HSI-DeNet [3] and HSID-CNN [48] ignored the global spectral information. Following, Dong *et al.* proposed the separable 3D convolution with Unet architecture [14] to efficiently capture the spatial-spectral features of 3D HSI. Fu *et al.* introduced an efficient spectrum based recurrent architecture [42]. For HSI imaging reconstruction, the spatial-spectral attention [31, 30] and plug-and-play framework [54, 49] are also introduced to enhance the spectral information, and improve the restoration performance.

In summary, 3D convolution [9, 24] is appearing to explore spatial-spectral information, but the balance analysis between spatial and spectral modes is missing. Furthermore, most of existed works simply focus on one specific task and lose the transferability. The proposed STAS learns an adaptive spatial-spectral 3D convolution, and analyzes the transferability across different tasks.

One-shot neural architecture search (NAS). Since proposed in [56], NAS techniques have achieved great progress, and have been utilized in various applications, such as classification [56, 26, 29], object detection [37], and restoration [17, 51, 25, 22], among others.

At the outset, reinforcement learning [56], evolution [34], and sequential model-based optimization [26] were utilized to optimize the NAS. However, these algorithms are time-consuming and cost more than hundreds of GPU days for training on a small dataset. Subsequently, one-shot NAS was proposed to share weights during the training progress and significantly save search time [27, 46, 10, 40]. Specifically, one-shot NAS builds a super-net [27] to subsume all candidate architectures for the whole search space. Each candidate architecture is regarded as a discrete structure from the super-net and inherits the weight of the super-net, which is trained only once. In this case, the training time is significantly reduced.

We also utilize the one-shot NAS to design efficient architecture for HSI restoration. To make NAS workable for this problem, we design a spectrum-aware super-net that includes spatial, spectral and spatial-spectral convolutions. The gradient-based algorithm [27] is utilized to optimize the proposed super-net.

3 Method

In this section, we introduce the proposed spectrum-aware and transferable architecture search. We first discuss the limitation of existing methods for HSI

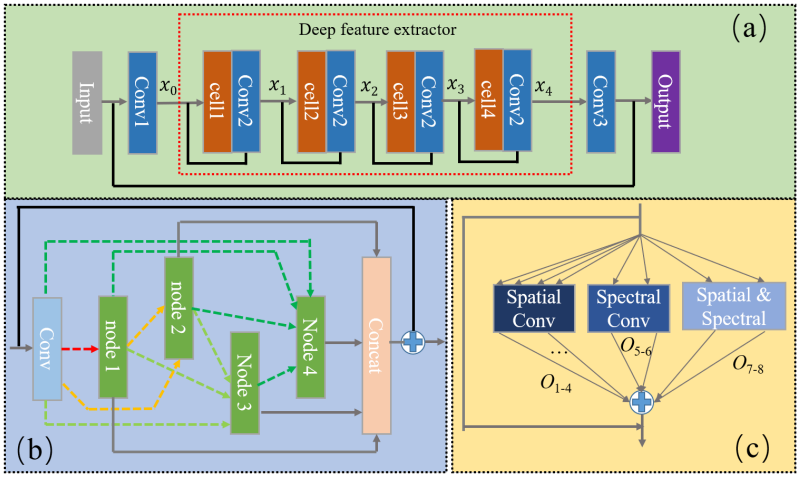

Fig. 1. Overview of the proposed STAS. (a) Global architecture used in STAS. (b) The architecture of super-cell. (c) MRB to connect the nodes in the super-cell.

restoration. Subsequently, we introduce the spectrum-aware search space design for STAS, including global architecture (Fig. 1(a)), micro architecture (Fig. 1(b)), and mixed residual block (Fig. 1(c)). Finally, we present the Noise level independent one-shot search algorithm for the STAS.

Until now, many works have been proposed for HSI restoration [52, 2, 3, 31, 24]. Inspired by these works, we summarize a baseline global architecture for the HSI restoration task, as presented in Fig. 1(a). The baseline global architecture consists of a shallow feature extractor, a deep feature extractor, a feature reconstruction module, and a global skip connection [52, 3]. The channel number of the input x is c. The shallow feature extractor `Conv1` outputs x_0 with $8 \times c$ channels. The extracted x_0 will be distilled by the deep feature extractor of n-layer repeated modules, with each module composed of the `cell` block and `Conv2`. We fix the layer of repeated modules as 4. The procedure is formulated as the following:

$$x_i = \texttt{Conv2}\big\{\texttt{Concat}(\texttt{cell}(x_{i-1}), x_{i-1})\big\}, \tag{1}$$

where $i = 1, \cdots, 4$. `cell` is the main block with $8 \times c$ channels for the input/output features. The input and output of the current `cell` block are stacked together, followed by `Conv2` to generate the input of the next `cell`. Subsequently, the distilled features from the deep feature extractor are processed by feature reconstruction `Conv3` to output the residual image. Finally, the output residual and the input images are combined by the global skip connection to formulate the final reconstructed HSI y.

