Stochastic Deep Image Prior for Multishot Compressive Spectral Image Fusion

Roman Jacome, Brayan Monroy, Jorge Bacca and Henry Arguello
Universidad Industrial de Santander
{roman2162474@correo,brayan2180034@correo,jorge.bacca1@correo.,henarfu@}.uis.edu.co

Abstract

The Deep Image Prior (DIP) technique has been successfully employed in Compressive Spectral Imaging (CSI) as a non-data-driven deep model approach. DIP methodology updates the deep network's weights by minimizing a loss function that considers the difference between the measurements and the forward operator of the network's output. However, this method often yields local minima as all the measurements are evaluated at each iteration. This paper proposes a stochastic deep image prior (SDIP) approach, which stochastically trains DIP networks using random subsets of measurements from different CSI sensors in a CSI fusion (CSIF) setting, resulting in the improvement of the convergence through stochastic gradient descent optimization. The proposed SDIP method improves upon the deterministic DIP and requires less computational time since fewer forward operators are required per iteration. The SPID method provides comparable performance against the state-of-the-art CSIF techniques based on supervised data-driven and unsupervised methods, achieving up to 5 dB in the reconstruction.

1. Introduction

Deep neural networks (DNNs) for solving inverse problems have gathered significant interest in the research community since their high recovery performance [3, 17, 18, 24, 32]. These methods can be classified into two subcategories: supervised and unsupervised. The first involves training a DNN using a paired dataset to transform the observed measurements into the target data. Several DNN-based methods have been proposed in a wide range of inverse problems [8, 19–21, 25, 34]. [17, 24]. For CSI, blackbox DNNs have been proposed [30], or recently unrolling-based methods which leads to interpretable DNNs from optimization approaches [12, 22, 23, 31]. Despite their outstanding results in image recovery, supervised DNNs are highly dependent on the quality and quantity of the training data. In fact, if the training data is limited, such as in CSI, the performance of these methods can be sub-optimal.

Also, unsupervised methods do not employ paired measurements and target data. The seminal work on deep image prior (DIP) [28] has opened new horizons on unsupervised methods. The DIP approach consists of a DNN as a generative model which obtains the target image from a latent space (usually sampled from a Gaussian distribution). In particular, the DNN parameters are fitting by minimizing the loss function, which measures the discrepancy between the generated image projected by the acquisition operator and the acquired measurements. The DIP framework has effectively adapted to CSI where the latent space can be learned based on a Tucker representation [5], or the DNN architecture can be designed based on the Linear Mixture Model [11]. One of the main characteristics of DIP frameworks consists of only one-sample training, whereby solely the measurements are necessary to compute the error of the network predictions. However, this approach inhibits the use of stochastic gradient descent (SGD) techniques where the network inputs are randomly sampled during each iteration, which presents a challenge for evading local minima. SDG has shown remarkable results in improving over parametrized DNN training convergence [1, 33] and these ideas have been extended to convex optimization for imaging inverse problems in imaging [27].

In CSI, it is possible to acquire several measurements of the scene (multishot acquisition); specifically, the CSI systems are able to capture the coded information of the whole scene. Even we can employ multiple CSI systems following a CSI fusion (CSIF) scheme. CSIF uses two compressive acquisition systems that employ synthetic apertures to acquire measurements at different resolutions, which are then used in a recovery algorithm that fuses the spectral encoded information in a single estimation of the spectral image. To solve the CSIF problem; [29] presents an algorithm based on the alternating direction method of multipliers (ADMM) with sparsity and total variation priors, and [9] promotes non-local low-rank prior on the abundance of the spectral image via an ADMM formulation. Recent DNN-based methods have adopted a model-based formulation, such as networks designed from the iterations of half-quadratic splitting in [13] and ADMM-inspired formulation in [14, 26]. However, state-of-the-art methods currently utilize the complete observation set to optimize the objective function, which can be computationally expensive when estimating high-dimensional signals like spectral images. Additionally, complete optimization may rely on early sub-optimal solutions due to bias towards specific measurements.

