HIGH-QUALITY SELF-SUPERVISED SNAPSHOT HYPERSPECTRAL IMAGING

Yuhui Quan Xinran Qin Mingqin Chen Yan Huang

School of Computer Science and Engineering, South China University of Technology, China

ABSTRACT

Hyperspectral image (HSI) reconstruction is about recovering a 3D HSI from its 2D snapshot measurements, to which deep models have become a promising approach. However, most existing studies train deep models on large amounts of organized data, the collection of which can be difficult in many applications. This paper leverages the image priors encoded in untrained neural networks (NNs) to have a self-supervised learning method which is free from training datasets while adaptive to the statistics of a test sample. To induce better image priors and prevent the NN overfitting undesired solutions, we construct an unrolling-based NN equipped with fractional max pooling (FMP). Furthermore, the FMP is used with randomness to enable self-ensemble for reconstruction accuracy improvement. In the experiments, our self-supervised learning approach enjoys high-quality reconstruction and outperforms recent methods including the supervised ones.

Index Terms— Hyperspectral imaging, Self-Supervised learning, Image reconstruction, Untrained network priors

1. INTRODUCTION

A hyperspectral image (HSI) is a 3D image containing many spectral bands over the same spatial area, which provides rich spectral information for identifying specific materials, with applications ranging from computer vision to remote sensing. Coded aperture snapshot spectral imaging (CASSI) [1] is one efficient approach to collecting HSIs. It encodes the 3D HSI into a 2D snapshot via mixing different wavelength signals modulated by a physical mask and a disperser. Then, the 3D HSI is reconstructed from the 2D compressive snapshot.

Let $\boldsymbol{X} \in \mathbb{R}^{M\times N\times\Lambda}$ denote an HSI with its element denoted by $\boldsymbol{X}(m,n,\lambda)$, where m,n are the spatial indices and λ is the spectral index. The generation of the snapshot $\boldsymbol{Y} \in \mathbb{R}^{(M+\Lambda-1)\times N}$ from CASSI can be formulated as [2]:

$$\boldsymbol{Y}(m,n)=\sum_{\lambda=1}^{\Lambda}\rho(\lambda)\varphi(m-\lambda,n)\boldsymbol{X}(m-\lambda,n,\lambda)+\boldsymbol{N}(m,n), \tag{1}$$

where $\rho(\cdot)$ is the spectral response of the camera, $\varphi(\cdot,\cdot)$ represents the spatial coded aperture pattern, and $\boldsymbol{N}$ is the measurement noise. We can rewrite (1) in a matrix-vector form:

$$\boldsymbol{y}=\boldsymbol{\Phi}\boldsymbol{x}+\boldsymbol{n}, \tag{2}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{N(M+\Lambda-1)\times MN\Lambda}$ is the measurement matrix determined by $\rho(\lambda)$, and $\boldsymbol{x},\boldsymbol{y},\boldsymbol{n}$ are the vectorized form of $\boldsymbol{X},\boldsymbol{Y},\boldsymbol{N}$, respectively. It is a challenging ill-posed problem, due to the significant compression caused by $\boldsymbol{\Phi}$ and the existence of $\boldsymbol{n}$. One approach to resolving the ill-posedness is designing shallow models with handcrafted priors, *e.g.*, smoothness [3], sparsity [4], low-rank structure [5] and self-similarity [6]. However, their performance is not satisfactory.

In recent years, deep models have promised effective solutions to HSI reconstruction. Most existing works use an external dataset containing ground truths to supervisedly train deep neural networks (NNs) for reconstruction. The trained NN encodes data-adaptive priors of HSIs which lead to noticeable improvement over shallow models. A few studies (*e.g.* [7–10]) fit a direct mapping from snapshots to HSIs using an NN, but without making full use of the imaging model (1) to avoid overfitting of the NN. Deep unrolling is a promising alternate which unrolls the iterative process of some shallow model and replaces the intermediate steps related to image priors with denoising NNs; see *e.g.* [2, 11–13]. In addition to the difference in the process selected to unroll, existing methods mainly differ in the implementation of image priors used in the unrolled process, *e.g.*, deep spatial-spectral prior [11], non-local prior [12], tensor low-rank prior [2], and Gaussian scale mixture prior [13], leading to various NN structures,.