By replacing `cell` with different operations, the baseline can approach different hard-crafted networks, such as HSI-DeNet [3] with spatial convolution and SIP [24] with separable 3D convolution. Regarding the operation choice for `cell`,

3D convolution has been proved efficient for the spatial-spectral exploration of 3D HSI data [14]. However, it is always associated with a huge computation burden and prevents the large-scale processing of HSIs. SIP combines a 2D depthwise convolution and a 2D pointwise convolution to formulate the separable 3D convolution, but the simple combination of the two cannot learn the spatial-spectral features well. To confirm this finding, we formulate 4 different nets, *i.e.*, `spe` net, `spa` net (in HSI-DeNet [3]), `SS1` of separable convolution (in SIP [24]), and finally `SS2` net with another composition of depthwise and pointwise convolutions. The 4 nets are illustrated in Fig. 7 of Supplementary Materials. The experiments presented in Fig. 8 of Supplementary materials shows that well composed spatial and spectral architecture of `SS2` can achieve better learning ability. That is to say, *the composition strategy between* ***`spa`*** *and* ***`spe`*** *is also important to efficiently learning the HSI features. Therefore, we are inspired to decouple the* ***`spe`*** *and* ***`spa`****, and learn an adaptive 3D convolution through the use of NAS.*

HSIs are always corrupted by different noise due to sensors and imaging condition; meanwhile different tasks always introduce different type of model errors. Due the uncertain noise in different dataset/tasks, the essence features exploration for HSIs are blocked by the limited hard-crafted architectures. On the other hand, from [20], different HSI restoration tasks can be integrated in an unified paradigm with different types of noise. This motivates us to explore the intrinsic architectures across HSI datasets/tasks by injecting the noise-independent analysis, with the intention of extending the applicability of proposed STAS.

3.1 Spectrum-aware search space

In this subsection, we present the search space design of the proposed STAS. The search space can be represented by a super-net, which consists of global architecture (Fig. 1(a)), micro architecture (Fig. 1(b)) and mixed residual block (MRB) (Fig. 1(c)). Similar to [27], we consider a fixed global architecture as the same of baseline but search for a better micro architecture and MRB, because cheap of the search in computational resource requirements. To implement the global architecture, we define the details of convolution in Fig. 1(a), *i.e.*, `Conv1` composed of 1×1 convolution, ReLU (rectified linear units), and BN (batch normalization); `Conv2` composed of spatial depthwise 3×3 convolution, spectral pointwise 1×1 convolution, ReLU, and BN; and `Conv3` composed of 1×1 convolution and 3×3 convolution.

Micro level. The cell topology structure is based on [27] with appropriate adjustment for HSI restoration. As presented in Fig. 1(b), we build a super-cell to integrate all possible convolutions. The super-cell is composed of one `Conv` and 4 nodes. The `Conv` is a series of operations that include the 1×1 convolution, ReLU, and BN. Specifically, the channel of input features is $8 \times c$, and `Conv` reduces it to $2 \times c$. We denote the features after `Conv` as x^0. The super-cell adopts x^0 as input, utilizes a directed acyclic graph to collect the sequence of 4

nodes, and outputs a tensor x_i by concatenating the outputs of 4 nodes with a local skip connection.

Mixed residual block (MRB). As presented in Fig. 1(c), we introduce the MRB to connect the possible two nodes, and fully exploit the spatial-spectral information. Our designed operation search space consists of 4 spatial convolutions, 2 spectral convolutions, and 2 spatial-spectral convolutions, as presented in Table 1.

Table 1. Operations utilized in the MRB.

Conv Type	Label	Operation
Spatial	o1	3×3 convolution
	o2	5×5 convolution
	o3	3×3 dilated convolution
	o4	3×3 dilated group convolution
Spectral	o5	1×1 convolution
	o6	1×1 group convolution
Spatial -spectral	o7	3×3 separable dilated convolution
	o8	3×3 separable dilated group convolution

3.2 Noise level independent search algorithm

Based on the above introduction, the overall search complexity of our proposed STAS is estimated as 9^{10}, which is challengeable for optimization. Inspired by [27, 40, 12], we are motivated to use one-shot NAS, which greatly reduce search time by training a super-net that contains all candidate architectures.

Search objective. The connection between the paired nodes (j, k) $(0 \leq j \leq 3, 1 \leq k \leq 4, j < k)$ using MRB can be formulated as

$$\mathrm{MRB}(x^j) = x^j + \sum\nolimits_{l=1}^{8} o_l(x^j), \tag{2}$$

where x^j is the output of node j, and l indicates the operation from Table 1. To discriminate the importance of different operations in Table 1 and make the search space continuous, a weight vector α^{jk} with softmax is utilized to improve MRB as the following:

$$\mathrm{MRB}^{jk}(x^j) = x^j + \sum\nolimits_{l=1}^{8} \frac{\exp(\alpha_l^{jk})}{\sum_{p=1}^{8} \exp(\alpha_p^{jk})} o_l(x^j). \tag{3}$$

Taking into consideration all the node pairs presented in Fig. 1(c), the task of STAS is to learn a set of continuous variables $\alpha = \{\alpha^{jk}\}$. After the training of the super-net STAS, for each MRB, the operator with the largest weight α_l^{jk} $(1 \leq l \leq 8)$, denoted as o^{jk}, is selected to replace the MRB, and finally a discrete operation of MRB is obtained.

Given a set of images $\mathcal{D} = \{x_i(\sigma), y_i\}$ where y_i is the clean image and $x_i(\sigma)$ is the noisy version of y_i with noise level σ. Let $\texttt{net}_\alpha$ be the STAS net with architecture represented by α. The goal of STAS net is to recover the clean image from $x(\sigma)$. We use the square loss to measure the recovered image to the clean one. Subsequently, the recovery performance of $\texttt{net}_\alpha$ on $\mathcal{D}$ is

$$\mathcal{L}(\theta, \alpha, \mathcal{D}) = 1/2 \sum\nolimits_{y \in \mathcal{D}} \left\| \texttt{net}_\alpha\big(x(\sigma); \theta\big) - y \right\|_2^2. \tag{4}$$

Thus, we have the following bi-level objective function for searching α:

$$\min_{\alpha} \quad \mathbb{E}_\sigma[\mathcal{L}(\theta^\star(\alpha, \sigma), \alpha, \mathcal{D}_{\text{val}})], \tag{5}$$

$$\text{s.t.} \quad \theta^\star(\alpha, \sigma) = \arg\min_{\theta} \mathcal{L}(\theta, \alpha, \mathcal{D}_{\text{tra}}) \tag{6}$$

where $\mathcal{D}_{\text{tra}}$ and $\mathcal{D}_{\text{val}}$ denote training set and validation set respectively.

