Lightweight Hyperspectral Image Reconstruction Network with Deep Feature Hallucination

Kazuhiro Yamawaki[1 [0000−0003−4670−548X]] and Xian-Hua Han[1 [0000−0002−5003−3180]]

Yamaguchi University, Yamaguchi, 753-8511, Japan

Abstract. Hyperspectral image reconstruction from a compressive snapshot is an dispensable step in the advanced hyperspectral imaging systems to solve the low spatial and/or temporal resolution issue. Most existing methods extensively exploit various hand-crafted priors to regularize the ill-posed hyperspectral reconstruction problem, and are incapable of handling wide spectral variety, often resulting in poor reconstruction quality. In recent year, deep convolution neural network (CNN) has became the dominated paradigm for hyperspectral image reconstruction, and demonstrated superior performance with complicated and deep network architectures. However, the current impressive CNNs usually yield large model size and high computational cost, which limit the wide applicability in the real imaging systems. This study proposes a novel lightweight hyperspectral reconstruction network via effective deep feature hallucination, and aims to construct a practical model with small size and high efficiency for real imaging systems. Specifically, we exploit a deep feature hallucination module (DFHM) for duplicating more features with cheap operations as the main component, and stack multiple of them to compose the lightweight architecture. In detail, the DFHM consists of spectral hallucination block for synthesizing more channel of features and spatial context aggregation block for exploiting various scales of contexts, and then enhance the spectral and spatial modeling capability with more cheap operation than the vanilla convolution layer. Experimental results on two benchmark hyperspectral datasets demonstrate that our proposed method has great superiority over the state-of-the-art CNN models in reconstruction performance as well as model size.

Keywords: Hyperspectral image reconstruction · Lightweight network· Feature hallucination.

1 Introduction

Hyperspectral imaging (HSI) systems is able of capturing the detailed spectral distribution with decades or hundreds of bands at each spatial location of a scene. The abundant spectral signature in HSI possesses the deterministic attributes about the lighting and imaged object/material, which greatly benefits the characterization of the captured scene in wide fields, including remote sensing [14, 4], vision inspection [23, 24], medical diagnosis [3, 20] and digital forensics [10]. To capture a full 3D HSI, the conventional hyperspectral sensors have to employ multiple exposures to scan the target scene [9, 6, 5, 26], and require long imaging time failing in dynamic scene capturing.

To enable the HS image measurement for moving objects, various snapshot hyperspectral imaging systems [12, 18] have been exploited by mapping different spectral bands (either single narrow band or multiplexed ones) to different positions and then collecting them by one or more detectors, which unavoidably cause low-resolution in spatial domain. Motivated by the compressive sensing theory, a promising imaging modality: coded aperture snapshot spectral imaging (CASSI) [29, 13, 1] has attracted increasing attention. With the elaborated optical design, CASSIs encode the 3D HSI into a 2D compressive snapshot measurement, and require a reconstruction phase to recover the underlying 3D cubic data on-line.

Although extensive studies [17, 39, 27, 25, 2, 43, 44, 32] have been exploited, to faithfully reconstruct the desirable HSI from its compressive measurement is still the bottleneck in the CASSIs. Due to the ill-posed nature of the reconstruction problem, traditional model-based methods widely incorporate various hand-crafted priors of the underlying HSIs, such as the total variation [33, 34, 40, 2], sparsity [11, 30, 2] and low-rankness [19, 42], and demonstrate some improvements in term of the reconstruction performance. However, the hand-crafted priors are empirically designed, and usually deficient to model the diverse attributes of the real-world spectral data.

Recently, deep convolutional neural network (DCNN) [36, 8, 22, 35, 31] has popularly been investigated for HSI reconstruction by leveraging its powerful modeling capability and automatically learning of the inherent priors in the latent HSI using the previously collected external dataset. Compared with the model-based methods, these deep learning-based paradigms have prospectively achieved superior performance, and been proven to provide fast reconstruction in test phase. However, the current researches mainly focus on designing more complicated and deeper network architecture for pursuing performance gain, and thus cause a large-scale reconstruction model. However, the large-scale model would restrict wide applicability for being implanted in real HSI systems. More recently, incorporating the deep learned priors with iterative optimization procedure has been investigated to increase the flexibility of deep reconstruction model, and the formulated deep unrolling based optimization methods e.g., LISTA [15] ADMMNet [21, 37] and ISTA-Net [41] have manifested acceptable performance for the conventional compressive sensing problem but still have insufficient spectral recovery capability for the HSI reconstruction scenario.

To this end, this study aims to exploit a practical deep reconstruction model with small size and high efficiency for being easily embedded in the real imaging systems, and proposes a novel lightweight hyperspectral reconstruction network (LWHRN) via hallucinating/duplicating the effective deep feature from the already learned ones. As proven in [16], the learned feature maps in the well-trained deep models such as in the ResNet-50 using the ImageNet dataset may have abundance or even redundant information, which often guarantees the comprehensive understanding of the input data, and some features can be obtained with a more cheap transformation operation from other feature maps instead of the vanilla convolution operation. Inspired by the above insight, we specifically exploit a deep feature hallucination module (DFHM) for synthesizing more features with cheap operations as the main components of our LWHRN model, and stack multiple of them to gradually reconstruct the the residual component un-recovered in the previous phase. Concretely, the DFHM consists of spectral hal-

