HYPERSPECTRAL IMAGE SUPER-RESOLUTION VIA ADJACENT SPECTRAL FUSION STRATEGY

*Qiang Li, Qi Wang *, Xuelong Li*

School of Computer Science and School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, P.R. China

ABSTRACT

Hyperspectral image exhibits low spatial resolution due to the limitation of imaging system. Improving it without an auxiliary high resolution (HR) image still remains a challenging problem. Recently, although many deep learning-based hyperspectral image super-resolution (SR) methods have been proposed, they make the insufficient utilization of adjacent bands to improve the reconstruction performance. To address this issue, we explore a new structure for hyperspectral image SR via adjacent spectral fusion strategy. Inspired by the high similarity among adjacent bands, neighboring band partition is proposed to divide the adjacent bands into several groups. Through the current band, the adjacent bands is guided to enhance the exploration ability. To explore more complementary information, an alternative fusion mechanism, i.e., intra-group fusion and inter-group fusion, is designed, which helps to recover the missing details in the current band. Experiments demonstrate that our approach produces the state-of-the-art results over the existing approaches.

Index Terms— Hyperspectral image, super-resolution (SR), adjacent bands, group fusion

1. INTRODUCTION

Hyperspectral image is obtained through dozens to hundreds of contiguous spectral bands using imaging system. Since more bands are divided within a limited spectral range, the spatial resolution of hyperspectral image is lower than that of natural or multispectral image [1], which affects the subsequent interpretation and application. For accurate descriptions of the image, hyperspectral image super-resolution (SR) is proposed, which aims at producing a high-resolution (HR) image from its corresponding low-resolution (LR) version.

Recently, deep learning-based hyperspectral image SR methods [2, 3, 4, 5, 6] provide great success due to the powerful representational ability of CNNs. At present, hyperspectral image SR algorithms are mainly divided into two categories: CNNs using 2D convolution [7, 3, 8] and CNNs using 3D convolution [9, 6]. As for the network with only 2D convolution, the researchers refer to the natural image SR algorithms [10, 11] to design the model. Typical methods are GDRRN [7] and SSPSR [3]. Because 2D convolution cannot effectively explore spectral information, this type of algorithm generally has poor performance.

Based on the fact that rich spectral knowledge can improve the performance for spatial resolution, employing 3D convolution to study SR has become a hot research. Mei *et al.* [9] first use regular 3D convolution to design the network (3D-FCNN). Compared with the models adopting 2D convolution, it produces promising results. Later, many CNNs using 3D convolution are proposed [6, 12, 13]. For example, Yang *et al.* [14] develop multi-scale wavelet 3D convolution neural network. Li *et al.* [15] present 3D generative adversarial network. The above algorithms all utilize regular 3D convolution to explore the features of hyperspectral image. Usually, using regular 3D convolution directly produces a large number of parameters. To solve this issue, the researchers [16] modify the filter $k \times k \times k$ for 3D convolution as $k \times 1 \times 1$ and $1 \times k \times k$. As a result, the number of network parameters is significantly reduced. Existing works applying this manner include SSRNet [12], MCNet [6], [17], etc. Among these approaches, MCNet is by far the best performance. This method mainly uses mixed 2D/3D convolution through sharing space features to process SR task. Inspired by this mixed convolution, we also adopt this manner to design the network.

The hyperspectral image has the remarkable characteristics of high similarity between adjacent bands [18]. When reconstructing the current band, if the adjacent bands are employed effectively, the complementary information would be beneficial to recover more missing details. Nevertheless, the above algorithms ignores this point. Besides, in a certain spectral range, the sharpness of the edge in the image varies with the bands. It indicates the information of different bands complements each other. Therefore, the central issue is *how to make effective use of adjacent bands to improve the reconstruction performance*. In general, the similarity of adjacent bands is higher than that of nonadjacent bands. Such distant neighboring bands are not explicitly guided by the current band. Considering this issue, in our paper, the adjacent bands

*Qi Wang is the corresponding author. This work was supported by the National Key R&D Program of China under Grant 2018YFB1107403, National Natural Science Foundation of China under Grant U1864204, 61773316, U1801262, and 61871470.

