Letter Vol. 50, No. 4 / 15 February 2025 / *Optics Letters* 1281

Optics Letters

Fourier single-pixel spectral imaging via local low-rank tensor nuclear norm and deep tensor priors

ZIXIN TANG,[1,2,3,†] **YEXUN HU,**[1] **CHEN DUO,**[4,5,6] **GUOWEI YANG,**[1] **TAIXIANG JIANG,**[1,2,3,7,†] **AND DAQING GUO**[5,8,†]

[1] *School of Computing and Artificial Intelligence, Southwestern University of Finance and Economics, Chengdu, Sichuan Province, China*
[2] *Kash Institute of Electronics and Information Industry, Kash, China*
[3] *Engineering Research Center of Intelligent Finance, Ministry of Education, Southwestern University of Finance and Economics, China*
[4] *Chongqing University of Education, School of Artificial Intelligence, China*
[5] *School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan Province, China*
[6] *Chongqing University Industrial Technology Research Institute, Department of Science and Technology Development, China*
[7] *taixiangjiang@gmail.com*
[8] *dqguo@uestc.edu.cn*
[†] *These authors contributed equally to this work.*

Received 21 November 2024; revised 15 January 2025; accepted 15 January 2025; posted 16 January 2025; published 10 February 2025

The imaging quality of single-pixel spectral imaging (SSI) methods is poor at a low sampling ratio (SR). To tackle this problem, a new Fourier single-pixel spectral imaging (FSSI) technique is proposed. Firstly, we introduce the low-rank tensor nuclear norm (TNN) to characterize the correlation between spectral images. Compared with the conventional method, TNN reconstructs image details better but brings image artifacts simultaneously. Therefore, local low-rank TNN (LTNN) constraint is proposed to ameliorate global ones and to reduce the distortion caused by TNN and low SR. Secondly, to make full use of the spectral information, the proposed constraint is used as the coarse prior, and the deep tensor prior (DTP) is introduced as the fine one to construct the joint priors. Different from the single prior, the joint method can make the two priors benefit and improve each other and further enhance the imaging quality. Finally, an efficient and high-quality SSI technique is achieved by deducing the closed-form solution algorithm. Experimental results show that our method significantly improves the quality of FSSI as much as 7–10 dB when compared to 3DTV at the SR of 5%.

https://doi.org/10.1364/OL.549558

Single-pixel imaging (SPI) technology has been widely concerned by researchers with its special imaging style in recent years. It has been commonly used in various imaging systems, such as infrared imaging [1], single-photon imaging [2], and multi-spectral imaging [3–5]. SPI captures scene information only with a single-point detector, which dramatically reduces the manufacturing cost of non-visible light or single-photon detectors. However, it faces the problems of slow imaging speed and poor imaging quality at low SR. To solve these problems, researchers have proposed a large number of solutions, including Hadamard basis scanning, Fourier spectrum sampling, and random mask imaging methods [6–8]. In addition to the encoding technique, the reconstruction algorithm is also an essential factor affecting the performance of SPI. In recent years, various image reconstruction methods have been proposed to solve the problem of poor image quality at a low SR, including optimization-based methods, end-to-end deep learning, and untrained deep learning algorithms [9–13]. However, these methods all have various flaws. For instance, optimization-based methods often improve imaging quality by handcraft priors such as total variation and l_1 regularization, but those methods fail to adequately characterize the intrinsic features of the data, thus resulting in poor imaging quality at the low SR. The end-to-end deep learning techniques significantly improve image quality, but their performance depends on the selection of training samples, and the generalization of the network is poor. The untrained method essentially uses neural networks to solve optimization problems, and its performance depends on the image prior and network design. In this paper, we proposed a new technique that combines LTNN and DTP to facilitate FSSI. Compared to the conventional handcraft prior, such as 3DTV, our approach improves imaging detail by introducing and refining tensor techniques. Besides, the introduction of deep tensor prior and their combination with handcraft constraint fully exploit redundant information of images, thereby avoiding the problem of poor imaging quality in traditional single priors. Experimental results show that our method is superior.