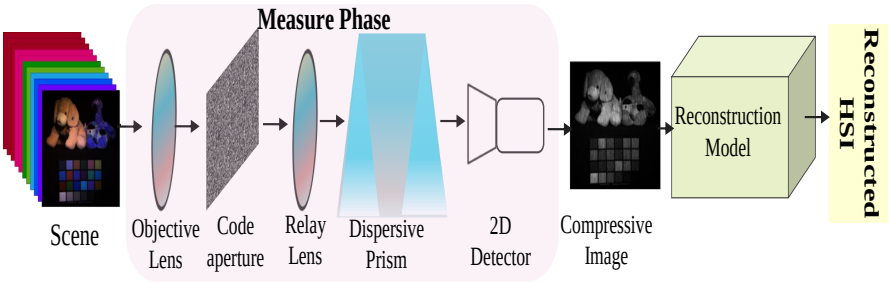


Fig. 1. The schematic concept of the CASSI system.

lucination block (SHB) for synthesizing more channel of features and spatial context aggregation block (SCAB) for exploiting various scales of contexts, where both SHB and SCAB are implemented using depth-wise convolution layer instead of the vanilla convolution layer, which can be expected to enhance the spectral and spatial modeling capability with the more cheap operation. Experimental results on two benchmark hyperspectral datasets demonstrate that our proposed LWHRN manifests great superiority over the state-of-the-art CNN models in reconstruction performance as well as model size.

In summary, the main contributions are three-fold:

1. We present a novel lightweight hyperspectral reconstruction network from a single snapshot measurement, which employs multiple reconstruction modules to gradually recover the residual HS components by alternatively incorporating spectral and spatial context learning.
2. We exploit a deep feature hallucination module (DFHM) as the main component of the multi-stage reconstruction module, which consists of a spectral hallucination block (SHB) for synthesizing more channel of features and a spatial context aggregation block (SCAB) for exploiting various scales of contexts using more cheap depth-wise convolution than the vanilla convolution.
3. We conduct extensive experiments on two benchmark HSI datasets, and demonstrate superior results over the SoTA reconstruction models in term of the reconstruction performance, model size and computational cost.

2 Related Work

Recently, the hyperspectral reconstruction in the computational spectral imaging have attracted extensive attention, and different kinds of methods, which are mainly divided into optimization-based methods and deep learning-based methods, have been proposed for improving reconstruction performance. In this section, we briefly survey the related work.

2.1 Optimization-based methods

The HSI reconstruction from a snapshot measurement is inherently a inverse problem, and can be intuitively formulated as the minimization problem of the reconstruction error of the the observed snapshot. Since the number of the unknown variables in the latent HSI is much larger than the known variables in the observed snapshot image, this inverse problem has severe ill-posed nature, and would cause quite un-stable solution via directly optimization. Existing methods have striven to incorporate various hand-crafted priors such as modeling the spatial structure and spectral characteristics of the latent HSI, into the inverse problem, and then formulate as a regularization term for robust optimization. Taking the high dimensionality of spectral signatures into accounting, different image local priors for characterizing the spectral image structure within a local region have popularly been exploited. For example, Wang et al. [33] exploited a Total Variation (TV) regularized model by imposing the first-order gradient smoothness prior for spectral image reconstruction while Yuan et. al. [40] proposed to employ a generalized alternating projection to solve the TV-regularized model (GAP-TV). Further, Kittle et. al. [17] explored two-step iterative shrinkage/thresholding method (TwIST) for optimization. Although the incorporation of the TV prior for the HSI reconstruction potentially benefits both boundary preservation and smooth region recovery, the reconstructed result may loss some detail structure. Motivated by the successful application in the blind compressed sensing (BCS) [25], sparse representation algorithms have been applied for HSI reconstruction from the snapshot image, which optimizes the representation coefficients with the sparsity prior constraint on the learned dictionary for local image patches [18]. Later, Yuan et. al. imposed the compressibility constraint instead of sparsity prior and proposed a global-local shrinkage prior to learn the dictionary and representation coefficients [39]. Moreover, Wang et. al. [30] incorporated the non-local similarity into a 3D non-local sparse representation model for boosting reconstruction performance. However, the hand-crafted image priors are not always sufficient to capture the characteristics in various spectral images, and thus cause unstable reconstruction performance. Furthermore, the suitable priors for different images would be varied, and to discover a proper prior for a specific scene is a hard task in the real scenario.

2.2 Deep learning-based methods

Benefiting from the powerful modeling capability, deep learning-based methods have been widely used for image restoration tasks including HSI reconstruction. The deep learning-based HSI reconstruction methods can implicitly learn the underlying prior from the previously prepared training samples instead of manually designing priors for modeling the spatial and spectral characteristics of the latent HSI, and then construct a mapping model between the compressed snapshot image and the desirable HSI. Various deep networks have been proposed for the HSI reconstruction problem. For example, Xiong et al [36] employed several vanilla convolution layer-based network (HSCNN) to learn a brute-force mapping between the latent HS image and its spectrally under-sampled projections, and demonstrated the feasibility for HSI reconstruction from a common RGB image or a compressive sensing (CS) measurement. Wang

et al. [35] proposed a joint coded aperture optimization and HSI reconstruction network for simultaneously learning the optimal sensing matrix and the latent HS image in an end-to-end framework while Miao et al. [22] developed a λ-net by integrating both the sensing mask and the compressed snapshot measurement for hierarchically reconstructing the HSI with a dual-stage generative model. Later Wang et al. [31] conducted multi-stage deep spatial-spectral prior (DSSP) modeling to incorporate both local coherence and dynamic characteristics for boosting the HSI reconstruction performance. Although promising performance has been achieved with the deep networks, the current researches mainly focus on designing more complicated and deeper network architecture for pursuing performance gain, and thus cause a large-scale reconstruction model. However, the large-scale model would restrict wide applicability for being implanted in real HSI systems.