Search algorithm. Unlike classical NAS, the network parameter θ in (6) depends on noise level, but α in (5) does not. In this way, we hope architectures can focus more on the statistics of the images not the noise. Thus, we do not need to search architectures for different noise levels. However, this requires us to propose new search algorithm to deal with $\mathbb{E}_\sigma$. Thus, we generate the inputs via adding random noise during the training and combine this process with DARTS, which leads to Algorithm 1.

Algorithm 1 Spectrum-aware and Transferable Architecture Search (STAS).

Input: Training data, validation data, MRB via (3)

Initialize network parameter θ and architecture parameter α
while not converged **do**
  Generate inputs via adding random noise on the training/validation data
  Update α by descending $\mathcal{L}(\theta(\alpha, \sigma), \alpha, \mathcal{D}_{\text{val}})$
  Update θ by descending $\nabla_\theta \mathcal{L}_{\text{train}}(\theta(\alpha), \alpha)$
return architecture α

Output: Discrete architecture from α.

Following our idea, more recent NAS works can also be adopted. Examples are DrNAS [6], SPOS [18], NASP [46], and DARTS- [10], which seek to improve the classical one-shot NAS method, *i.e.*, DARTS, from different directions. In experiments, we will further replace DARTS update rules in Algorithm 1 with SPOS as an usage of example..

3.3 Difference with existing works

Compared to representative hand-crafted architectures HSI-DeNet [3], SIP [24] and QRNN3D [42], we also expect that the learned architecture STAS can be utilized with different dataset and different noise cases. First, the proposed

spectrum-aware search space of STAS explores the domain-specific information of the HSI. Second, the learned architectures via STAS can be adapted to different noise levels. However, we can simply train STAS once, and apply the searched architecture to the restoration task on HSIs with similar spectral information, regardless of the noise level.

Table 2. Dataset introduction and implementation.

Task	Training set		Test set	
	patch size	number	size	number
STAS on CAVE ($STAS_C$)	$40\times40\times31$	22K	—	—
STAS on Pavia ($STAS_P$)	$40\times40\times60$	15K	—	—
CAVE denoising	$40\times40\times31$	120K	$300\times300\times31$	7
ICVL denoising	$40\times40\times31$	162K	$300\times300\times31$	7
Pavia denoising	$40\times40\times60$	15K	$200\times200\times60$	4
KAIST Imaging	$256\times256\times28$	5K	$256\times256\times28$	10

4 Experiments

In this section, we first introduce the datasets, including CAVE [6], ICVL [1], Pavia center, Pavia University [7], and KAIST [8]. Subsequently, we utilize the searched architecture for the different restoration tasks. Third, we analyze the learned discrete architectures via STAS, including the efficient understanding and transferability of the searched architecture.

4.1 Datasets and implementation details

The CAVE dataset contains 32 images of size $512 \times 512 \times 31$. We split it into training and test sets of 25 and 7 images, respectively. The images from ICVL are of size $1392 \times 1300 \times 31$. We select 120 images for the training, and 7 images of size $300 \times 300 \times 31$ for the test. Paiva datasets are from the remote sensing database, and we select the Pavia center of $1096 \times 715 \times 60$ for the training, and crop 4 patches of size $200 \times 200 \times 60$ from Pavia University for the test. In accordance with [30], we also choose 10 images of size $256 \times 256 \times 28$ from the KAIST dataset for the test of the imaging reconstruction task.

For the architecture search of STAS, we execute on the training sets of CAVE and Pavia Center. 2% of the training samples are chosen as the valid dataset, which is used to update the STAS architecture parameters. We train the STAS 100 epochs with a batch size of 8. The parameters of the network and the architecture are optimized separately by Adam. For the update of network parameters, we set the weight decay as $4e^{-4}$, and utilize the cosine annealing strategy to decay the learning rate from $1e^{-3}$ to $1e^{-6}$. For the update of architecture parameters, we set the weight decay as $1e^{-3}$, and the learning rate as $3e^{-4}$. In the training of STAS, we generate inputs by randomly adding noise of variations from $30, 50$, and 70 on the training/valid images.

For the denoising task, we retrain the discrete model learned from STAS on the training dataset, and apply the trained model to the related test dataset. We

[6] http://www1.cs.columbia.edu/CAVE/databases/
[7] http://www.ehu.eus/ccwintco/index.php/

keep the optimizer, weight decay, and learning rate the same as that of STAS. For the imaging reconstruction task, we train on the whole CAVE dataset, and test on the KAIST dataset. The imaging mask and the coded image are utilized to generate an initialization that is the same size as the output HSI. The input settings of this network are the same as that of [30].

We generate small patches from the training dataset, and utilize crop and rotation to augment the trained images. The patch size and related numbers for different tasks are presented in Table 2. We refer to the network obtained by STAS on CAVE as STAS_C, and on Pavia as STAS_P. The experiments are performed on a single NVIDIA Tesla V100 GPU.

In the following, we apply the searched architecture from CAVE dataset for denoising (Subsection 4.2) and imaging reconstruction (Subsection 4.3). Subsection 4.4 presents the understanding of the searched architecture, and Subsection 4.5 explains why the searched architecture STAS_C can be applied to different datasets and different HSI restoration tasks.