Usually, multi-spectral images can be represented by three-dimensional tensors. Thus, the single-pixel multi-spectral image reconstruction can be considered as a tensor recovery problem. Notably, adjacent spectral images have a considerable similarity. Therefore, the three-dimensional tensor of consecutive spectral images is low rank or sparse in the transformation domain. In general, low-rank-based reconstruction methods can be used to restore the original scene, but these methods often rearrange the three-dimensional tensor into a two-dimensional matrix and reconstruct it via the low-rank constraint of the matrix. This strategy will destroy the tensor's internal structure, resulting in the quality reduction of tensor recovery. To solve this problem, the tensor nuclear norm (TNN) based on tensor singular value decomposition (t-SVD) [14–16] is introduced to describe the correlation of multi-spectral images:

$$\hat{\mathcal{X}} = \underset{\mathcal{X}}{\arg\min} \, ||\mathcal{X}||_{TNN}, \tag{1}$$

where $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the three-dimensional tensor, which is composed of multi-spectral images and $||\mathcal{X}||_{TNN} = \sum_{i=1}^{n_3} ||\bar{X}^{(i)}||_*$, $\bar{X}^{(i)}$ represents the i-th front-slice of the tensor $\bar{\mathcal{X}}$, and $\bar{\mathcal{X}}$ is the Fourier transform of tensor $\bar{\mathcal{X}}$ along the mode-3 fibers. $|| * ||_*$ represents the nuclear norm. The sum of $||\bar{X}^{(i)}||_*$ can be obtained by the t-SVD of tensor $\mathcal{X}$ in the Fourier domain. Essentially, TNN is the correlation description of the multi-spectral images. If the images are highly correlated, then its TNN is very low. Generally, the multi-spectral images are highly similar and have redundancy features in the spectral domain, but they are not entirely identical due to the images' abundant details and spectral change. To solve this problem, a new method is proposed based on image local low-rank TNN (LTNN) to improve the performance further, which uses TNN constraints to every sub-region image of corresponding adjacent spectra. In this manner, it precisely utilizes spectral redundancy information to reconstruct the image and retain its internal structure and detail information simultaneously. Therefore, the new image reconstruction model can be rewritten as follows:

$$\hat{\mathcal{X}} = \underset{\mathcal{X}}{\arg\min} \, ||\mathcal{X}||_{LTNN}. \tag{2}$$

LTNN enhances TNN's modeling ability for tensors, but due to the abundant details in the adjoining spectrum image, the handcraft local low-rank tensor prior cannot describe the multi-spectral image sufficiently. To tackle this problem, a widely used deep prior in image super resolution and denoising [17,18] is introduced to further facilitate the model and improve the reconstructed image's details. Different from the traditional deep priors for a single image, we use the deep tensor prior to describe the correlation between spectral images. Compared to low-rank tensor constraints, the deep tensor prior uses the knowledge learned from the 3D data by the neural network to optimize the image reconstruction process. In this paper, deep and low-rank tensor constraints are integrated into the image reconstruction model simultaneously so that they can complement each other and fully describe the intrinsic features of the image. Therefore, the image reconstruction model can be further rewritten as follows:

$$\hat{\mathcal{X}} = \underset{\mathcal{X}}{\arg\min} \, \mu||\mathcal{X}||_{LTNN} + \lambda R(\mathcal{X}), \tag{3}$$

where $R(\mathcal{X})$ is the deep tensor prior. μ and λ are penalty coefficients, which can adjust the proportion of the two priors, enabling the model to achieve the best performance in describing the intrinsic features of the image. In addition, the image reconstruction process should conform to the Fourier single-pixel measurement equation. Therefore, the above formula can be rewritten as follows:

$$\hat{\mathcal{X}} = \underset{\mathcal{X}}{\arg\min} \, \mu||\mathcal{X}||_{LTNN} + \lambda R(\mathcal{X}), \quad s.t. \quad ||\mathcal{F}_p(\mathcal{X}) - \mathcal{Y}||_2^2 < \delta^2, \tag{4}$$

where $\mathcal{F}_p$ is the three-dimensional (3D) partial Fourier matrix and $\mathcal{Y}$ is the Fourier spectrum tensor obtained by a single-pixel detector. To solve the above problem, we change the $\mathcal{F}_p(\mathcal{X})$ into $S \odot \mathcal{F}(\mathcal{X})$ and introduce the alternating direction method of multipliers (ADMM) algorithm [19] to optimize the solving process, where S represents the 0–1 sampling matrix and $\odot$ represents the element-wise product. Under the framework of ADMM, we can introduce additional tensors $\mathcal{U}$ and $\mathcal{V}$ to optimize the solving process. Therefore, the above problems can be transformed into the following constraint:

$$\begin{aligned} &\hat{\mathcal{X}} = \underset{\mathcal{X}}{\arg\min} \, \mu||\mathcal{X}||_{LTNN} + \lambda R(\mathcal{X}) + \frac{\delta}{2}||S \odot F(\mathcal{X}) - \mathcal{Y}||_2^2) \\ &s.t. \mathcal{U} = \mathcal{X}, \mathcal{V} = \mathcal{X}. \end{aligned} \tag{5}$$