However, the requirement on a large number of HSIs for training lowers the practicability of these supervised learning-based methods, as the cost of collecting HSIs is usually high. In addition, as the spectral response characteristics and spectral wavelength range vary among spectral imaging devices, the NN trained upon the HSI dataset captured by specific devices may fail to perform well on the HSIs captured by a different device. There are a few plug-and-play methods [14, 15] using the denoising NNs pre-trained on natural images in the unrolled iterative process. While these methods avoid using HSIs for training, their performance may be limited due to the domain gap between natural images and HSIs.

This paper proposes a self-supervised learning approach to mitigate the issues arising from supervised learning, which

This work was supported in part by National Natural Science Foundation of China under Grants 61872151 and 61902130, and in part by Natural Science Foundation of Guangdong Province under Grant 2020A1515011128.
Corresponding author: Yan Huang (aihuangy@scut.edu.cn)

exploits the deep image prior (DIP) [16] encoded in an untrained deep convolutional NN (DCNN) so that no external data but only the snapshot $\boldsymbol{y}$ itself is required for NN training. We are motivated by the facts that HSIs often have strong internal self-similarities [5, 6, 13] and these properties can be well captured by untrained-DCNN-based image priors [17]. Let $\mathcal{G}$ denote an untrained DCNN. Based on DIP [16], the self-supervised learning can be done via the internal learning:

$$\min_{\mathcal{G}} \|\boldsymbol{y} - \boldsymbol{\Phi}\mathcal{G}(\boldsymbol{\epsilon})\|, \tag{3}$$

with a random seed $\boldsymbol{\epsilon}$ where $\mathcal{G}$ acts as an HSI generator. Nevertheless, in such an approach, the DCNN is easy to overfit even with early stopping and temporal averaging [16], leading to unsatisfactory performance in practice.A

The proposed approach distinguishes itself from standard untrained-NN-based methods (*e.g.* [18]) mainly in two aspects. First, we introduce fractional max pooling [19] to the untrained DCNN. It not only acts as an implicit regularizer, but also allows the test-time self-ensemble that combines multiple predictions from the NN to improve the reconstruction accuracy, similar to [20]. Second, we define $\mathcal{G}$ as a deep-unrolling-based DCNN with the snapshot $\boldsymbol{y}$ as input. The replacement of random seed $\boldsymbol{\epsilon}$ with the snapshot $\boldsymbol{y}$ as input is nontrivial, as it changes the role of $\mathcal{G}$ from a pure generator which is empirically designed without considering $\boldsymbol{\Phi}$, to a reconstructive NN where the imaging model (1) can be used for inducing structural regularization, *e.g.*, using deep unrolling. These two differences lead to significant performance improvement.

There are few studies on self-supervised deep learning for CASSI-based HSI reconstruction. Sun *et al.* [18] exploited DIP by a spatial-spectral attentive conditional DCNN with early stopping. They combine the snapshot with a random seed as input; however, its NN is still defined as an empirical generator without incorporating imaging models. Meng *et al.* [21] proposed to integrate DIP into the unrolled optimization process, not defining an unrolled NN for DIP as what our approach does. The improvement of our approach over DIP-based methods has been discussed above. Another two related works are [22, 23] for general compressive sensing reconstruction problems. Particularly, Pang *et al.* [23] uses a Bayesian DCNN that introduces randomness to the kernel weights in a DCNN for reducing the overfitting of DIP, while we introduce randomness to max pooling. In experiments, our method outperforms all aforementioned methods.

To conclude, our contributions are as follows:

- We proposed a self-supervised learning approach for HSI reconstruction, which is free from external training data and pre-trained models while adaptive to each test sample.
- We introduced model ensemble driven by FMP for image reconstruction and utilized an unrolling-based NN for untrained-NN-based image priors to handle the overfitting in the self-supervised learning of linear inverse problems.
- The proposed self-supervised learning approach showed superior performance over existing methods including the supervised ones, even without utilizing training datasets.