Consequently, we proposed stochastic deep image prior (SDIP) for multi-shots CSIF. The proposed method stochastically trains DIP networks employing random subsets of the CSI measurements from different CSI sensors. The network estimation is passes through the corresponding subset of sensing matrices. As a result, the entire set of measurements undergoes stochastic evaluations at each iteration, yielding analogous dynamics of employing stochastic training samples at each iteration in supervised methods, leading to improved convergence through SGD optimization. The SDIP improves the deterministic DIP and requires less computational time since fewer forward operators are needed for each iteration. Finally, the proposed method is compared with unsupervised fusion methods showing an improvement of up to 5 dB in PSNR.

2. Compressive Spectral Image Fusion Observation Model

Two well-known CSI sensors are used; the coded aperture snapshot spectral imager (CASSI) and the multispectral color filter array (MCFA). Similar to [14], CASSI is assumed to have a high spectral resolution but with low spatial resolution, and the MCFA provides a high spatial resolution and a low spectral resolution.

CASSI model: the linear model of the CASSI image formation is formulated as follows

$$\boldsymbol{y}_c^\ell = \boldsymbol{H}_c^\ell \boldsymbol{f} + \boldsymbol{\omega}_c, \tag{1}$$

where $\boldsymbol{f} \in \mathbb{R}^{MNL}$ is the vectorization of the high spatial-spectral resolution image with $M \times N$ spatial pixels and L spectral bands. Let us denote $M_d = \frac{M}{r_s}$ and $N_d = \frac{N}{r_s}$, then $\boldsymbol{y}_c^\ell \in \mathbb{R}^{M_d(N_d+L-1)s_c}$ is the compressed measurements, $r_s > 1$ is the spatial down-sampling factor, ℓ is the CASSI shot index $\boldsymbol{\omega}_c$ is an additive Gaussian noise, and $\boldsymbol{H}_c^\ell \in \mathbb{R}^{M_d(N_d+L-1)s_c \times MNL}$ is the sensing matrix for ℓ-th shot [4, 6] with s_c number of snpashots.

MCFA model: the linear model of the MCFA system is:

$$\boldsymbol{y}_m^\ell = \boldsymbol{H}_m^\ell \boldsymbol{f} + \boldsymbol{\omega}_m, \tag{2}$$

where $\boldsymbol{y}_c^\ell \in \mathbb{R}^{s_m MN}$ is the compressed measurements, $\boldsymbol{\omega}_c$ is an additive Gaussian noise, and $\boldsymbol{H}_m^\ell \in \mathbb{R}^{s_m MN \times MNL}$ is the sensing matrix for ℓ-th shot with s_m number of snapshots

Then, we define the whole set of measurements of the CSIF setting as follows

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{w}, \quad \boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_c \\ \boldsymbol{H}_m \end{bmatrix}, \tag{3}$$

where the whole sensing matrix $\boldsymbol{H}$ is composed of the CASSI $\boldsymbol{H}_c$ and the MCFA $\boldsymbol{H}_m$ sensing matrices. $\boldsymbol{H}_c$ is the vertical stack of the sensing matrices related to the CASSI system, i.e., $\boldsymbol{H}_c = [(\boldsymbol{H}_c^1)^\top, \dots, (\boldsymbol{H}_c^{s_c})^\top]^\top$, and similarly the $\boldsymbol{H}_m^k$ is the vertical stack of the sensing matrices related to the MCFA system, i.e., $\boldsymbol{H}_m = [(\boldsymbol{H}_m^1)^\top, \dots, (\boldsymbol{H}_m^{s_m})^\top]^\top$. Two important factors in CSI are the number total of snapshots for both systems $s = s_c + s_m$ and the compression ratio, which is defined as $\gamma = \frac{s_c(M_d(N_d+L-1)) + s_m(MN)}{MNL}$.