Then, we can use the Lagrange multiplier method to transform the above constrained problem into the following unconstrained problem:

$$\begin{aligned} \hat{\mathcal{X}} &= \underset{\mathcal{X}}{\arg\min} \, \mu||\mathcal{V}||_{LTNN} + \lambda R(\mathcal{U}) + \frac{\delta}{2}||S \odot F(\mathcal{X}) - \mathcal{Y}||_2^2 \\ &+ \frac{\beta}{2}\left(||\mathcal{X} - \mathcal{V} + \frac{\gamma_v}{\beta}||_2^2 + ||\mathcal{X} - \mathcal{U} + \frac{\gamma_u}{\beta}||_2^2\right), \end{aligned} \tag{6}$$

where γ_v and γ_u are the Lagrange multipliers and δ and β are the penalty coefficients. By introducing additional tensors and combining them with the variable splitting algorithm, it is possible to achieve step-by-step analytical solutions for each sub-problem, thereby improving the iteration efficiency of the algorithm. For $\mathcal{X}$ sub-problem, we can obtain the solution:

$$\mathcal{X} = F^{-1}\left(\frac{F(\delta F^{-1} S^T \odot \mathcal{Y} + (\mathcal{U} - \frac{\gamma_u}{\beta}) + (\mathcal{V} - \frac{\gamma_v}{\beta}))}{F(\delta F^{-1} S^T \odot SF + \gamma)}\right). \tag{7}$$

For the tensor $\mathcal{U}$ sub-problem, we can obtain the solution by solving for the following equation:

$$\hat{\mathcal{U}} = \underset{\mathcal{U}}{\arg\min} \, \frac{\beta}{2}||\mathcal{U} - \mathcal{X} + \frac{\gamma_u}{\beta}||_2^2 + \lambda||R(\mathcal{U})||_2^2. \tag{8}$$

The $\mathcal{U}$ sub-problem is a multi-dimensional image-denoising problem, which can be solved by the neural network. Since HSI-SDeCNN [20] has the advantages of real-time and sound denoising performance, we adopt it to solve the above problem. Thus, the solution of $\mathcal{U}$ can be obtained by the following:

$$\mathcal{U} = D\left(\mathcal{X} + \frac{\gamma_u}{\beta}, \delta\right), \tag{9}$$

where $\delta = \sqrt{\frac{\lambda}{\beta}}$ represents the noise level in HSI-SDeCNN, but is the error between ground truth and the estimation in the FSSI problem, and D represents the deep neural network denoising. The last sub-problem is the $\mathcal{V}$ sub-problem. Based on the variable splitting method, we can get it by the following:

$$\hat{\mathcal{V}} = \underset{\mathcal{V}}{\arg\min} \, \frac{\beta}{2}||\mathcal{X} - \mathcal{V} + \frac{\gamma_v}{\beta}||_2^2 + \lambda||\mathcal{V}||_{LTNN}. \tag{10}$$

To solve the above equation, it needs to be transformed into the following problem:

$$\hat{\mathcal{V}} = \underset{\mathcal{V}}{\arg\min} \, \frac{\beta}{2}\sum_i ||Q_i - \mathcal{P}_i||_2^2 + \lambda \sum_i ||\mathcal{P}_i||_{TNN}, \tag{11}$$

where $\mathcal{P}_i$ is the tensor composed of corresponding sub-regions between consecutive spectral tensor $\mathcal{V}$. Q_i represents the same sub-regions composed of tensor $\mathcal{X}$ with error. The closed-form

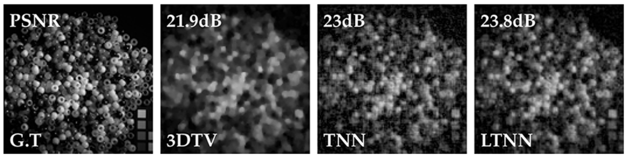

Fig. 1. The 660 nm spectral images (128 × 128) were reconstructed by 3DTV, TNN, and LTNN constraint, respectively.

solution of the sub-region can be obtained by the tensor singular value decomposition algorithm [21,22]. Meanwhile, the updated formula of Lagrange multipliers γ_v and γ_u can be obtained by $\gamma_v = \gamma_v + \beta(\mathcal{V} - \mathcal{X})$ and $\gamma_u = \gamma_u + \beta(\mathcal{U} - \mathcal{X})$.