2. METHODOLOGY

Let $\mathcal{G}_{\boldsymbol{\Phi}} : \mathbb{R}^{N(M+\Lambda-1)} \to \mathbb{R}^{MN\Lambda}$ denote an untrained DCNN that reconstructs the latent HSI $\boldsymbol{x}$ from the snapshot $\boldsymbol{y}$, given the snapshot process $\boldsymbol{\Phi}$. We aim to learn $\mathcal{G}_{\boldsymbol{\Phi}}$ from the snapshot itself by exploiting the internal statistics of $\boldsymbol{x}$. It is done by the following self-supervised training loss:

$$\min_{\mathcal{G}_{\boldsymbol{\Phi}}} \|\boldsymbol{\Phi}\mathcal{G}_{\boldsymbol{\Phi}}(\boldsymbol{y}) - \boldsymbol{y}\|_2^2. \tag{4}$$

Since $\mathcal{G}_{\boldsymbol{\Phi}}$ may overfit undesired solutions, we use two strategies to improve the prediction accuracy from $\mathcal{G}$: (a) fractional max pooling (FMP) based ensemble learning for both implicit regularization in training and model averaging in inference; and (b) an unrolling-based DCNN that incorporates the imaging model (1) into its architecture for implicit regularizations.

2.1. Unrolling-Based Network Architecture

Recall that in existing untrained-NN-based learning methods, the NN is used as a generator that focuses on how to map a random seed to an estimation of the latent HSI. In other words, $\mathcal{G}$ acts as an image model that incorporates the priors on their intermediate layers. Usually, such a generator has to be designed by experience without making full use of the knowledge from the imaging model, *e.g.* $\boldsymbol{\Phi}$. In the proposed approach, the input is set to the snapshot $\boldsymbol{y}$. Then the NN indeed acts as a reconstructive NN that tries to recover $\boldsymbol{x}$ from $\boldsymbol{y}$. This allows us to exploit the imaging model (1) to build up a more effective NN. Bearing this in mind, we adopt the deep unrolling approach to design $\mathcal{G}$. Note that while deep unrolling is popular for designing supervised models, it is not the case for designing untrained NNs for internal learning.

The unrolling is done using half-quadratic splitting on the following regularization model:

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda\mathcal{R}(\boldsymbol{x}), \tag{5}$$

where $\lambda \in \mathbb{R}$ is a weight, and $\mathcal{R}(\cdot)$ denotes some NN-based image prior. By introducing an auxiliary variable $\boldsymbol{h}$ and an auxiliary parameter μ, the problem (5) can be written as

$$\min_{\boldsymbol{x},\boldsymbol{h}} \ \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \mu\|\boldsymbol{x} - \boldsymbol{h}\|_2^2 + \lambda\mathcal{R}(\boldsymbol{h}). \tag{6}$$

The $\boldsymbol{x}$-subproblem is a solved by the numerical scheme used in [2, 12] which approximates the explicit solution:

$$\boldsymbol{x}^k = (\boldsymbol{\Phi}^\top\boldsymbol{\Phi} + \mu\boldsymbol{I})^{-1}(\boldsymbol{\Phi}^\top\boldsymbol{y} + \mu\boldsymbol{h}^{k-1}), \tag{7}$$

where $\boldsymbol{I}$ is an identity matrix with desired dimensions, and we define $\boldsymbol{x}^0 = \boldsymbol{\Phi}^\top\boldsymbol{y}$. The $\boldsymbol{h}$-subproblem at each iteration is solved by an individual DCNN acting artifact eliminator. We define the DCNNs as standard U-Nets with FMP and without weight sharing. See Fig. 1 for more details on the whole $\mathcal{G}$.

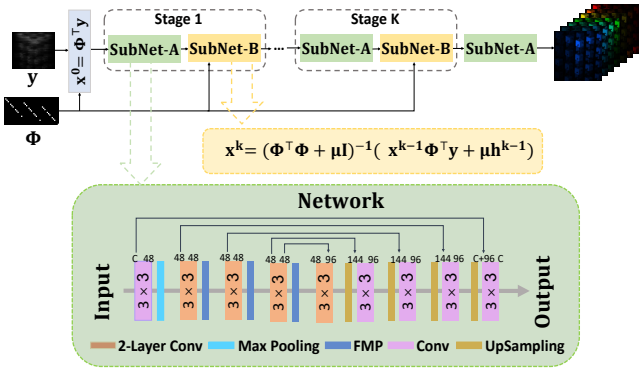

Fig. 1. Architecture of the NN used in proposed method.