In order to enhance the flexibility and interpretability of the deep reconstruction model, several works recently incorporated the deep learned priors into iterative optimization procedure, and proposed the deep unrolling based optimization methods in natural compressive sensing, e.g., LISTA [15] ADMMNet [21, 37] and ISTA-Net [41]. These methods unroll the iterative optimization procedure into a serial of learnable subproblems, and aim at simultaneously learning the network parameters for modelling the deep priors and the image updating parameters according to the reconstruction formula. However, they were proposed for solving natural compressive sensing problem via elaborately modelig the latent spatial structure, and are insufficient to capture the spectral prior in the high-dimensional HSIs. In order to effectively model the prior in the spectral domain, Choi et al. [8] proposed a convolutional auto-encoder network to learn spectral prior, and then incorporated the deep image priors learned in pretraining phase into the optimization procedure as a regularizer. Wang et al. [32] further conducted both spectral and non-local (NLS) prior learning, and combined the model-based optimization method with the NLS-based regularization for robust HSI reconstruction. Although these unrolling methods have manifested acceptable performance for the conventional compressive sensing problem but still have insufficient spectral recovery capability.

3 Proposed lightweight hyperspectral reconstruction network

In this section, we first present the formulation problem of the measure and reconstruction phases in the coded aperture snapshot spectral imaging (CASSI) system, and then introduce our lightweight hyperspectral reconstruction model including the overview architecture and the proposed residual reconstruction module: deep feature hallucination module.

3.1 CASSI observation model

CASSI [29, 13, 1] encodes the 3D hyperspectral data of a scene into a 2D compressive snapshot image. we denote the intensity of the incident light for a spectral scene as $X(h, w, \lambda)$, where h and w are the spatial index ($1 \leqq h \leqq H$, $1 \leqq w \leqq W$) and λ is the spectral index ($1 \leqq \lambda \leqq \Lambda$). The incoming light can be collected by the objective lens, and then spatially modulated using a coded aperture, which creates a transmission

function $T(h, w)$ for the mathematical implementation. Next the modulated scene is spectrally dispersed with a wavelength-dependent dispersion function $\psi(\lambda)$ by the disperser, and a charge-coupled device (CCD) is adopted to detect the spatial and spectral coded scene as a snapshot image. The schematic concept of this measurement phase in the CASSI is shown in Fig. 1. Mathematically, the observation procedure for measuring the 2D snapshot image can be formulated as:

$$Y(h, w) = \sum T(h - \psi(\lambda))X(h - \psi(\lambda), w, \lambda). \tag{1}$$

For simplicity, we reformulate the observation model in Eq. 1 as a matrix-vector form, which is expressed as:

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{X} \tag{2}$$

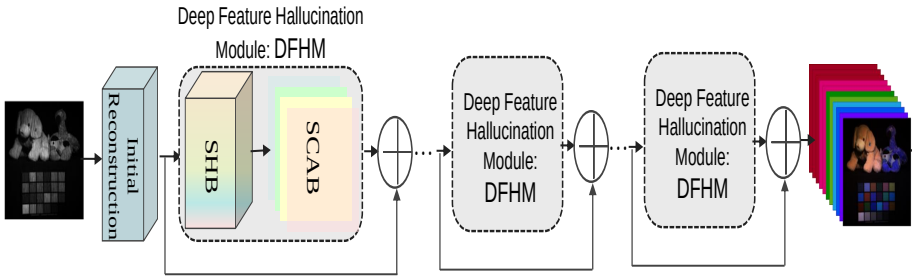
where $\mathbf{\Phi} \in \Re^{(W+\Lambda-1)H \times WH\Lambda}$ is the measurement matrix of CASSI, and is the combination operation jointly determined by $T(h, w)$ and $\psi(\lambda)$. $\mathbf{Y} \in \Re^{(W+\Lambda-1)H}$ and $\mathbf{X} \in \Re^{WH\Lambda}$ represent the vectorized expression of the compressive image and the full 3D HSI, respectively.

Give the observed compressive snapshot $\mathbf{Y}$, the goal of the HSI reconstruction in the CASSI is to recover the underlying 3D spectral image $\mathbf{X}$, which is a severe ill-posed inverse problem. The traditional model-based methods usually result in insufficient performance while the existing deep learning-based paradigms usually yield large-scale model and then restrict its wide applicability in the real HSI systems despite the promising performance. This study aims to exploit a lightweight deep reconstruction model for not only maintaining the reconstruction performance but also reducing model size and computational cost.