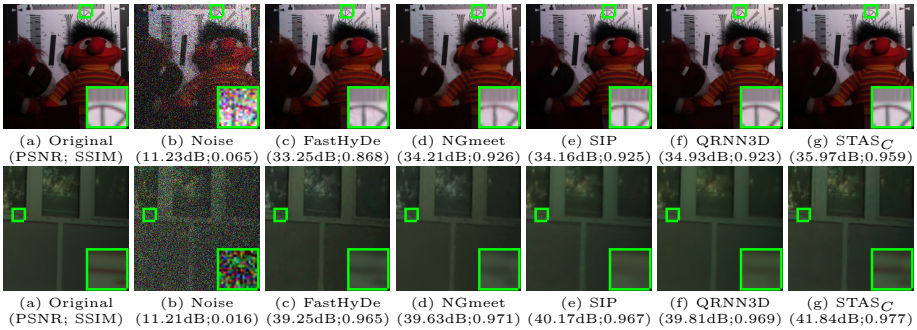

Fig. 2. Denoised results of different methods with noise variance 70. Top: CAVE toy image (R:28, G:11, B:6), bottom: ICVL image (R:30, G:11, B:6).

4.2 Denoising results

As elaborated in Section 4.4 and 4.5, STAS is spectrum-aware and independent of noise level. We only train the STAS on the CAVE images once, and apply the learned STAS_C to the denoising of different datasets (CAVE and ICVL from Table 2) with different noise levels (30, 50, and 70).

We compare the proposed STAS_C to classical SOTA methods FastHyDe [55], NGmeet [20], and learning-based methods, including SIP [24], and QRNN3D [42]. The peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) are adopted to evaluate the performance. Higher PSNR and SSIM mean better performance in spatial information.

Denoising on CAVE. Table 3 presents the average quantitative evaluation results of different methods on the 7 CAVE test images. The noise variance changes from $30, 50,$ to 70. In the low noise case, the NGmeet achieves the best performance. As the noise level increases, the learning-based methods, *i.e.*, QRNN3D and STAS_C, can beat NGmeet. In particular, in the noise case of

70, the proposed STAS$_C$ can improve 1.5dB on average compared to that of NGmeet. On the other hand, compared to other learning-based methods SIP and QRNN3D, our proposed STAS$_C$ can always achieve the best performance on average. Fig. 2 shows the visual results of different methods on the CAVE toy image with noise variance 70. It can be observed that the proposed STAS$_C$ achieves the best visual result in the balance of noise removal and details restoration.

Table 3. Denoising experiments on CAVE and ICVL datasets. The best result is in bold, while the second best is underlined.

	method	FastHyDe		NGmeet		SIP		QRNN3D		STAS$_C$	
	noise	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
CAVE	30	38.00	0.949	**39.05**	**0.963**	36.97	0.948	37.65	0.957	38.39	0.961
	50	35.53	0.911	36.38	0.941	35.81	0.937	35.84	0.935	**36.80**	**0.949**
	70	33.70	0.871	34.34	0.916	34.64	0.929	34.96	0.927	**35.83**	**0.940**
ICVL	30	42.96	0.971	43.42	0.973	41.58	0.960	42.08	0.967	**43.92**	**0.978**
	50	40.58	0.958	40.85	0.962	40.03	0.950	40.62	0.959	**42.01**	**0.969**
	70	38.86	0.941	39.21	0.950	38.88	0.930	39.21	0.943	**41.13**	**0.961**

Denoising on ICVL. Table 3 shows the average evaluation results of different methods on the 7 ICVL test images. Compared to NGmeet, QRNN3D can achieve similar PSNR values, but lower MSA values (in Supplementary Materials). That is to say, QRNN3D can achieve similar spatial restoration performance, but lower spectral performance. This is mainly because the advantage of NGmeet in spectral regularization. However, the proposed STAS$_C$ beats the classical methods and learning-based methods in the three evaluation indices of various noise levels. This indicates that our learned STAS$_C$ on CAVE can be successfully transferred to the ICVL denoising problem. Fig. 2 shows the visual results of different methods on ICVL images with noise variance 70. From the enlarged rectangle, we can see that our proposed method can clearly recover window edge, while other compared methods produce blurry details. Detailed evaluation results are presented in Supplementary Materials.

Table 4. Imaging reconstruction on the KAIST dataset. The best result is in bold, while the second best is underlined.

method	DeSCI		HSSP		λ-net		TSA		GSM-based		STAS$_C$	
index	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	27.15	0.787	31.07	0.851	30.82	0.874	31.26	0.882	32.17	0.915	**32.52**	**0.918**
2	22.26	0.687	26.31	0.799	26.30	0.845	26.88	0.856	27.20	0.897	**28.55**	**0.898**
3	26.56	0.873	29.00	0.879	29.42	0.915	**30.03**	0.921	30.02	**0.925**	29.57	0.915
4	39.00	**0.964**	38.24	0.926	37.37	0.961	**39.90**	0.963	39.20	0.963	38.83	0.955
5	24.66	0.774	27.98	0.831	27.84	0.865	28.89	0.878	28.19	0.882	**29.62**	**0.921**
6	24.80	0.742	29.16	0.819	30.69	0.892	31.30	0.891	32.84	**0.937**	**33.47**	0.933
7	20.03	0.763	24.11	0.854	24.20	0.874	25.16	0.886	25.29	0.886	**25.81**	**0.890**
8	23.98	0.725	27.94	0.804	28.86	0.873	29.69	0.880	31.38	0.923	**31.65**	**0.936**
9	25.94	0.809	29.14	0.829	29.33	0.898	30.03	0.898	29.67	0.911	**31.23**	**0.917**
10	24.28	0.647	26.44	0.731	27.66	0.836	28.32	0.841	**30.52**	**0.925**	30.13	0.911
Average	25.86	0.777	28.94	0.832	29.25	0.883	30.15	0.890	30.65	0.916	**31.14**	**0.919**

4.3 Imaging reconstruction results

In accordance with [30], we directly utilize the proposed STAS$_C$ instead of TSA net for the HSI imaging reconstruction task, with the CAVE dataset for training and the KAIST dataset for testing. We slightly adjust the input and output

channels of STAS_C to fit the KAIST images. We compare our proposed method to DeSCI [28], HSSP [38], λ-net [31], TSA [30], and GSM-based [23]. Table 4 presents the evaluation results on 10 images from the KAIST dataset. From the table, it can be observed that in most cases, our proposed STAS_C can achieve the best accuracy compared to the previous methods. Specifically, our proposed STAS_C can improve nearly 1dB in PSNR compared to TSA and 0.5dB compared to GSM-based. The proposed STAS_C has $1.43M$ parameters, much less than the $44.3M$ of TSA and the $3.76M$ of GSM-based, further indicating the efficiency of our STAS architecture. The visual results of different methods are presented in Supplementary Materials.