3. Deep Image Prior for CSI Recovery

We aim to recover the spectral image $\mathbf{f}$ from the compressed measurements of the CASSI and MCFA systems. The main idea of DIP for CSI is that untrained DNN is optimized to map a random vector sampled from a Gaussian distribution, $\boldsymbol{z} \in \mathbb{R}^m$, to the desired spectral image as $\widehat{\boldsymbol{f}} = \mathcal{N}_{\widehat{\theta}}(\boldsymbol{z})$ where $\mathcal{N}_{\widehat{\theta}}(\cdot)$ represents the DNN, and $\widehat{\theta}$ are its optimal parameters (see Fig. 1(a,b)) obtained by solving the following optimization problem

$$\widehat{\theta} = \arg\min_{\theta} \|\boldsymbol{y} - \boldsymbol{H}\mathcal{N}_\theta(\boldsymbol{z})\|_2^2. \tag{4}$$

Inspired by our previous work [5], we promote low-rank over the latent space via learning a tucker representation of the spectral image as DNN input. Specifically, $\mathcal{Z} \in \mathbb{R}^{M \times N \times L}$ is the low-rank tensor DNN input of the form $\mathcal{Z} = \mathcal{Z}_0 \times_1 \boldsymbol{U} \times_2 \boldsymbol{V} \times_3 \boldsymbol{W}$ where $\times_1, \times_2, \times_3$ are the tensor mode product in the first, second and third dimension, respectively. The core tensor, $\mathcal{Z}_0 \in \mathbb{R}^{M_\rho \times N_\rho \times L\rho}$, and the matrices $\boldsymbol{U} \in \mathbb{R}^{M_\rho \times M}$, $\boldsymbol{V} \in \mathbb{R}^{N_\rho \times N}$, $\boldsymbol{W} \in \mathbb{R}^{L_\rho \times L}$ have the following relationship with a rank factor ρ as $\frac{M_\rho}{M} = \frac{N_\rho}{N} = \frac{L_\rho}{L} = \rho$, thus promoting low-rank representation, which has proven to be a valid assumption in this type of data [9, 10]. Moreover, the variables $\Omega = \{\widehat{\theta}, \widehat{\mathcal{Z}}_0, \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}, \widehat{\boldsymbol{W}}\}$ can be trained following the DIP methodology by solving

$$\widehat{\Omega} \in \arg\min_{\Omega} \|\boldsymbol{y} - \boldsymbol{H}\mathcal{N}_\theta(\mathcal{Z})\|_2^2 \quad \text{subject to} \quad \mathcal{Z} = \mathcal{Z}_0 \times_1 \boldsymbol{U} \times_2 \boldsymbol{V} \times_3 \boldsymbol{W}. \tag{5}$$

In the baseline DIP framework, the loss function is computed with the whole set of observations in every iteration of the network training, which can lead to falling into local minima.

4. Stochastic Deep Image Prior

The proposed method stochastically trains DIP networks employing random subsets of the CSI measurements from

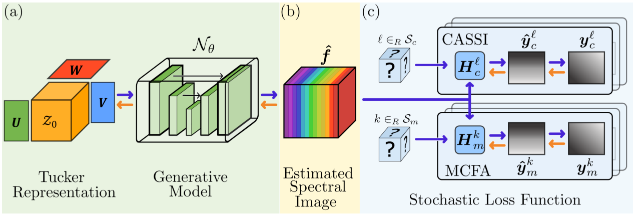

Figure 1. Stochastic Deep Image Prior. (a, b) The estimated spectral image is generated from the tucker representation $\hat{f} = \mathcal{N}_\theta$. (c) Then, the matrices subset H_c^ℓ and H_m^k are randomly selected for the stochastic loss function computation.

different CSI sensors, as shown in Fig. 1. Therefore, let us decompose the main loss function $g(\boldsymbol{f}) = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{f}\|_2^2$ as

$$g(\boldsymbol{f}) = \frac{1}{s}\sum_{i=1}^{s} g_i(\boldsymbol{f})$$

where $g_i(\boldsymbol{f}) = \|\boldsymbol{y}^i - \boldsymbol{H}^i\boldsymbol{f}\|_2^2$. Note that depending on the index i, the function can take either the observation of the CASSI system or the MCFA sensor. Thus, we now consider a subset of index $\mathcal{S}_i = \{i_1, i_2, \dots, i_B\}$ where $B = |\mathcal{S}_i|$ is the minibatch size and the indexes $i_1, i_2, \dots, i_B$ are uniformly sampled from $\mathcal{T} = \{1, 2, \dots, s\}$. Then, the stochastic approach solves the following optimization problem