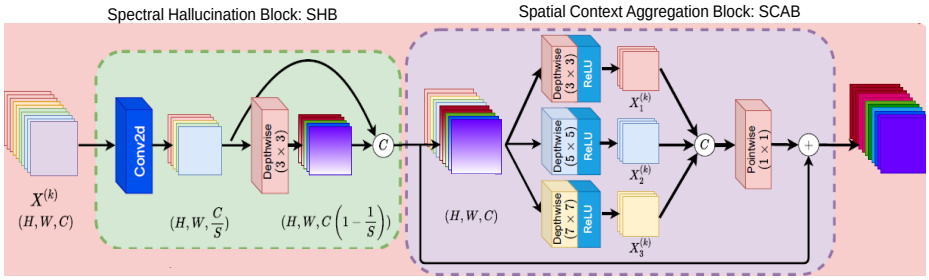
3.2 Overview of the lightweight reconstruction model

The conceptual architecture of the proposed lightweight reconstruction model (LWHRN) is illustrated in Fig. 2(a). which includes an initial reconstruction module and multiple lightweight deep feature hallucination modules (DFHM) for hierarchically reconstructing the un-recovered residual spatial and spectral components with cheaper operation than the vanilla convolution layers. The DFHM module is composed of a spectral hallucination block (SHB) for duplicating more spectral feature maps using depth-wise convolution and a spatial context aggregation block (SCAB) for exploiting the multiple contexts in various receptive fields. In order to reduce the complexity and the model parameter, we elaborately design both SHB and SCAB with more cheap operation but maintaining the amount of the learned feature maps for guaranteeing the reconstruction performance, and construct the lightweight model for practical application in real HSI systems. Moreover, we employ the residual connection structure to learn the unrecovered component in the previous module, and gradually estimate the HS image from coarse to fine.

Concretely, given the measured snapshot image $\mathbf{Y}$, the goal is to recover the full spectral image $\mathbf{X}$ using the LWHRN model. Firstly, as shown in Fig 2(a), an initial



(a) The overall architecture of the proposed lightweight reconstruction model



(b) The deep hallucination module

Fig. 2. The conceptual architecture of the proposed lightweight reconstruction model.

reconstruction module, which consist of several vanilla convolution layers, transforms the 2D compressed image $\mathbf{Y}$ into multi-channel of feature maps, and then predict an initial HSI: $\mathbf{X}_0$ with Λ spectral bands. The initial reconstruction can be formulated as:

$$\mathbf{X}^{(0)} = f_{Ini-rec}(\mathbf{Y}), \tag{3}$$

where $f_{Ini-rec}(\cdot)$ represents the overall transformation of the initial reconstruction module. In our experiments, we simply employ 3 convolution layers with kernel size 3×3, and a RELU activation layer follows after each convolution. Then, multiple deep feature hallucination modules (DFHM) with cheap operation and residual connection are stacked to form our backbone architecture, which can hierarchically predict the residual components to reconstruct the latent HSI from coarse to fine. Let $\mathbf{X}_k$ denotes the output of the $k-th$ DFHM module, the $(k+1)-th$ DFHM module with the residual connection aims to learn a more reliable reconstruction of the latent HSI, which is expressed as

$$\mathbf{X}_{k+1} = \mathbf{X}_k + f_{DFHM}(\mathbf{X}_k), \tag{4}$$

where $f_{DFHM}(\cdot)$ denotes the transformation operators in the MFHM module. The MFHM module consists of a spectral hallucination blocks and a spatial context aggregation block, which are implemented to capture sufficient channel of feature maps based on cheap depth-wise convolution operation instead of the vanilla convolution,

and is expected to reconstruct more reliable structures in both spatial and spectral directions. Moreover, we adopt the residual connection in the MFHM module to model only the un-recovered components in the previous module as shown in Fig. 2(a). Next, we would present the detail structure of our proposed DFHM module.

3.3 The DFHM Module

In the HSI reconstruction task from a snapshot image, it require to simultaneously model detail spatial structure and abundant spectral characteristics for reconstructing more plausible HSIs. It is an extreme challenging task to reliably reconstruct the high-dimensional signal in both spectral and spatial dimensions. The existing deep models generally deepen and widen the network architecture to learn large amount of feature maps for boosting the recovering performance, which unavoidably causes large-scale model size and high computational cost. Inspired by the insight that some feature maps in the well-trained networks may be obtained by employing specific transformation operations on the already learned features, we deploy the vanilla convolution layer to learn feature maps with small number of channels (reduced spectral), and then adopt the more cheap depth-wise convolution operation transforming the previously learned ones to obtain more hallucinated spectral information, dubbed as spectral hallucination block (SHB). Moreover, with the concatenated spectral reduced and hallucinated feature maps, we further conduce the depth-wise convolution with various kernel sizes to capture multi-scale spatial context, and then aggregate them as the final feature map, dubbed as spatial context aggregation block (SCAB). Since the SCAB mainly is composed of depth-wise convolution, it also can greatly decrease the parameter compared with vanilla convolution. Finally, a point-wise convolution layer is used to estimate the un-recover residual component in the previous module. To this end, we construct the deep feature hallucination module (DFHM) with a SHB and a SCAB to reciprocally hallucinate more spectral information and spatial structure with various scale of contexts, following a point-wise convolution to achieve the output. The DFHM structure is ahown in Fig. 2(b). Next, we embody the detailed description of the SHB and SCAB.