We first test the performance of the proposed method by comparing the LTNN constraint with the original TNN and the traditional three-dimensional total variation (3DTV) regularization method in multi-spectral SPI. The test images used in this paper are from a well-known CAVE dataset [23]. In every simulation experiment, 31 spectral images are used simultaneously for testing, and the sampling rate is 5%. The parameters of μ and β in LTNN and TNN are set to 100 and 1, respectively. The parameter λ is set to 0 for removing DIP. The test results are shown in Fig. 1, and the sampling method is variable density random sampling [24].

From Fig. 1, it can be seen that the reconstructed image of 3DTV is blurry, with many image details lost, while the result of TNN has better image details. However, what cannot be ignored is that the image obtained by TNN is also mixed with more artifacts. In contrast, the images reconstructed by LTNN have the best clarity and artifact suppression. The above results preliminarily demonstrate the effectiveness of the LTNN prior. Still, it can be seen that the improvement of the LTNN prior is limited (about 1–2 dB's PSNR higher than TNN in numerous tests). In response to this issue, we introduce DTP to complement LTNN. The essence of the DTP is to use the knowledge learned by deep neural networks to improve imaging quality. Thus, it has a specific dependence on training data but describes the intrinsic features of the data more fundamentally. On the contrary, LTNN do not rely on any training data and have better generalization. Based on the advantages and disadvantages of the above two approaches, this paper proposes joint priors to fully integrate the advantages of both methods and improve imaging quality.

To verify the performance of the proposed techniques, we use multiple multi-spectral images (128 × 128 × 31) for testing. In the simulation experiment, the parameters μ, β, and δ were set to 40, 0.5, and 1, respectively. At the same time, the ablation test of the proposed method is carried out: the performance of the LTNN prior and DTP is tested separately and compared with the joint priors. The parameters of every method are set the same, except for the μ and λ in LTNN and DTP that are set to 0, respectively. The experimental results are shown in Fig. 2.

From the results, it can be seen that the image quality reconstructed by the 3DTV is the lowest among the four methods. In comparison, the images reconstructed by the other three constraints are superior, especially regarding image details. Specifically, the LTNN-based method outperforms the 3DTV method in terms of image quality but is inferior to the results reconstructed by the DTP or neural network method. Compared to the LTNN method, the neural network based learned prior reconstructs more clear images with richer details. This indicates that constraint models based on deep prior can better recover image information. The reason is that the image prior knowledge obtained from a large amount of training data can characterize the intrinsic features of images more efficiently. The constraint method combining LTNN and DTP has the best performance, and the PSNR value of the reconstructed image exceeds the single deep prior by 4–5 dB at a sampling rate of 5%. From the image quality perspective, the joint method outperforms single ones such as LTNN or DTP regarding image details and artifact suppression. The reason is that when $\mathcal{X}$ with errors is reconstructed through Eq. (7), then two restoration models (9, 11) will be used consecutively to restore $\mathcal{X}$. Thus, the performance of joint prior is better.

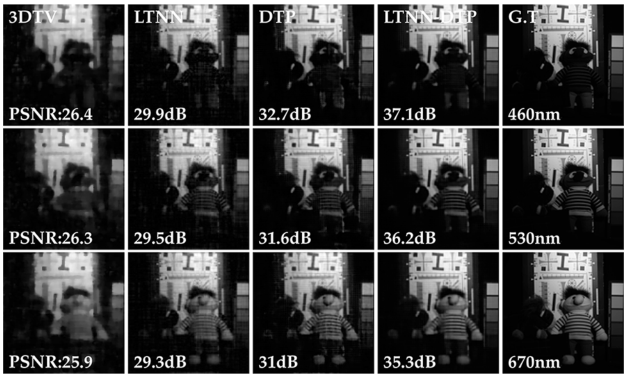

Fig. 2. Three different wave band spectral images (128 × 128) were reconstructed by 3DTV, LTNN, DTP, and LTNN–DTP joint priors (5% sampling ratio), respectively.