2.2. FMP-Based Ensemble

Our DCNN is equipped with FMP instead of the standard max pooling. Taking a $M_{\text{in}} \times N_{\text{in}}$ feature map as input, FMP returns a smaller one with size $M_{\text{out}} \times N_{\text{out}}$. Let $\{m_i\}_{i=1}^{M_{\text{out}}}$ and $\{n_i\}_{i=1}^{N_{\text{out}}}$ be two ascending integer sequences starting at 1 and ending at M_{out} and N_{out} respectively, with increment less than 3. The pooling regions in FMP can be disjoint, denoted by $P_{i,j} = [m_{i-1}, m_i - 1] \times [n_{j-1}, n_j - 1]$ or overlapping, denoted by $P_{i,j} = [m_{i-1}, m_i] \times [n_{j-1}, n_j], \forall i, j$. Each pooling region on an FMP layer is generated with a pseudo random manner for stability [19]: $m_i, n_i = \text{Ceil}(\alpha(i + u))$, where $\alpha \in (1, 2)$ and $u \in (0, 1)$. See [19] for more details. The FMP in our DCNN is used not for fractional-ratio downscaling, but for introducing regularization and uncertainty. Our DCNN downscales feature maps by half using consecutive FMP layers for higher randomness.

The randomness from FMP has implicit regularization effects on the NN during training. In addition, FMP with randomness can also introduce uncertainty to the trained model. Indeed, an NN using FMP with above random pooling region generation can be thought of as a set of similar NNs with different max pooling regions. In other words, training such an NN simulates training multiple models simultaneously. Then, we can also have multiple models in test by turning random FMP on. As a result, we can have many hypotheses on $\boldsymbol{x}$ with a certain degree of diversity, denoted by $\mathcal{G}_{\boldsymbol{\Phi}}^1, \cdots, \mathcal{G}_{\boldsymbol{\Phi}}^K$, which are aggregated for the final reconstruction:

$$\boldsymbol{x}^{\text{est}} = \frac{1}{K} \sum_{k=1}^{K} \mathcal{G}_{\boldsymbol{\Phi}}^k(\boldsymbol{y}). \tag{8}$$

Such an ensemble scheme is very useful. Recall that the prediction of an NN can be measured by the mean squared error decomposed into bias and variance. In self-supervised internal learning, the bias is not the concern as the DCNN is often over-parameterized even with a specific design. However, the prediction variance can be very high as only one sample is used for training the NN. Such variance can be effectively reduced by the proposed FMP-based ensemble.

3. EXPERIMENTS

The proposed method is implemented as follows. On all convolutional layers, kernel size is set to 3×3 and both the stride and padding number are set to 1. The training is done by Adam with learning rate of 10^{-4} and maximal epoch number of 1×10^5. Ten times inferences are used for self-ensemble.

3.1. Performance Evaluation

Following [13], we extract 10 scenes of spatial size 256×256 from the KAIST dataset [24] for testing and employ a measurement system from [10, 13] for synthesis. To be consistent with the wavelengths of real systems [10], the wavelengths of the training and testing data are unified by spectral interpolation to have 28 spectral bands ranging from 450nm to 650nm. The methods selected for comparison include (a) Supervised learning-based ones: HSSP [11], DNU [12], TSA-Net [10] and DGSM [13]; and (b) Untrained NN-based ones: BCNN [23], HCS^2-Net [18] and PnP-DIP [21]. For BCNN, we use its GitHub code to obtain the results. For HCS^2-Net, it has neither codes nor results published, and we implement it with the default setting suggested in [18]. For other compared methods, we quote the results from existing literature. On the whole dataset, our method takes 7.5h and 14.7s for the learning and prediction respectively, with an RTX 2080Ti GPU. In comparison, PnP-DIP takes 10.2h and 12.7min for learning and prediction respectively.