4.4 Understanding of STAS for HSIs

Effectiveness of the search space. Table 5 presents the ablation study of proposed STAS with different attributes. Firstly, compared to *baseline* HSI-DeNet [3] and search space DARTS [27], our proposed search space of STAS can achieve better accuracy. Secondly, from *architecture design*, the local skip connections used in Figs. 1(b-c), and the global skip connection are proved to be efficient.

Table 6 shows the results of STAS replaced with different operations and improved algorithms. Regarding *operation*, we replace the learned operations of STAS with other operations, including `spe`, `spa`, `SS1` (in subsection 3) and random selection (in [27]). It is obvious that the learned adaptive spatial-spectral convolution via STAS is more suitable for the restoration of HSIs. Regarding *algorithm*, although we replace different improved algorithms [10, 18] to optimize STAS, the improvement is limited. It indicates that the designed search space of STAS is good enough for HSI processing. Since the skip connections has been adjusted in STAS, The DARTS- [10] decreases the result on our designed search space.

Table 5. Ablation study of the proposed STAS on the CAVE with noise 50. In *baseline*, "DARTS" means search space used in [27].

attribute	*baseline*		*architecture design*				
	HSI-DeNet	DARTS	STAS	STAS	STAS	STAS	STAS
skip in Fig. 1(b)					√		√
skip in Fig. 1(c)				√	√	√	√
global *skip*	√					√	√
PSNR(dB)	33.85	32.18	33.87	34.52	35.73	35.22	36.80

Table 6. Ablation study of the proposed STAS with different operations and algorithms on the CAVE with noise 50.

attribute	*operation*				*algorithm*		
	random	`spe`	`spa`	`SS1`	DARTS-	SPOS	STAS
PSNR(dB)	35.25	32.81	34.72	35.67	36.42	36.89	36.80

Fig. 3 presents the PSNR values with the GPU hours of the proposed search space STAS and the well-known search space DARTS on the CAVE dataset.

STAS(SPOS) means the proposed search space optimized by SPOS [18]. We slightly adjust the last layer of DARTS for the HSI restoration task. In fact, STAS(SPOS) achieves the best accuracy faster, but the final accuracy is similar to STAS. Overall, the performance of proposed search space is much better than that of DARTS. Therein, it motivates us to pay more attention to the search space design for HSIs, but not the optimization algorithms.

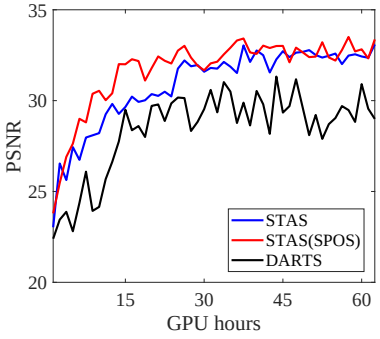

Fig. 3. PSNR values with the GPU hours on different search spaces STAS/DARTS.

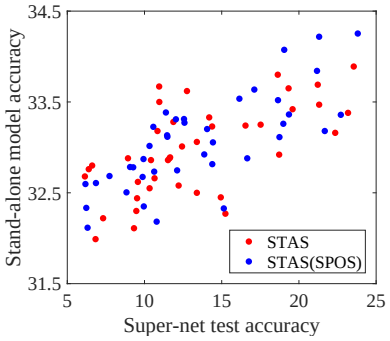

Fig. 4. Comparison of super-nets and stand-alone model accuracies from a statistical sample of architectures.

Quality of the trained super-net. We attempt to access the quality of the trained super-net STAS. Based on [47], we randomly select the operations and connections from the super-cell (Fig. 1(b)) to formulate the discrete architecture (such as Fig. 5), and remove the rest from the super-cell. With this randomly selected architecture, we can obtain a super-net test accuracy by using the weight of the super-net STAS trained on CAVE. From another perspective, we retrain this random architecture on a small training set, and obtain the test accuracy, which we call the stand-alone model accuracy. We randomly select 40 architectures; the obtained accuracy pairs of the super-net STAS and the stand-alone model are presented in Fig. 4. It can be seen that higher super-net test accuracy means higher stand-alone model accuracy, indicating that the searched architectures from super-net STAS can predict the importance of the weight and operations in the final discrete architecture. We also utilize SPOS to optimize the proposed designed architecture, and achieve similar results compared to that STAS.

Table 7. Quantitative evaluation with noise variance 50 on CAVE/Pavia denoising.

Method	Index	FastHyDe	NGmeet	STAS$_C$	STAS$_P$
CAVE	PSNR	33.53	36.38	**36.80**	35.92
	SSIM	0.911	0.941	**0.949**	0.936
	MSA	9.33	6.12	**5.34**	6.04
Pavia	PSNR	33.93	34.80	33.10	**34.96**
	SSIM	0.913	0.926	0.917	**0.933**
	MSA	4.86	3.98	4.20	**3.64**

Table 8. Training (hour) and testing (second) cost of different methods on CAVE denoising and KAIST imaging tasks.