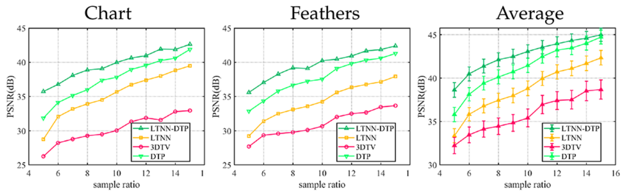

Fig. 3. Reconstruction results of two different (Chart and Feathers) multi-spectral images and the average PSNR curves of all the multi-spectral images in CAVE with different priors.

Without loss of generality, we tested the algorithm's performance at different scene and sampling rates and selected all the multi-spectral images in the CAVE dataset as test samples (SR, 5–15%; resolution, 128 × 128 × 31). The above results (Fig. 3) indicates that the method proposed in this paper has a significant lead over the 3DTV prior at low sampling rates. Precisely, at a sampling rate of 5%, the joint priors reconstruct images with a PSNR 7–10 dB higher than the 3DTV, which fully reveals the excellent performance of the proposed method. In addition, compared with other single prior constraints such as LTNN or DTP, the joint model also shows better implementation in different sampling rates and imaging scenes. But as the SR increases, the image quality gradually approaches the GT, and the parameters δ, μ, and the network model trained with noisy images are no longer applicable to the new input images, resulting in a relative performance decrease between the joint and single priors. However, it can be improved by adjusting parameters and retraining the model. In terms of computational time, the proposed method reconstructs 31 images simultaneously and takes an average of 65 s, which is achieved on an outdated Intel laptop

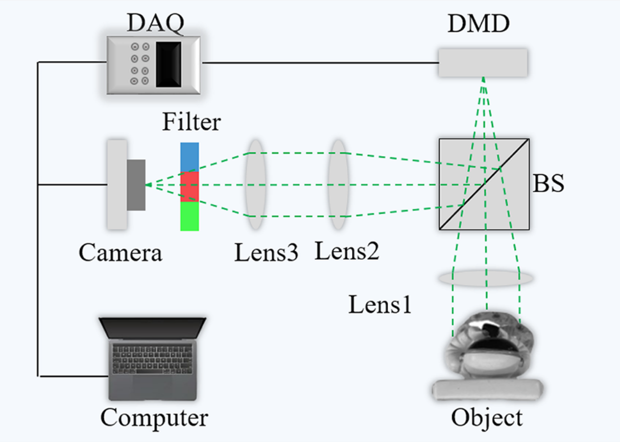

Fig. 4. Schematic diagram of our optical imaging system.

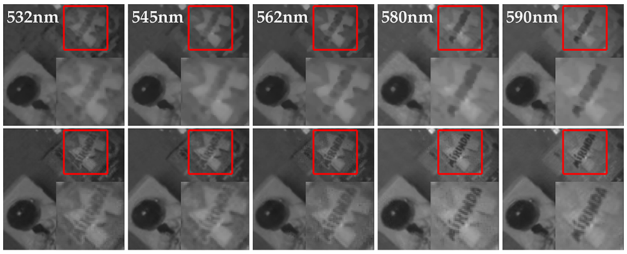

Fig. 5. Multi-spectral images obtained in optical experiments (10% SR). The first row is the result of the 3DTV constraint, and the second row is the reconstruction result of the joint priors.

CPU. If more advanced hardware is adopted, it is expected to reduce the reconstruction time significantly.

We designed an optical system to test the imaging performance of the proposed method in the physical environment. The optical path diagram is shown in Fig. 4. In the diagram, the reflected light of the object is collected by the imaging lens and projected through a beam splitter on the digital micro-mirror device (DMD, Vialux v-7001). The DMD encodes the scene images and then reflects light through the beam splitter again to the camera by two relay lens. The camera is used as a single-pixel detector, and the image information on the CCD is integrated as a measurement signal. We use a four-step phase-shifting method [7] to encode images and obtain a Fourier spectrum based on variable density random sampling in the experiment. To acquire different spectral images, we used a six-channel wheel installed with six wave filters, and the spectral range is from 532 to 600 nm. Besides, a data acquisition card (DAQ, NI-6211) controls the DMD and camera for synchronous encoding and imaging. In this experiment, our hardware system is CPU: Intel i5-9300H with 32 GB memory. The software platform is MATLAB 2022, and the deep learning environment is MatConvNet. The imaging results are shown in Fig. 5. As can be seen from the results, the method proposed in this paper achieves better imaging quality than the traditional method at a low sampling rate, especially the text in the image that is more clearer, which proves that the joint algorithm performs better than the traditional method in real imaging scene.