Spectral hallucination block (SHB): Given the reconstructed HSI $\mathbf{X}_k \in \mathbb{R}^{H\times W\times \Lambda}$ at the $k-th$ DFHM module, the DFHM first transforms it to a feature map with C channels: $\mathbf{X}^{(k)} \in \mathbb{R}^{H\times W\times C}$, where large number channel (spectral) should have better representative capability. The SHB aims to further learn deeper representative features with the same channel number. Instead of directly learning the deeper feature with the required spectral number, the SHB first employs a pair of vanilla convolution/RELU layer to obtain a feature map with the reduced spectral channel number $\frac{C}{S}$, and then adopts a set of linear operations on the reduced spectral feature to hallucinate more spectral features. Finally, the hallucinated spectral features by linear operations and the the spectral reduced feature have been stacked together as the final learned feature map of the SHB. Specifically, we use the depth-wise convolution, which is much cheaper operation than the vanilla convolution layer, to implement the linear operation. The mathematical formula of the SHB can be expressed as:

$$\mathbf{X}_{RSF}^{(k)} = f_{RSF}(\mathbf{X}^{(k)}), \tag{5}$$

$$\mathbf{X}_{SH}^{(k)} = f_{SH}(\mathbf{X}_{RSF}^{(k)}), \tag{6}$$

$$\mathbf{X}_{SHB}^{(k)} = Concat(\mathbf{X}_{RSF}^{(k)}, \mathbf{X}_{SH}^{(k)}) \tag{7}$$

where $\mathbf{X}_{RSF}^{(k)} \in \mathbb{R}^{H\times W\times \frac{C}{S}}$, $\mathbf{X}_{SH}^{(k)} \in \mathbb{R}^{H\times W\times \frac{C(S-1)}{S}}$ and $\mathbf{X}_{SHB}^{(k)} \in \mathbb{R}^{H\times W\times C}$ represent the spectral reduced feature, the hallucnated spectral feature, and the outputed feature map of the SHB, respectively. $f_{RSF}(\cdot)$ denotes the transformation of a vanilla convolution/RELU layer with the spatial kernel size 3×3 to reduce the spectral channel number from C to $\frac{C}{S}$ while f_{SH} is the transformation of a set of depth-wise convolution layers with the spatial kernel size 3×3.

It should noted if a vanilla convolution with kernel size $d \times d$ is employed to transform the feature map $\mathbf{X}^{(k)}$ into a deeper feature with the same spectral number, the number of the parameters by ignoring the bias term for simplicity would be $C\cdot d\cdot d\cdot C$. While the parameter number in the proposed SHB with the spectral reduced vanilla convolution and the spectral hallucinated depth-wise convolution is $C\cdot d\cdot d\times \frac{C}{S}+d\cdot d\cdot (C-\frac{C}{S})$. Thus, the compression ratio of the parameters with the SHB can be calculated as

$$r_p = \frac{C\cdot d\cdot d\cdot C}{C\cdot d\cdot d\times \frac{C}{S}+d\cdot d\cdot (C-\frac{C}{S})} \approx S \tag{8}$$

Similarly, we can obtain the theoretical speed-up ratio by replacing the vanilla convolution layer with the SHB as S. Therefore, the proposed SHB can not only learn the same amount of feature map but also greatly reduce the parameter number as wells as speed-up the computation.

Table 1. Performance comparisons on the CAVE and Harvard datasets (3% compressive ratio). The best performance is labeled in **bold**, and the second best is labeled in underline.

Method	CAVE			Harvard			Params (MB)	Flops(G)
	PSNR	SSIM	SAM	PSNR	SSIM	SAM		
TwIST	(−)	(−)	(−)	27.16	0.924	0.119	(−)	(−)
3DNSR	(−)	(−)	(−)	28.51	0.940	0.132	(−)	(−)
SSLR	(−)	(−)	(−)	29.68	**0.952**	**0.101**	(−)	(−)
HSCNN[36]	24.94	0.736	0.452	35.09	0.936	0.145	312	87
HyperReconNet [35]	25.18	0.825	0.332	35.94	0.938	0.160	581	152
λ-Net [22]	24.77	0.816	0.314	36.73	0.947	0.141	58247	12321
DeepSSPrior[31]	25.48	0.825	0.324	37.10	0.950	0.137	341	89
Our	**27.44**	**0.830**	**0.302**	**37.26**	0.951	0.133	197	49

Spatial context aggregation block (SCAB): It is known that the reliable spectral recovery of a specific pixel would greatly depend on the around spatial context, and the

required spatial range may be changed according to the physical characteristics of the pixel.The ordinary convolution networks usually carry out the context exploitation of the same spatial range for all pixels regardless to the pixel characteristic, which may yield non-optimal spectral reconstruction. To this end, we attempt to learn the feature by exploiting multiple spatial contexts under various receptive fields with the cheap depth-wise convolution operation, and adaptively aggregate them to a compact representation with a point-wise convolution, dubbed as Spatial context aggregation block (SCAB). Given the spectral hallucinated features $\mathbf{X}_{SHB}^{(k)}$ of the SHB, the SCAB firstly adopts a mixed depth-wise convolutional layer (dubbed as MixConv) [28] to adaptively exploit the spatial dependency in different sizes of local spatial regions for parameter reduction. In the detailed implementation, the spectral hallucinated feature map $\bar{\mathbf{X}}_{SHB}^{(k)} \in \mathbb{R}^{H\times W\times C}$ is partitioned into M groups: $\bar{\mathbf{X}}_{SHB}^{(k)} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_M]$ via evenly dividing the channel dimension, where $\mathbf{X}_m \in \mathbb{R}^{H\times W\times L_m}$ ($L_m = L/M$) represents the feature maps in the $m-th$ group. The MixConv layer is deployed to exploit different spatial contexts for different groups via using depth-wise convolution layers. Let's denote the parameter set of the MixConv layer as $\mathbf{\Theta}_{Mix}^{(k)} = \{\theta_1, \theta_2, \cdots, \theta_M\}$ in the M group of depth-wise convalution layers, where the parameters for different groups have various spatial kernel sizes for exploring spatial contexts in different local regions with $\theta_m \in \mathbb{R}^{s_m\times s_m\times L_m}$, the MixConv layer is formulated as:

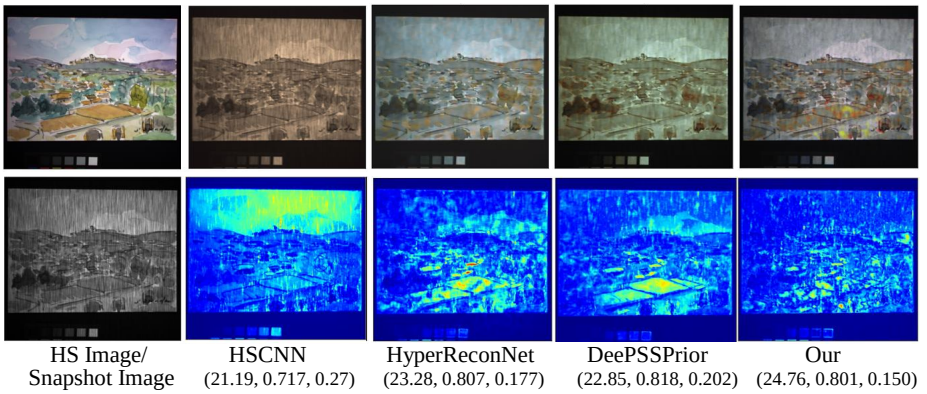
$$\mathbf{X}_{Mix}^{(k)} = Concat(f_{dp}^{\theta_1}(\mathbf{X}_1), f_{dp}^{\theta_2}(\mathbf{X}_2), \cdots, f_{dp}^{\theta_M}(\mathbf{X}_M)), \tag{9}$$

where $f_{dp}^{\theta_m}(\cdot)$ represents the depth-wise convolutional layer with the weight parameter θ_m (for simplicity, we ignore the bias parameters). With the different kernel spatial sizes at different groups, the spatial correlation in various local regions is simultaneously integrated for extracting high representative features in one layer. Moreover, we employ the depth-wise convolution operations in all groups, which can greatly reduces the parameters ($\frac{1}{L_m}$) compared with a vanilla convolution layer for being easily implanted in the real imaging systems, and expect more reliable spatial structure reconstruction via concentrating on spatial context exploration. Finally, a point-wised convolution layer is employed to estimate the residual component of the $k-th$ DFHM module, and is expressed as:

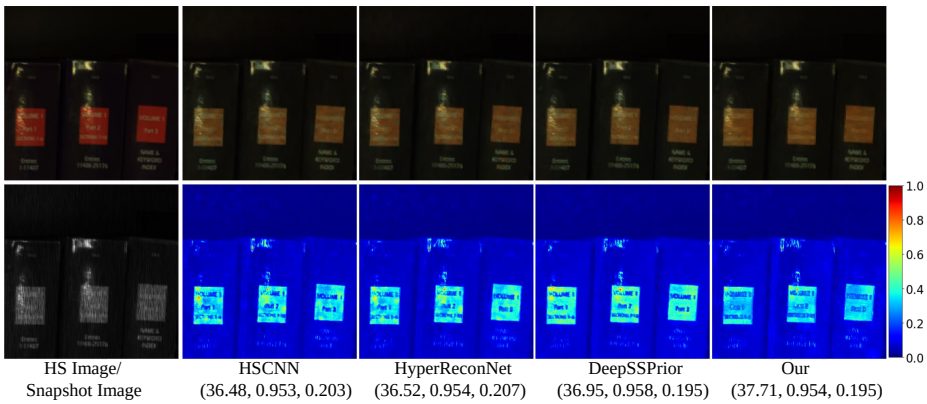
$$\bar{\mathbf{X}}_k = f_{PW}(\mathbf{X}_{Mix}^{(k)}). \tag{10}$$

4 Experimental results

To demonstrate the effectiveness of our proposed lightweight reconstruction model, we conduct comprehensive experiments on two hyperspectral datasets including the CAVE [38] dataset and the Harvard dataset [7]. The CAVE dataset consists of 32 images with spatial resolution 512 $\times$ 512 and 31 spectral channels ranging from 400nm to 700nm while the Harvard dataset is composed of 50 outdoor images captured under daylight conditions with the spatial resolution are 1040 $\times$ 1392 and the spectral wavelength



(a)



(b)

Fig. 3. Visualization results of two example images compared with the SoTA deep learning models: HSCNN [36], HyperReconNet [35], DeepSSPrior [31], and our proposed lightweight models, where the three values under the reconstruction represents the PSNR, SSIM and SAM, respectively.

ranging from 420nm to 720nm. In our experiment, we randomly select 20 images in the CAVE and 10 image in the Harvard dataset as the testing samples and the rest for training. For simulating the 2D snapshot image, we synthesize the transmission function $T(h, w)$ of the coded aperture in the HS imaging system via randomly generating a binary matrix according to a Bernoulli distribution with p = 0.5, and then create the snapshot measurements by transforming the original HSI with the synthesized transformation function. To prepare training samples, we extract the corresponding snapshot/HSI patches with spatial size of 48 $\times$ 48 from the training images. We implement our overall network by stacking 9 DFHM modules following the same number of stages in the conventional deep models: HSCNN [36] and DeepSSPrior [31] for fair comparison in model parameter and computational cost. Moreover, we quantitatively evaluate

Table 2. Ablation study on the CAVE dataset. The best performance is labeled in **bold**, and the second best is labeled in underline.