Time (s)	FastHyDe	NGmeet	SIP	QRNN3D	STAS
			train (h)/test	train (h)/test	train (h)/test
CAVE denoising	21	68	35h/1.5s	23h/1.2s	12h/0.8s
Time (s)	GAP-TV	DeSCI	lambda net	TSA	STAS
			train (h)/test	train (h)/test	train (h)/test
KAIST imaging	120	7,928	31h/1.2s	25h/1.6s	11h/0.6s

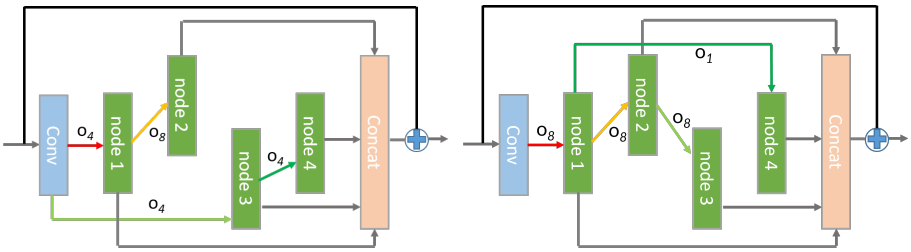

Fig. 5. The cell learned on CAVE dataset. **Fig. 6.** The cell learned on Pavia dataset.

Computational efficiency. Table 8 presents the time cost of learned STAS$_C$ and the compared methods on CAVE denoising and KAIST imaging reconstruction tasks. From the table, it can be observed that the deep-learning-based methods can be very fast in the test stage compared to the classical methods. Compared to other deep-learning-based methods, the proposed STAS$_C$ can achieve better performance with many fewer parameters and less training time, demonstrating the efficiency of the searched architecture.

Visualization of searched architectures. The searched cell structure of STAS on CAVE denoising is presented in Fig. 5, while on Paiva denoising is illustrated in Fig. 6. From the comparison of these two cell structures on two HSI datasets, we make the following observations.

- For the CAVE dataset with fewer spectral bands, the discrete operators are dominated by the spatial group dilated convolution. This indicates that the spatial convolution is enough to explore the prior knowledge of CAVE images.
- For the Paiva with larger spectral bands, the separable convolution dominates the cell block, indicating the necessity of joint spatial/spectral exploration.
- The proposed STAS is dependent on the spectral information, and independent of the noise level. We will further illustrate this in Subsection 4.5.

4.5 Transferability

Dependency of spectral information. The experimental results of STAS$_C$ on CAVE denoising, ICVL denoising, and KAIST imaging reconstruction demonstrate the transferability of the spectrum-aware searched architecture. We also extend our learned STAS$_C$ and STAS$_P$ to cross-validate the performance on CAVE and Pavia image denoising tasks. We adjust the input/output channels of STAS$_C$ to denoise Pavia images with different spectral information. From Table 7, it can be observed that the STAS$_C$ performs worse to the baseline

NGmeet on the Pavia denoising task, whereas STAS_P can achieve much better performance. For the CAVE image denoising task, the STAS_C can obtain much better performance compared to STAS_P. We can see that the transferability of STAS_C from CAVE to Pavia datasets gets stuck. This is mainly because the spectral information of the two datasets is different. By comparing the architectures of STAS_C (Fig. 5) and STAS_P (Fig. 6), we can conclude that the transferability is dependent on spectral information. For HSIs with fewer spectral bands, the spatial convolution is important, while, the spatial-spectral convolution is efficient for HSIs with higher spectral bands.

Table 9. Denoising results on CAVE test images with different noise levels. $\text{STAS}_{\text{adap}}$ means the network learned by STAS on the corresponding noise level.

Noise	30	50	70
STAS_C	38.39	**36.80**	35.83
$\text{STAS}_{\text{adap}}$	**38.52**	36.67	**35.99**

Independency of the noise level. We train STAS only once with the mixed noise, and apply the learned STAS_C to different noise level denoising of the CAVE and ICVL datasets. The experiments presented in subsection 4.2 prove the efficiency of STAS_C on different noise levels. Furthermore, we retrain STAS on the CAVE dataset with the fixed noise level, *e.g.* train the STAS on CAVE with noise variance 30 and apply the searched architecture (marked as $\text{STAS}_{\text{adap}}$) for the CAVE denoising with 30 of noise variance. Table 9 presents the comparison of STAS_C and $\text{STAS}_{\text{adap}}$. We can see that the differences between the two searched architectures are not that significant, further demonstrating the independency of the noise level.

5 Conclusions

In this work, we propose a spectrum-aware and transferable architecture search for HSI restoration across different datasets. To make the proposed STAS workable for HSI restoration, we design a spectrum-aware search space that includes spatial, spectral, and spatial-spectral convolutions. Furthermore, the proposed super-net STAS is equipped with local and global skip connections to boost performance. The fruitful experiments also demonstrate that the proposed STAS_C can be successfully transferred to different dataset (ICVL) denoising and different tasks (KAIST imaging reconstruction), and achieve SOTA performance. The transferability of searched architecture is dependent on the spectral information, and independent of the noise levels. Specifically, the spatial convolution is efficient for fewer spectral band HSI processing, whereas the spatial-spectral convolution is meaningful for higher spectral band HSI.

6 Acknowledgments

This work is partially supported by the National Natural Science Foundation of China under grants 61871298 and 42071322, and partially supported by the JST, FOREST Grant Number JPMJFR206S, Japan.