In conclusion, to solve the problem of low imaging quality in multi-spectral SPI technology at a low sampling rate, this paper proposes a new FSSI reconstruction method that combines LTNN and DTP. Experimental results show that the technique proposed in this paper outperforms the traditional prior models, such as 3DTV, as much as 7–10 dB in PSNR at 5% SR. Our joint framework not only improves the imaging quality of FSSI at low SR but can also be extended to application scenarios such as video imaging where there is a high correlation between consecutive frames. In view of the complexity of our method, in the future, the computational efficiency can be significantly improved by optimizing algorithms or introducing hardware devices such as GPUs, thereby further enhancing practicality.

Funding. Natural Science Foundation of Xinjiang Uygur Autonomous Region (2024D01B06, 2024D01A18); Sichuan Provincial Science and Technology Support Program (24NSFSC1452, 2024ZYD0147); Natural Science Foundation of Chongqing Municipality (CSTB2024NSCQ-MSX0627); Science and Technology Research Program of Chongqing Municipal Education Commission (KJZD-K202401603); National Key Research and Development Program of China (2023YFF1204200); The Graduate Representative Achievement Cultivation Project of SWUFE (JGS2024069); China Postdoctoral Science Foundation (2024M763876).

Disclosures. The authors declare no conflicts of interest.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. G. M. Gibson, B. Sun, M. P. Edgar, *et al.*, Opt. Express **25**, 2998 (2017).
2. Y. Wang, K. Huang, J. Fang, *et al.*, Nat. Commun. **14**, 1073 (2023).
3. Y. Xu, L. Lu, V. Saragadam, *et al.*, Nat. Commun. **15**, 1456 (2024).
4. Z. Li, J. Suo, X. Hu, *et al.*, Sci. Rep. **7**, 41435 (2017).
5. Y. Wang, J. Suo, J. Fan, *et al.*, IEEE Photon. Technol. Lett. **28**, 288 (2015).
6. E. Hahamovich, S. Monin, Y. Hazan, *et al.*, Nat. Commun. **12**, 4516 (2021).
7. Z. Zhang, X. Ma, and J. Zhong, Nat. Commun. **6**, 6225 (2015).
8. A.-X. Zhang, Y.-H. He, L.-A. Wu, *et al.*, Optica **5**, 374 (2018).
9. S. Rizvi, J. Cao, K. Zhang, *et al.*, Opt. Express **28**, 7360 (2020).
10. F. Wang, C. Wang, M. Chen, *et al.*, Light: Sci. Appl. **11**, 1 (2022).
11. Y. Tian, Y. Fu, and J. Zhang, Opt. Laser Eng. **154**, 106970 (2022).
12. C. Li, W. Yin, H. Jiang, *et al.*, Comput. Optim. Appl. **56**, 507 (2013).
13. M. F. Duarte, M. A. Davenport, D. Takhar, *et al.*, IEEE Signal Process Mag. **25**, 83 (2008).
14. T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, *et al.*, J. Comput. Appl. Math. **372**, 112680 (2020).
15. C. Lu, J. Feng, Y. Chen, *et al.*, in *2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2016), pp. 5249–5257.
16. A. Wang, Z. Lai, and Z. Jin, Neurocomputing **330**, 267 (2019).
17. K. Zhang, W. Zuo, and L. Zhang, in *2019 IEEECVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2019), pp. 1671–1681.
18. K. Zhang, Y. Li, W. Zuo, *et al.*, IEEE Trans. Pattern Anal. Mach. Intell. **44**, 6360 (2022).
19. S. Boyd, Found Trends Mach. Learn. **3**, 1 (2010).
20. A. Maffei, J. M. Haut, M. E. Paoletti, *et al.*, IEEE Trans. Geosci. Remote Sensing **58**, 2516 (2020).
21. C. Lu, J. Feng, Y. Chen, *et al.*, IEEE Trans. Pattern Anal. Mach. Intell. **42**, 925 (2020).
22. J. F. Candés, E. J. Shen, and Z. Cai, SIAM J. Optim. **20**, 1956 (2010).
23. Columbia Imaging and Vision Laboratory, "Multi-spectral image dataset," CAVE (2008), https://cave.cs.columbia.edu/repository/Multispectral.
24. M. Wenwen, S. Dongfeng, H. Jian, *et al.*, Opt. Express **27**, 31490 (2019).