Table 1 compares the reconstruction results of these methods quantitatively. Our method is the best performer overall, and it outperforms other compared methods noticeably in 8 out of 10 scenarios, with performance gain larger than 1dB and even up to 3dB. The improvement of our method over BCNN, HCS^2-Net and PnP-DIP has demonstrated the effectiveness of the proposed techniques on resolving the overfitting in self-supervised internal learning. What surprised us is, our method also performs competitively with the supervisedly-trained models and even shows superior performance over them in many scenarios and on average. See also Fig. 2 for a visual comparison on reconstructed spectral frames and the spectral curves. Our method not only achieved high visual quality in reconstruction, but also shows a higher correlation with the ground-truth spectra than other methods.

We also evaluate on two real snapshots of size 660×714 from [10, 13], which are captured by a real CASSI system with 28 wavelengths ranging from 450nm to 650nm and with 54-pixel dispersion in the column dimension. See Fig. 3 for a visual comparison to PnP-DIP, where the proposed achieved higher reconstruction quality.

3.2. Ablation Study

See Table 2 for the results of ablation studies. First, we replace all FMP layers with the standard max pooling and disable the test-time ensemble. The results become much worse,

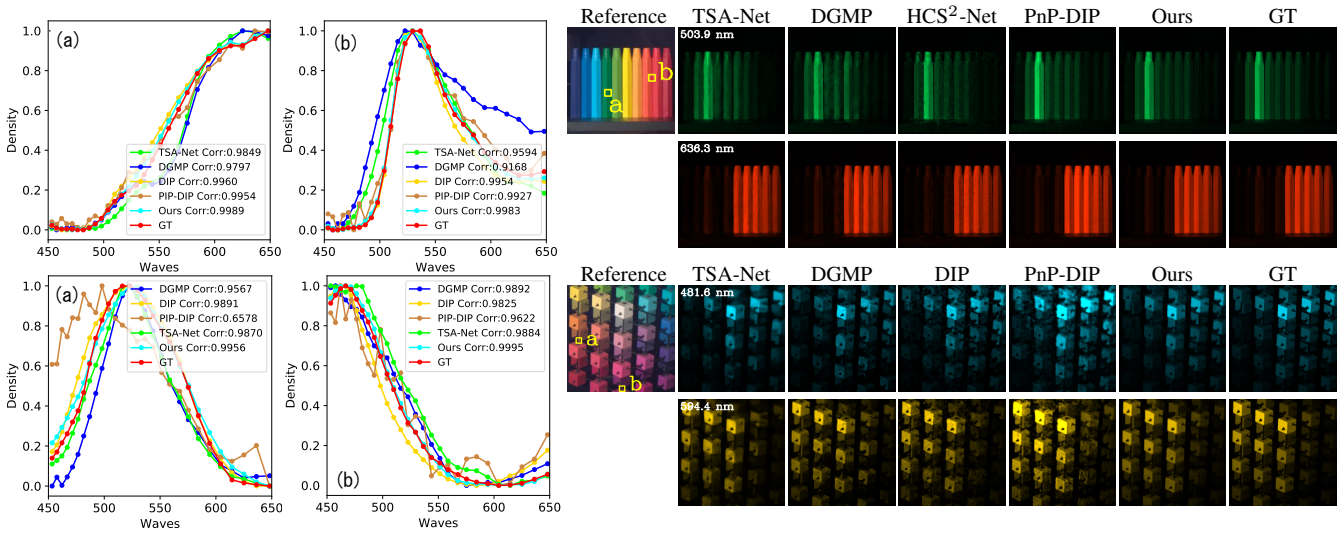

Fig. 2. Visual comparison. Left: Recovered spectra of regions a and b. Right: Reconstruction on 2/28 spectral channels.