Metrics	w/o SCAB ($S = 2$)	w/o SCAB ($S = 3$)	w/o SCAB ($S = 4$)	SHB + SCAB ($S = 2$)
PSNR	26.98	26.63	26.13	27.44
SSIM	0.813	0.805	0.813	0.830
SAM	0.307	0.324	0.327	0.302
Parameter(MB)	40	35	32	42

the HS image reconstruction performance using three metrics including the peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and spectral angle mapper (SAM).

Comparison with the SoTA methods: We compare our proposed method with several state-of-the-art HSI reconstruction methods, including three traditional methods with hand-crafted prior modeling , *i.e*, TwIST with TV prior [40], 3DNSR and SSLR with NLS prior [30, 11], and four deep learning-based methods, *i.e*., HSCNN [36], HyperReconNet [35], λ-net [22] and Deep Spatial Spectral Prior (DeepSSPrior) [31]. Our lightweight model was implemented by stacking 9 MFHM modules with the compression ratio : $S = 2$ (in the SHB). The compared quantitative results are illustrated in Table 1, which verifies that our proposed lightweight models can not only achieve promising reconstruction performance but also greatly reduce the parameter as well as computational cost. Moreover, we also provide the compared visualization results of our lightweight model with the HSCNN[36], HyperReconNet [35] and DeepSSPrior [31] in Fig. 3, which also demonstrated the better reconstruction performance by our method.

Ablation study: As introduced in Section 2, we compress the spectral channel from C to $\frac{C}{S}$, and then hallucinate more spectral features using cheap depth-wise convolution operation in the SHB, where the hyper-parameter S can be adjusted according to the compression ratio of the parameter in the SHB. What is more, the following SCAB is incorporated for exploiting multi-scale spatial contexts with cheap operation, which can be plugged in or ignored in the DFHM module. To verify the effect of the compression ratio S and the additional SCAB, we carried out experiments by varying S from 2 to 4, and w/o the incorporation of the SCAB on the CAVE dataset. The ablation results are shown in Table 2. From Table 2, we observe that the compressive ratio 2 achieves the best performance while increasing the compression ratio will yield a little performance drop but with smaller parameter number. Moreover, the incorporation of the SCAB can further boost the reconstruction performance, whilst causes few parameter raising.

5 Conclusions

This study proposed a novel lightweight model for efficiently and effectively reconstruct a full hyperspectral image from a compressive snapshot measurement. Although the existing deep learning based models have achieved remarkable performance improvement compared with the traditional model-based methods for hyperspectral image reconstruction, it still confronts the difficulties to embed the deep models in real HSI

systems due to large-scale model size. To this end, we exploited an efficient deep feature hallucination module (DFHM) to construct our lightweight models. Specifically, we elaborated the DFHM by a vanilla convolution-based spectral reduced layer and a depth-wise convolution-based spectral hallucination layer to learn sufficient feature maps with cheap operation. Moreover, we further incorporated a spatial context aggregation block to exploit multi-scale context in various receptive fields for boosting reconstruction performance. Experiments on two datasets demonstrated that our proposed method achieved superior performance over the SoTA models as well as greatly reduced the parameters and computational cost.

Acknowledgements This work was supported in part by MEXT under the Grant No. 20K11867, and JSPS KAKENHI Grant Number JP12345678.

References

1. Arce, G., Brady, D., Carin, L., Arguello, H., Kittle, D.: Compressive coded aperture spectral imaging: An introduction. IEEE Signal Processing Magazine **31**, 105–115 (2014)
2. Bioucas-Dias, J., Figueiredo, M.A.T.: A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. IEEE Transactions on Image Processing **16**, 2992–3004 (2007)
3. Bjorgan, A., Randeberg, L.L.: Towards real-time medical diagnostics using hyperspectral imaging technology (2015)
4. Borengasser, M., Hungate, W.S., Watkins, R.: Hyperspectral remote sensing: Principles and applications (2007)
5. Cao, X., Du, H., Tong, X., Dai, Q., Lin, S.: A prism-mask system for multispectral video acquisition. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**, 2423–2435 (2011)
6. Cao, X., Yue, T., Lin, X., Lin, S., Yuan, X., Dai, Q., Carin, L., Brady, D.: Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. IEEE Signal Processing Magazine **33**, 95–108 (2016)
7. Chakrabarti, A., Zickler, T.E.: Statistics of real-world hyperspectral images. CVPR 2011 pp. 193–200 (2011)
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**, 1 – 13 (2017)
9. Cui, Q., Park, J., Smith, R.T., Gao, L.: Snapshot hyperspectral light field imaging using image mapping spectrometry. Optics letters **45 3**, 772–775 (2020)
10. Devassy, B.M., George, S.: Forensic analysis of beverage stains using hyperspectral imaging. Scientific reports **11**, 6512 (2021)
11. Fu, Y., Zheng, Y., Sato, I., Sato, Y.: Exploiting spectral-spatial correlation for coded hyperspectral image restoration. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 3727–3736 (2016)
12. Gao, L., Kester, R., Hagen, N., Tkaczyk, T.: Snapshot image mapping spectrometer (ims) with high sampling density for hyperspectral microscopy. Optics Express **18**, 14330–14344 (2010)
13. Gehm, M., John, R., Brady, D., Willett, R., Schulz, T.: Single-shot compressive spectral imaging with a dual-disperser architecture. Optics express **1521**, 14013–27 (2007)
14. Goetz, A., Vane, G., Solomon, J., Rock, B.: Imaging spectrometry for earth remote sensing. Science **228**, 1147 – 1153 (1985)