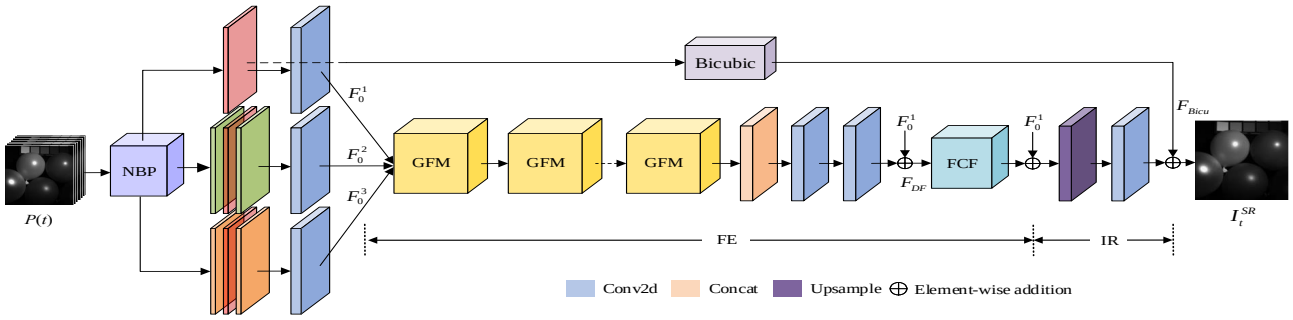

Fig. 1. The overall architecture of the proposed network for hyperspectral image SR.

are divided into several groups based on the position of the current band. To generate the missing information in current band form adjacent bands, a novel adjacent spectral fusion strategy is proposed to enable the model to obtain missing information from groups. In summary, our main contributions of this paper are as follows:

• Neighboring band partition is proposed to divide multiple bands with high similarity to the current band into several groups, which guides the distant bands to explore potential content.

• An alternative fusion mechanism contains intra-group and inter-group fusion is designed, which borrows complementary information from neighboring bands, recovering the more missing details.

• Experiments demonstrate that our method outperforms the state-of-the-art methods across datasets and scales in terms of spatial reconstruction and spectral fidelity.

2. PROPOSED METHOD

In this section, we detail the architecture of the proposed method, including network structure, group fusion, and feature context fusion.

2.1. Network Architecture

Now we present the proposed architecture, which is shown in Fig. 1. The network mainly contains feature extraction (FE) and image reconstruction (IR). Different from previous works [6, 3], our method adopt the single band and its adjacent bands for analysis. Let I_t^{LR} denote the t-th LR band. Four adjacent bands are I_{t-1}^{LR}, I_{t-2}^{LR}, I_{t+1}^{LR}, and I_{t+2}^{LR}, respectively. To obtain complementary information, five bands are fed into designed network simultaneously, which is denoted as

$$P(t) = \begin{cases} \left[I_1^{LR}, I_2^{LR}, I_3^{LR}, I_4^{LR}, I_5^{LR}\right], & 1 \le t < 3 \\ \left[I_{t-2}^{LR}, I_{t-1}^{LR}, I_t^{LR}, I_{t+1}^{LR}, I_{t+2}^{LR}\right], & 3 \le t \le L-3, \\ \left[I_{L-4}^{LR}, I_{L-3}^{LR}, I_{L-2}^{LR}, I_{L-1}^{LR}, I_L^{LR}\right], & L-3 < t \le L \end{cases} \tag{1}$$

where L is the total number of bands in hyperspectral image, and $P(t)$ represents the set of input bands. Since the sharpness of the edge in the image varies with the bands, in our work, the set $P(t)$ is divided into three groups. After constructing the groups, we obtain the corresponding shallow features through 3×3 convolution, respectively. Then, the initial features are input into group fusion modules (GFMs). These modules guide the model to extract complementary information from neighboring bands, allowing effective information extraction and fusion. Under the action of concatenation, two 2D convolution layers, and long skip connection, we acquire the output F_{DF} of FE. To make effectively use of the learned features from previous band, it is transferred to the reconstruction task of the current band. Subsequently, we upsample the features into HR space by sub-pixel convolution according to scale r, which is followed by a 3×3 convolution layer. Finally, the reconstructed hyperspectral image I_t^{SR} is obtained after an additional cross-space skip residual F_{Bicu}, i.e.,

$$I_t^{SR} = F_{IR} + F_{Bicu}. \tag{2}$$

2.2. Group Fusion

Inspired by the fact that high similarity exists in adjacent bands [18, 19], an alternative fusion mechanism is designed to guide the model to obtain more useful information from intra/inter-groups. Next, we will describe the details about fusion mechanism.