References

1. Arad, B., Ben-Shahar, O.: Sparse recovery of hyperspectral signal from natural rgb images. In: European Conference on Computer Vision. pp. 19–34. Springer (2016)
2. Bodrito, T., Zouaoui, A., Chanussot, J., Mairal, J.: A trainable spectral-spatial sparse coding model for hyperspectral image restoration. Advances in Neural Information Processing Systems **34** (2021)
3. Chang, Y., Yan, L., Fang, H., Zhong, S., Liao, W.: Hsi-denet: Hyperspectral image restoration via convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing **57**(2), 667–682 (2019). https://doi.org/10.1109/TGRS.2018.2859203
4. Chang, Y., Yan, L., Zhao, X.L., Fang, H., Zhang, Z., Zhong, S.: Weighted low-rank tensor recovery for hyperspectral image restoration. IEEE transactions on cybernetics **50**(11), 4558–4572 (2020)
5. Chang, Y., Yan, L., Zhong, S.: Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In: CVPR. pp. 4260–4268 (2017)
6. Chen, X., Wang, R., Cheng, M., Tang, X., Hsieh, C.J.: Drnas: Dirichlet neural architecture search. In: ICLR (2021)
7. Chen, Y., Huang, T.Z., He, W., Zhao, X.L., Zhang, H., Zeng, J.: Hyperspectral image denoising using factor group sparsity-regularized nonconvex low-rank approximation. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–16 (2021)
8. Choi, I., Jeon, D.S., Nam, G., Gutierrez, D., Kim, M.H.: High-quality hyperspectral reconstruction using a spectral prior. ACM Transactions on Graphics (TOG) **36**(6), 1–13 (2017)
9. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
10. Chu, X., Wang, X., Zhang, B., Lu, S., Wei, X., Yan, J.: Darts-: robustly stepping out of performance collapse without indicators. International Conference on Learning Representations (2020)
11. Dian, R., Fang, L., Li, S.: Hyperspectral image super-resolution via non-local sparse tensor factorization. In: Proc. CVPR. pp. 3862–3871 (July 2017). https://doi.org/10.1109/CVPR.2017.411
12. Ding, Y., Yao, Q., Zhao, H., Zhang, T.: Diffmg: Differentiable meta graph search for heterogeneous graph neural networks. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 279–288. KDD'21, New York, NY, USA (2021). https://doi.org/10.1145/3447548.3467447
13. Dong, W., Li, G., Shi, G., Li, X., Ma, Y.: Low-rank tensor approximation with laplacian scale mixture modeling for multiframe image denoising. In: ICCV. pp. 442–449 (2015)
14. Dong, W., Wang, H., Wu, F., Shi, G., Li, X.: Deep spatial spectral representation learning for hyperspectral image denoising. IEEE Transactions on Computational Imaging **5**(4), 635–648 (2019). https://doi.org/10.1109/TCI.2019.2911881
15. Dong, Y., Liang, T., Zhang, Y., Du, B.: Spectral spatial weighted kernel manifold embedded distribution alignment for remote sensing image classification. IEEE Transactions on Cybernetics **51**(6), 3185–3197 (2021). https://doi.org/10.1109/TCYB.2020.3004263
16. Fu, Y., Zheng, Y., Sato, I., Sato, Y.: Exploiting spectral-spatial correlation for coded hyperspectral image restoration. In: Proc. CVPR. pp. 3727–3736 (Jun 2016). https://doi.org/10.1109/CVPR.2016.405

17. Gou, Y., Li, B., Liu, Z., Yang, S., Peng, X.: Clearer: Multi-scale neural architecture search for image restoration. Advances in Neural Information Processing Systems **33** (2020)
18. Guo, Z., Zhang, X., Mu, H., Heng, W., Liu, Z., Wei, Y., Sun, J.: Single path one-shot neural architecture search with uniform sampling. In: ECCV (2020)
19. He, W., Chen, Y., Yokoya, N., Li, C., Zhao, Q.: Hyperspectral super-resolution via coupled tensor ring factorization. Pattern Recognition **122**, 108280 (2022)
20. He, W., Yao, Q., Li, C., Yokoya, N., Zhao, Q., Zhang, H., Zhang, L.: Non-local meets global: An integrated paradigm for hyperspectral image restoration. IEEE Trans. Pattern Anal. Mach. Intell. **44**(4), 2089–2107 (2022). https://doi.org/10.1109/TPAMI.2020.3027563
21. Hong, D., He, W., Yokoya, N., Yao, J., Gao, L., Zhang, L., Chanussot, J., Zhu, X.: Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing. IEEE Geosci. and Remote Sens. Mag. **9**(2), 52–87 (Jun 2021). https://doi.org/10.1109/MGRS.2021.3064051
22. Huang, H., Shen, L., He, C., Dong, W., Huang, H., Shi, G.: Lightweight image super-resolution with hierarchical and differentiable neural architecture search. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021)
23. Huang, T., Dong, W., Yuan, X., Wu, J., Shi, G.: Deep gaussian scale mixture prior for spectral compressive imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16216–16225 (June 2021)
24. Imamura, R., Itasaka, T., Okuda, M.: Zero-shot hyperspectral image denoising with separable image prior. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops (Oct 2019)
25. Lee, R., Dudziak, Ł., Abdelfattah, M., Venieris, S.I., Kim, H., Wen, H., Lane, N.D.: Journey towards tiny perceptual super-resolution. In: European Conference on Computer Vision. pp. 85–102. Springer (2020)
26. Liu, C., Zoph, B., Neumann, M., Shlens, J., Hua, W., Li, L.J., Fei-Fei, L., Yuille, A., Huang, J., Murphy, K.: Progressive neural architecture search. In: Proceedings of the European conference on computer vision (ECCV). pp. 19–34 (2018)
27. Liu, H., Simonyan, K., Yang, Y.: Darts: Differentiable architecture search. In: International Conference on Learning Representations (2019)
28. Liu, Y., Yuan, X., Suo, J., Brady, D.J., Dai, Q.: Rank minimization for snapshot compressive imaging. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(12), 2990–3006 (2019). https://doi.org/10.1109/TPAMI.2018.2873587
29. Ma, A., Wan, Y., Zhong, Y., Wang, J., Zhang, L.: Scenenet: Remote sensing scene classification deep learning network using multi-objective neural evolution architecture search. ISPRS Journal of Photogrammetry and Remote Sensing **172**, 171–188 (2021)
30. Meng, Z., Ma, J., Yuan, X.: End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In: Proc. ECCV. pp. 187–204. Springer (2020)
31. Miao, X., Yuan, X., Pu, Y., Athitsos, V.: λ-net: Reconstruct hyperspectral images from a snapshot measurement. In: Proc. ICCV (Oct 2019)
32. Pan, Z., Healey, G., Prasad, M., Tromberg, B.: Face recognition in hyperspectral images. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12), 1552–1560 (Dec 2003). https://doi.org/10.1109/TPAMI.2003.1251148
33. Peng, Y., Meng, D., Xu, Z., Gao, C., Yang, Y., Zhang, B.: Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In: CVPR. pp. 2949–2956 (2014)