$$\min_{\Omega} g(\mathcal{N}_\theta(\mathcal{Z})) = \mathbb{E}_{\mathcal{S}_i \sim \mathcal{T}}\left[\|\boldsymbol{y}^{i_k} - \boldsymbol{H}^{i_k}\mathcal{N}_\theta(\mathcal{Z})\|_2^2\right] \tag{6}$$
$$\text{subject to } \mathcal{Z} = \mathcal{Z}_0 \times_1 \boldsymbol{U} \times_2 \boldsymbol{V} \times_3 \boldsymbol{W},$$

for $k = 1, \dots, B$. In this sense, B can be related to the batch size in SGD optimization. This formulation brings two main advantages. The first one is that it reduces the computational complexity of the training since only a portion of the whole sensing matrix is needed to compute the loss function. The second one is convergence acceleration given by the properties of first-order stochastic optimization. Particularly, the optimization of (6) performed by SGD with a randomly selected snapshot of the compressed measurements set $\mathcal{S}_i$ is given by

$$\widehat{\Omega}^{i+1} = \widehat{\Omega}^{i} - \eta\frac{1}{B}\sum_{i_k \in S_i}\frac{\partial}{\partial\Omega} g_{i_k}(\mathcal{N}_{\theta^{(i)}}(\mathcal{Z}^i)) \tag{7}$$

where $\mathcal{Z}^{(i)} = \mathcal{Z}_0^{(i)} \times_1 \boldsymbol{U}^{(i)} \times_2 \boldsymbol{V}^{(i)} \times_3 \boldsymbol{W}^{(i)}$ and η is the gradient step size. It is important to highlight that the gradient direction will depend directly on the set of measurements randomly chosen at each iteration $\mathcal{S}_i \sim \mathcal{T}$. There, the DNN weights are updated according to the stochastic gradient iteration, and the desired recovery image is obtained as $\widehat{\boldsymbol{f}} = \mathcal{N}_{\widehat{\theta}}(\widehat{\mathcal{Z}})$.

5. Results

To validate the proposed stochastic DIP framework for CSIF, we employ the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) in two main experiments. The first one is an ablation study varying the number of batches per iteration, and it is directly compared concerning the deterministic DIP.The second compares the state-of-the-art methods for CSIF, we compared two deep supervised learned-based methods, LADMM [26], and D^{2}UF [14], which are unrolling networks based on ADMM iterations and one unsupervised optimization-based spectral image fusion from compressive measurements (SIFCM) which is an ADMM algorithm with sparsity and total variation prior [29]. As DIP networks, we employ a ResNet and U-Net according to the author's implementation in [5]. All experiments were performed in GPU NVIDIA RTX 3090.

5.1. Stochastic DIP experiments

In order to evaluate the importance of the batch size, we employed the testing images of the KAIST dataset [7] with a size of $M = N = 512$ and $L = 31$. The number of snapshots acquired by the CASSI and MCFA systems was set to $s_h = s_m = 4$ yielding $s = 8$ total number of snapshots of the scene. The spectral and spatial decimation factors were set to $r_s = 2$, and $r_\lambda = 10$; particularly, the value r_λ was chosen such that the MCFA sensor works as an RGB camera with arbitrary CFA. Then, the compression ratio of the CSIF model is $\gamma = 0.165$. The synthetic apertures of both systems are drawn from a Bernoulli distribution with $p = 0.5$. The network employed here was the ResNet, the number of iterations was chosen to be 15000, and the network optimizer was the Adam algorithm [16] with a learning rate of $\eta = 10^{-3}$. All the results reported in this experiment are the mean of 5 runs.

The proposed SDIP was tested with varying batch *batch* size B, $B = \{3, \dots, 8\}$ where $B = 8$ represents the vanilla DIP. Fig. 2(a) shows the reconstruction performance of the SDIP in terms of PSNR along the running time for different values of B. Notably, with $B = 6$, the performance improvement with respect to the vanilla DIP is up to 5 dB, and the time required for stochastic training is noticeably less than for deterministic optimization. It is important to remark that the same number of gradient iterations was fixed

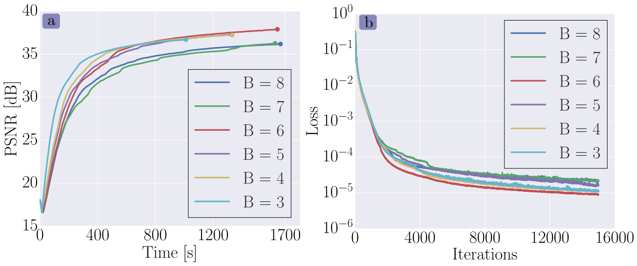

Figure 2. a) PSNR score along running time in seconds of SDIP with different values of B. b) Loss function for different values of B along training iterations. The y-axis is in logarithmic scale.