15. Gregor, K., LeCun, Y.: Learning fast approximations of sparse coding (2010)
16. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C.: Ghostnet: More features from cheap operations
17. Kittle, D.S., Choi, K., Wagadarikar, A.A., Brady, D.J.: Multiframe image estimation for coded aperture snapshot spectral imagers. Applied optics **49 36**, 6824–33 (2010)
18. Lin, X., Liu, Y., Wu, J., Dai, Q.: Spatial-spectral encoded compressive hyperspectral imaging. ACM Trans. Graph. **33**, 233:1–233:11 (2014)
19. Liu, Y., Yuan, X., Suo, J.L., Brady, D., Dai, Q.: Rank minimization for snapshot compressive imaging. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**, 2990–3006 (2019)
20. Lu, G., Fei, B.: Medical hyperspectral imaging: a review. Journal of Biomedical Optics **19** (2014)
21. Ma, J., Liu, X.Y., Shou, Z., Yuan, X.: Deep tensor admm-net for snapshot compressive imaging. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 10222–10231 (2019)
22. Miao, X., Yuan, X., Pu, Y., Athitsos, V.: lambda-net: Reconstruct hyperspectral images from a snapshot measurement. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 4058–4068 (2019)
23. Nguyen, H.V., Banerjee, A., Chellappa, R.: Tracking via object reflectance using a hyperspectral video camera. IEEE Computer Vision and Pattern Recognition Workshops pp. 44–51 (2010)
24. Pan, Z., Healey, G., Prasad, M., Tromberg, B.: Face recognition in hyperspectral images. IEEE Transactions Pattern Analysis and Machine Intelligence **25**, 1552–1560 (2003)
25. Rajwade, A., Kittle, D.S., Tsai, T.H., Brady, D.J., Carin, L.: Coded hyperspectral imaging and blind compressive sensing. SIAM J. Imaging Sciences **6**, 782–812 (2013)
26. Schechner, Y., Nayar, S.: Generalized mosaicing: Wide field of view multispectral imaging. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 1334–1348 (2002)
27. Tan, J., Ma, Y., Rueda, H., Baron, D., Arce, G.: Compressive hyperspectral imaging via approximate message passing. IEEE Journal of Selected Topics in Signal Processing **10**, 389–401 (2016)
28. Tan, M., Le, Q.V.: Mixconv: Mixed depthwise convolutional kernels (2019)
29. Wagadarikar, A.A., John, R., Willett, R.M., Brady, D.J.: Single disperser design for coded aperture snapshot spectral imaging. Applied optics **4710**, B44–51 (2008)
30. Wang, L., Xiong, Z., Shi, G., Wu, F., Zeng, W.: Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**, 2104–2111 (2017)
31. Wang, L., Sun, C., Fu, Y., Kim, M.H., Huang, H.: Hyperspectral image reconstruction using a deep spatial-spectral prior. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 8024–8033 (2019)
32. Wang, L., Sun, C., Zhang, M., Fu, Y., Huang, H.: Dnu: Deep non-local unrolling for computational spectral imaging. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1658–1668 (2020)
33. Wang, L., Xiong, Z., Gao, D., Shi, G., Wu, F.J.: Dual-camera design for coded aperture snapshot spectral imaging. Applied optics **54 4**, 848–58 (2015)
34. Wang, L., Xiong, Z., Gao, D., Shi, G., Zeng, W., Wu, F.: High-speed hyperspectral video acquisition with a dual-camera architecture. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 4942–4950 (2015)
35. Wang, L., Zhang, T., Fu, Y., Huang, H.: Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. IEEE Transactions on Image Processing **28**, 2257–2270 (2019)

36. Xiong, Z., Shi, Z., Li, H., Wang, L., Liu, D., Wu, F.: Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. 2017 IEEE International Conference on Computer Vision Workshops (ICCVW) pp. 518–525 (2017)
37. Yang, Y., Sun, J., Li, H., Xu, Z.: Admm-csnet: A deep learning approach for image compressive sensing. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**, 521–538 (2020)
38. Yasuma, F., Mitsunaga, T., Iso, D., Nayar, S.: Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. IEEE Transactions on Image Processing **19**, 2241–2253 (2010)
39. Yuan, X., Tsai, T., Zhu, R., Llull, P., Brady, D., Carin, L.: Compressive hyperspectral imaging with side information. IEEE Journal of Selected Topics in Signal Processing **9**, 964–976 (2015)
40. Yuan, X.: Generalized alternating projection based total variation minimization for compressive sensing. 2016 IEEE International Conference on Image Processing (ICIP) pp. 2539–2543 (2016)
41. Zhang, J., Ghanem, B.: Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 1828–1837 (2018)
42. Zhang, S., Wang, L., Fu, Y., Zhong, X., Huang, H.: Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 10182–10191 (2019)
43. Zhang, X., Lian, Q., Yang, Y.C., Su, Y.: A deep unrolling network inspired by total variation for compressed sensing mri. Digital Signal Processing **107**, 102856 (2020)
44. Zhou, S., He, Y., Liu, Y., Li, C.: Multi-channel deep networks for block-based image compressive sensing. ArXiv **abs/1908.11221** (2019)