34. Real, E., Aggarwal, A., Huang, Y., Le, Q.V.: Regularized evolution for image classifier architecture search. In: Proceedings of the aaai conference on artificial intelligence. vol. 33, pp. 4780–4789 (2019)
35. Shi, Q., Tang, X., Yang, T., Liu, R., Zhang, L.: Hyperspectral image denoising using a 3-d attention denoising network. IEEE Transactions on Geoscience and Remote Sensing pp. 1–16 (2021). https://doi.org/10.1109/TGRS.2020.3045273
36. Stein, D.W., Beaven, S.G., Hoff, L.E., Winter, E.M., Schaum, A.P., Stocker, A.D.: Anomaly detection from hyperspectral imagery. IEEE Signal Process Mag. **19**(1), 58–69 (2002)
37. Sun, P., Zhang, W., Wang, H., Li, S., Li, X.: Deep rgb-d saliency detection with depth-sensitive attention and automatic multi-modal fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1407–1417 (2021)
38. Wang, L., Sun, C., Fu, Y., Kim, M.H., Huang, H.: Hyperspectral image reconstruction using a deep spatial-spectral prior. In: Proc. CVPR. pp. 8032–8041 (2019)
39. Wang, L., Xiong, Z., Shi, G., Wu, F., Zeng, W.: Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. IEEE Trans. Pattern Anal. Mach. Intell. **39**(10), 2104–2111 (Oct 2016)
40. Wang, R., Cheng, M., Chen, X., Tang, X., Hsieh, C.J.: Rethinking architecture selection in differentiable nas. International Conference on Learning Representations (2021)
41. Wang, Y., Peng, J., Zhao, Q., Leung, Y., Zhao, X.L., Meng, D.: Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens **11**(4), 1227–1243 (2018)
42. Wei, K., Fu, Y., Huang, H.: 3d quasi recurrent neural network for hyperspectral image denoising. IEEE Transactions on Neural Networks and Learning Systems **32**(1), 363–375 (2021). https://doi.org/10.1109/TNNLS.2020.2978756
43. Xie, Q., Zhao, Q., Meng, D., Xu, Z.: Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. IEEE Trans. Pattern Anal. Mach. Intell. **40**(8), 1888–1902 (2018)
44. Xie, T., Li, S., Fang, L., Liu, L.: Tensor completion via nonlocal low-rank regularization. IEEE Transactions on Cybernetics **49**(6), 2344–2354 (2019). https://doi.org/10.1109/TCYB.2018.2825598
45. Xiong, F., Zhou, J., Qian, Y.: Material based object tracking in hyperspectral videos. IEEE Transactions on Image Processing **29**, 3719–3733 (2020). https://doi.org/10.1109/TIP.2020.2965302
46. Yao, Q., Xu, J., Tu, W.W., Zhu, Z.: Efficient neural architecture search via proximal iterations. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 6664–6671 (2020)
47. Yu, K., Ranftl, R., Salzmann, M.: An analysis of super-net heuristics in weight-sharing nas. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021)
48. Yuan, Q., Zhang, Q., Li, J., Shen, H., Zhang, L.: Hyperspectral image denoising employing a spatial spectral deep residual convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing **57**(2), 1205–1218 (2019). https://doi.org/10.1109/TGRS.2018.2865197
49. Yuan, X., Liu, Y., Suo, J., Durand, F., Dai, Q.: Plug-and-play algorithms for video snapshot compressive imaging. arXiv preprint arXiv:2101.04822 (2021)
50. Zhang, H., He, W., Zhang, L., Shen, H., Yuan, Q.: Hyperspectral image restoration using low-rank matrix recovery. IEEE Trans. Geosci. Remote Sens. **52**(8), 4729–4743 (Aug 2014)

51. Zhang, H., Li, Y., Chen, H., Shen, C.: Memory-efficient hierarchical neural architecture search for image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3657–3666 (2020)
52. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing **26**(7), 3142–3155 (2017)
53. Zhang, S., Wang, L., Fu, Y., Zhong, X., Huang, H.: Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery. In: Proc. ICCV (Oct 2019)
54. Zheng, S., Liu, Y., Meng, Z., Qiao, M., Tong, Z., Yang, X., Han, S., Yuan, X.: Deep plug-and-play priors for spectral snapshot compressive imaging. Photonics Research **9**(2), B18–B29 (2021)
55. Zhuang, L., Bioucas-Dias, J.M.: Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations. IEEE J. Sel.Topics Appl. Earth Observ. Remote Sens. **11**(3), 730–742 (Mar 2018)
56. Zoph, B., Le, Q.V.: Neural architecture search with reinforcement learning. International Conference on Learning Representations (2017)