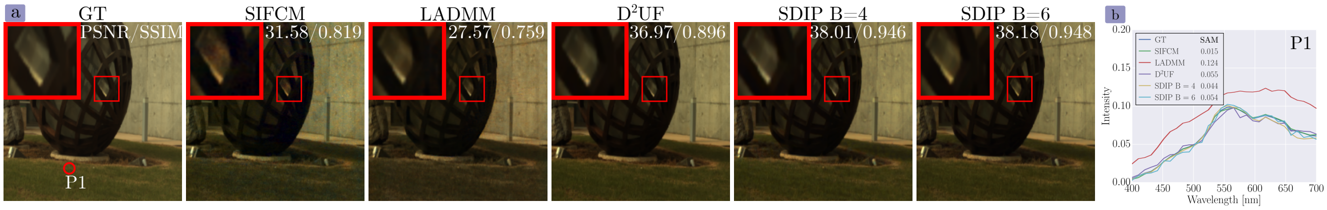

Figure 4. a) RGB visualization comparison with state-of-the-art methods, SIFCM [29], LADMM-NET [26], D^{2}UF [14], and the proposed method SDIP for $B = \{4, 6\}$ on the ICVL dataset, upper-right numbers show the PSNR and SSIM scores. b) Spectral reflectance and SAM score comparison of a random point compared to the reference signature.

Method	Unsupervised	SNR = 20 [dB]			SNR = 25 [dB]			SNR = 30 [dB]			SNR = 35 [dB]			SNR = ∞		
		PSNR ↑	SSIM↑	SAM ↓	PSNR↑	SSIM ↑	SAM ↓	PSNR ↑	SSIM ↑	SAM ↓	PSNR ↑	SSIM ↑	SAM ↓	PSNR↑	SSIM ↑	SAM ↓
LADMM-Net [26]	X	24.33	0.486	0.390	26.42	0.661	0.277	27.51	0.768	0.208	27.91	0.815	0.176	28.13	0.840	0.157
D^{2}UF [14]	X	30.02	0.701	0.333	33.111	0.839	0.239	35.192	0.904	0.193	36.326	0.936	0.168	37.02	0.951	0.151
SIFCM [29]	✓	20.33	0.245	0.658	24.22	0.407	0.486	27.15	0.570	0.337	29.02	0.701	0.210	30.19	0.816	0.170
SDIP $B = 4$	✓	25.63	0.212	0.695	28.77	0.414	0.510	31.49	0.695	0.283	32.58	0.845	0.197	35.65	0.934	0.136
SDIP $B = 6$	✓	26.59	0.138	0.765	29.46	0.312	0.575	30.95	0.703	0.278	34.83	0.907	0.157	35.59	0.913	0.139

Table 1. Numerical results of SDIP with state-of-the-art CSIF methods for the ICVL with compressed measurements corrupted by different levels of SNR. The highlighted green values are the best performance and the blue ones are the second best.

for all scenarios. Results suggest a trade-off between the total recovery time and the final spectral image quality estimation with respect to the B-value. Furthermore, Fig. 2(b) shows that the improvement in running time is coupled with an improvement in loss function optimization convergence for $B < 8$, as the cost function achieves lower values than the vanilla DIP ($B = 8$). Therefore, we selected 4 and 6 for the following experiments, comparing our SDIP with the state-of-the-art CSIF.

5.2. State-of-the-art comparison

We conducted a comparison with state-of-the-art CSIF methods using the ICVL dataset [2]. To train the supervised LADMM-Net, and D^{2}UF, we use 140 images for training and 20 for testing and the parameters suggested by the authors. A central crop of the image was taken with a size of $M = N = 256$. The MCFA and CASSI systems were set to $s_m = s_h = 4$ shots, similar to the previous experiment. For this experiment, the U-Net was used as a DNN generator.