Methods	Scene1	Scene2	Scene3	Scene4	Scene5	Scene6	Scene7	Scene8	Scene9	Scene10	Average
HSSP [11]	31.48/.86	31.09/.84	28.96/.86	34.56/.90	28.53/.81	30.83/.88	28.71/.82	30.09/.88	30.43/.87	28.78/.84	30.35/.85
DNU [12]	31.72/.86	31.13/.85	29.99/.84	35.34/.91	29.03/.83	30.87/.89	28.99/.84	30.13/.88	31.03/.88	29.14/.85	30.74/.86
TSA-Net [10]	32.03/.89	31.00/.86	32.25/.91	39.19/.95	29.39/.88	31.44/.91	30.32/.88	29.35/.89	30.01/.89	29.59/.87	31.46/.89
DGSM [13]	33.26/.92	32.09/.90	33.06/.93	40.54/.96	28.86/.88	**33.08/.94**	30.74/.89	**31.55/.92**	31.66/.91	31.44/**.92**	32.63/.92
BCNN [23]	32.12/.87	30.06/.79	32.75/.92	42.72/.97	29.07/.84	28.67/.86	34.27/.95	26.88/.85	37.03/.95	26.97/.75	32.05/.88
HCS^2-Net	33.24/.89	33.20/.89	35.26/.93	42.31/.97	30.12/.87	32.45/.90	30.23/.87	32.99/.91	37.15/.94	29.04/.85	33.60/.90
PnP-DIP [21]	32.80/.89	33.37/.89	35.77/.93	43.13/.98	30.83/.87	32.00/.91	33.78/.92	30.48/.87	37.68/.94	28.90/.85	33.91/.91
Ours	**36.04/.93**	**36.07/.93**	**37.69/.97**	**44.91/.99**	**34.89/.94**	32.24/.93	**37.01/.96**	31.22/**.92**	**38.86/.96**	**31.67/.92**	**36.05/.95**

Table 1. Average PSNR (dB) and SSIM values of compared methods on KAIST dataset.

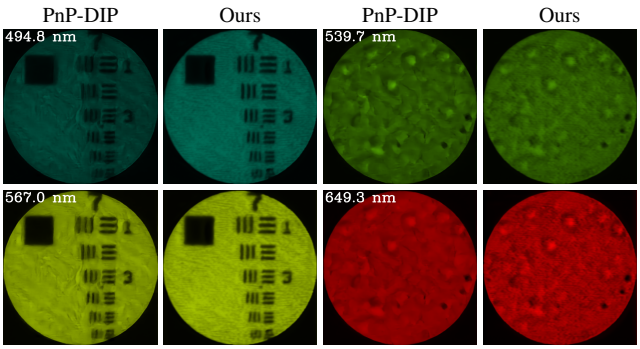

Fig. 3. Visual reconstruction results on real data.

Ours	w/o Ensemble	U-Net	Rand. Seed	w/o All	HCS^2-Net
36.05/.95	35.03/.93	35.16/.93	35.56/.94	33.22/.89	33.60/.90

Table 2. PSNR/SSIM in ablation study on KAIST dataset.

which demonstrates the effectiveness of FMP-based ensemble for self-supervised learning. Second, we replace our whole NN with the U-Net used in [16] with its max pooling layers set to FMP and its channel number enlarged to have a similar size our NN. The results become noticeably worse, which demonstrates the benefits of using unrolling-based NN to exploit the imaging model for alleviating overfitting. Next, we replace our NN's input $\boldsymbol{y}$ with a random seed as [16, 23] did. In this case, the NN acts as a generator with degenerated performance. Last, removing all modifications proposed in our method makes the results significantly worse. To conclude, each proposed component in our approach is necessary.

4. CONCLUSION

We proposed an effective self-supervised internal learning approach for snapshot-based HSI reconstruction, which does not require external data for deep learning and can adapt to each test sample. The proposed method trains an unrolling-based DCNN with FMP layers to match the snapshot measurement process. Both the unrolling-based NN architecture and FMP have implicit regularization effects on the learning process. More importantly, the FMP with randomness enables to apply model ensemble to significantly improve the reconstruction accuracy. The experiments have demonstrated the effectiveness of the proposed method.

5. REFERENCES

[1] Ashwin A Wagadarikar, Nikos P Pitsianis, Xiaobai Sun, and David J Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," *Opt. Lett.*, vol. 17, no. 8, pp. 6368–6388, 2009.

[2] Shipeng Zhang, Lizhi Wang, Lei Zhang, and Hua Huang, "Learning tensor low-rank prior for hyperspectral image reconstruction," in *Proc. CVPR*, 2021, pp. 12006–12015.