Table 1, provides a quantitative comparison of the evaluated methods under different levels of additive Gaussian noise SNR=$\{20, 25, 30, 35, \infty\}$. While the supervised approach D^{2}UF yields the best results, on average, for unsupervised methods, SDIP achieves the best results, with comparable results to supervised methods for SNR of 35 and noiseless scenarios. One of the main reasons for poor performance under low SNR values is that the baseline DIP loss function is based only on the noisy measurements, yielding a noisy estimation of the image. Recent works have been proposed to improve DIP under noisy observation by changing the loss function as Stein's unbiased risk estimator metric [15] that considers the noise level.

Figure 4(a) shows a testing RGB visualization of state-of-the-art supervised, unsupervised, and SDIP methods, for $B = \{4, 6\}$ in the noiseless case (SNR=∞). The proposed method obtains the best results in PSNR and SSIM scores. Interestingly, the proposed method achieves comparable results even when compared to supervised-based deep learning methods. Furthermore, the proposed method has the added benefit of being able to recover spectral signatures that fit the behavior of the reference signatures (as shown in Fig. 4(b)).

6. Conclusion

The proposed SDIP method for multishot CSIF improves the vanilla DIP method in terms of estimation quality and convergence rate. Stochastic optimization is achieved by randomly selecting snapshots to compute the DIP loss function. This optimization strategy offers two main benefits: the first is an improvement on the DIP running time since fewer forward sensing operators pass are required at each iteration, and the second is the benefits of the SGD optimization for DIP, leading to improve reconstruction performance. Simulation shows that a proper choice of batch size allows a significant improvement in recovery of up to 5 dB compared to the vanilla DIP with less running time. Moreover, we compared SDIP with state-of-the-art CSIF, providing comparable results with supervised methods and performing significantly better with respect to unsupervised approaches in different noise levels. It is worth noting that the proposed SDIP method can be generalized beyond multishot acquisition and can be employed in imaging inverse problems by selecting random partitions of the measurements, thus providing a new method to train DIP.

References

[1] Zeyuan Allen-Zhu, Yuanzhi Li, and Zhao Song. A convergence theory for deep learning via over-parameterization. In *International Conference on Machine Learning*, pages 242–252. PMLR, 2019. 1

[2] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer, 2016. 4

[3] Henry Arguello, Jorge Bacca, Hasindu Kariyawasam, Edwin Vargas, Miguel Marquez, Ramith Hettiarachchi, Hans Garcia, Kithmini Herath, Udith Haputhanthri, Balpreet Singh Ahluwalia, et al. Deep optical coding design in computational imaging. *arXiv e-prints*, pages arXiv–2207, 2022. 1

[4] Henry Arguello, Hoover Rueda, Yuehao Wu, Dennis W. Prather, and Gonzalo R. Arce. Higher-order computational model for coded aperture spectral imaging. *Appl. Opt.*, 52(10):D12–D21, Apr 2013. 2

[5] Jorge Bacca, Yesid Fonseca, and Henry Arguello. Compressive spectral image reconstruction using deep prior and low-rank tensor representation. *Applied optics*, 60(14):4197–4207, 2021. 1, 2, 3

[6] Jorge Bacca, Tatiana Gelvez-Barrera, and Henry Arguello. Deep coded aperture design: An end-to-end approach for computational imaging tasks. *IEEE Transactions on Computational Imaging*, 7:1148–1160, 2021. 2

[7] Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2017)*, 36(6):218:1–13, 2017. 3

[8] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 1

[9] Tatiana Gelvez and Henry Arguello. Nonlocal low-rank abundance prior for compressive spectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 2020. 1, 2

[10] Tatiana Gelvez, Hoover Rueda, and Henry Arguello. Joint sparse and low rank recovery algorithm for compressive hyperspectral imaging. *Appl. Opt.*, 56(24):6785–6795, Aug 2017. 2

[11] Tatiana Gelvez-Barrera, Jorge Bacca, and Henry Arguello. Mixture-net: Low-rank deep image prior inspired by mixture models for spectral image recovery. *arXiv preprint arXiv:2211.02973*, 2022. 1