[3] Xin Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *Proc. ICIP*. IEEE, 2016, pp. 2539–2543.

[4] Xin Yuan, Tsung-Han Tsai, Ruoyu Zhu, Patrick Llull, David Brady, and Lawrence Carin, "Compressive hyperspectral imaging with side information," *IEEE J Sel Top Signal Process*, vol. 9, no. 6, pp. 964–976, 2015.

[5] Shipeng Zhang, Lizhi Wang, Ying Fu, Xiaoming Zhong, and Hua Huang, "Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery," in *Proc. CVPR*, 2019, pp. 10183–10192.

[6] Lizhi Wang, Zhiwei Xiong, Guangming Shi, Feng Wu, and Wenjun Zeng, "Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 10, pp. 2104–2111, 2016.

[7] Xin Miao, Xin Yuan, Yunchen Pu, and Vassilis Athitsos, "l-net: Reconstruct hyperspectral images from a snapshot measurement," in *Proc. CVPR*, 2019, pp. 4059–4069.

[8] Lizhi Wang, Tao Zhang, Ying Fu, and Hua Huang, "Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2257–2270, 2018.

[9] Tao Zhang, Ying Fu, Lizhi Wang, and Hua Huang, "Hyperspectral image reconstruction using deep external and internal learning," in *Proc. CVPR*, 2019, pp. 8559–8568.

[10] Ziyi Meng, Jiawei Ma, and Xin Yuan, "End-to-end low cost compressive spectral imaging with spatial-spectral self-attention," in *Proc. ECCV*. Springer, 2020, pp. 187–204.

[11] Lizhi Wang, Chen Sun, Ying Fu, Min H Kim, and Hua Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *Proc. CVPR*, 2019, pp. 8032–8041.

[12] Lizhi Wang, Chen Sun, Maoqing Zhang, Ying Fu, and Hua Huang, "Dnu: deep non-local unrolling for computational spectral imaging," in *Proc. CVPR*, 2020, pp. 1661–1671.

[13] Tao Huang, Weisheng Dong, Xin Yuan, Jinjian Wu, and Guangming Shi, "Deep gaussian scale mixture prior for spectral compressive imaging," in *Proc. CVPR*, 2021, pp. 16216–16225.

[14] Ernest Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. ICML*. PMLR, 2019, pp. 5546–5557.

[15] Haiquan Qiu, Yao Wang, and Deyu Meng, "Effective snapshot compressive-spectral imaging via deep denoising and total variation priors," in *Proc. CVPR*, 2021, pp. 9127–9136.

[16] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Deep image prior," in *Proc. CVPR*, 2018, pp. 9446–9454.

[17] Yosef Gandelsman, Assaf Shocher, and Michal Irani, "Double-dip: Unsupervised image decomposition via coupled deep-image-priors," in *Proc. CVPR*, 2019, pp. 11026–11035.

[18] Yubao Sun, Ying Yang, Qingshan Liu, and Mohan Kankanhalli, "Unsupervised spatial-spectral network learning for hyperspectral compressive snapshot reconstruction," *IEEE Trans Geosci Remote Sens*, 2021.

[19] Benjamin Graham, "Fractional max-pooling," *arXiv preprint arXiv:1412.6071*, 2014.

[20] Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji, "Self2self with dropout: Learning self-supervised denoising from single image," in *Proc. CVPR*, 2020, pp. 1890–1898.

[21] Ziyi Meng, Zhenming Yu, Kun Xu, and Xin Yuan, "Self-supervised neural networks for spectral snapshot compressive imaging," *Proc. CVPR*, 2021.

[22] Dave Van Veen, Ajil Jalal, Mahdi Soltanolkotabi, Eric Price, Sriram Vishwanath, and Alexandros G Dimakis, "Compressed sensing with deep image prior and learned regularization," *arXiv preprint arXiv:1806.06438*, 2018.

[23] Tongyao Pang, Yuhui Quan, and Hui Ji, "Self-supervised bayesian deep learning for image recovery with applications to compressive sensing," in *Proc. ECCV*. Springer, 2020, pp. 475–491.

[24] Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 218:1–13, 2017.