[12] Tao Huang, Weisheng Dong, Xin Yuan, Jinjian Wu, and Guangming Shi. Deep gaussian scale mixture prior for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16216–16225, 2021. 1

[13] Roman Jacome, Jorge Bacca, and Henry Arguello. Deep-fusion: An end-to-end approach for compressive spectral image fusion. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 2903–2907. IEEE, 2021. 1

[14] Roman Jacome, Jorge Bacca, and Henry Arguello. D2uf: Deep coded aperture design and unrolling algorithm for compressive spectral image fusion. *IEEE Journal of Selected Topics in Signal Processing*, pages 1–11, 2022. 2, 3, 4

[15] Yeonsik Jo, Se Young Chun, and Jonghyun Choi. Rethinking deep image prior for denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5087–5096, 2021. 4

[16] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 3

[17] Alice Lucas, Michael Iliadis, Rafael Molina, and Aggelos K Katsaggelos. Using deep neural networks for inverse problems in imaging: beyond analytical methods. *IEEE Signal Processing Magazine*, 35(1):20–36, 2018. 1

[18] Michael T McCann, Kyong Hwan Jin, and Michael Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95, 2017. 1

[19] Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, pages 187–204. Springer, 2020. 1

[20] Christopher Metzler, Phillip Schniter, Ashok Veeraraghavan, and Richard Baraniuk. prdeep: Robust phase retrieval with a flexible deep network. In *International Conference on Machine Learning*, pages 3501–3510. PMLR, 2018. 1

[21] Xin Miao, Xin Yuan, Yunchen Pu, and Vassilis Athitsos. l-net: Reconstruct hyperspectral images from a snapshot measurement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4059–4069, 2019. 1

[22] Vishal Monga, Yuelong Li, and Yonina C Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021. 1

[23] Brayan Monroy, Jorge Bacca, and Henry Arguello. Jr2net: a joint non-linear representation and recovery network for compressive spectral imaging. *Applied Optics*, 61(26):7757–7766, 2022. 1

[24] Gregory Ongie, Ajil Jalal, Christopher A Metzler, Richard G Baraniuk, Alexandros G Dimakis, and Rebecca Willett. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020. 1

[25] Zhen Qin, Qingliang Zeng, Yixin Zong, and Fan Xu. Image inpainting based on deep learning: A review. *Displays*, 69:102028, 2021. 1

[26] Juan Marcos Ramirez, José Ignacio Martínez-Torre, and Henry Arguello. Ladmm-net: An unrolled deep network for spectral image fusion from compressive data. *Signal Processing*, 189:108239, 2021. 2, 3, 4

[27] Junqi Tang, Karen Egiazarian, Mohammad Golbabaee, and Mike Davies. The practicality of stochastic optimization in imaging inverse problems. *IEEE Transactions on Computational Imaging*, 6:1471–1485, 2020. 1

[28] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. 1

[29] Edwin Vargas, Oscar Espitia, Henry Arguello, and Jean-Yves Tourneret. Spectral image fusion from compressive measurements. *IEEE Transactions on Image Processing*, 28(5):2271–2282, 2018. 1, 3, 4

[30] Lizhi Wang, Tao Zhang, Ying Fu, and Hua Huang. Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. *IEEE Transactions on Image Processing*, 28(5):2257–2270, 2018. 1

[31] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu. Multispectral and hyperspectral image fusion by ms/hs fusion net. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1585–1594, 2019. 1

[32] Jong Chul Ye, Yoseob Han, and Eunju Cha. Deep convolutional framelets: A general deep learning framework for inverse problems. *SIAM Journal on Imaging Sciences*, 11(2):991–1048, 2018. 1

[33] Chiyuan Zhang, Qianli Liao, Alexander Rakhlin, Karthik Sridharan, Brando Miranda, Noah Golowich, and Tomaso Poggio. Musings on deep learning: Properties of sgd. Technical report, Center for Brains, Minds and Machines (CBMM), 2017. 1

[34] Jiawei Zhang, Jinshan Pan, Wei-Sheng Lai, Rynson WH Lau, and Ming-Hsuan Yang. Learning fully convolutional networks for iterative non-blind deconvolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3817–3825, 2017. 1