2.2.1. Neighboring Band Partition

Previous works [14, 9] apply 3D convolution directly to hyperspectral image. Such distant adjacent bands are not explicitly guided by the current band, resulting in insufficient information fusion. This hinders access to effective complementary information from distant bands. To address this issue, we partition the current band and its adjacent bands into several groups based on the similarity to current band by neighboring band partition (NBP). Suppose only the case of $3 \le t \le L-3$ is considered. We regard the band I_t^{LR} that needs to be reconstructed as the current band, and divide the adjacent bands and current band into three groups by GP f_{GP}, i.e.,

$$f_{GP}(P(t)) = \begin{cases} I_t^{LR} \\ \left[I_{t-1}^{LR}, I_t^{LR}, I_{t+1}^{LR}\right], 3 \le t \le L-3. \\ \left[I_{t-2}^{LR}, I_t^{LR}, I_{t+2}^{LR}\right] \end{cases} \tag{3}$$

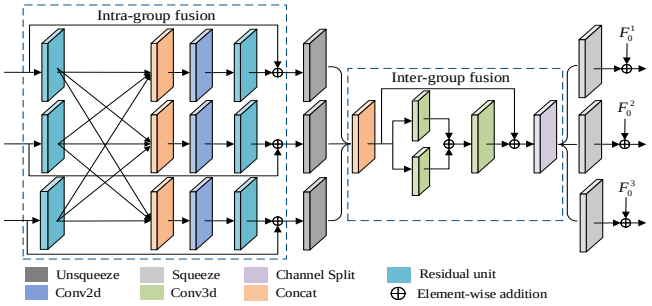

Fig. 2. The architecture of group fusion module (GFM).

Note that the current band I_t^{LR} appears in each group.

2.2.2. *Intra-group Fusion*

After constructing three groups, we input the corresponding initial features F_0^1, F_0^2, F_0^3 into three identical sub-networks to conduct intra-group fusion, respectively (see Fig. 2). Let take the first sub-network as an example. The sub-network within intra-group fusion (Intra_GF) mainly consists of three parts: residual unit, concatenation, and local skip connection. Specifically, the features F_0^1 are first imported into a residual unit. As for residual unit, it is made up of two identical blocks. Assuming that x is the input feature of block, the block is defined as

$$f_B(x) = f_{2D}(ReLU(f_{2D}(x))) + x, \tag{4}$$

where $f_{2D}(\cdot)$ represents 3×3 convolution operation, and $ReLU(\cdot)$ is the ReLU activation function. To study the valuable knowledge within group, the local features from other groups are then concatenated in this group, which is followed by 1×1 convolution layer. Finally, the group-wise features are produced through another residual unit and local skip connection. In this way, the current band guides the adjacent bands to explore potential spatial content by 2D convolution, improving the learning ability of spatial domain.

2.2.3. *Inter-group Fusion*

The existing networks [20] only applying 2D convolution cannot analyze information except for spatial dimension. Thus, to collect spectral features from different groups, we adopt 3D convolution to integrate the information by inter-group fusion (inter_GF) after performing *unsqueeze* operation. Since the regular 3D convolution yields a large number of parameters, we refer to [6, 12] and employ $3 \times 1 \times 1$ and $1 \times 3 \times 3$ to study the spectral and spatial features, respectively. Through an addition operation, the information between them is effectively fused, which enhances the ability of spatial exploration. For hyperspectral SR, its purpose is to improve spatial resolution while minimizing spectral distortion. Therefore, we utilize 3D convolution with size of $1 \times 3 \times 3$ again to mine more spatial features from the fused result. Besides, we add a local skip connection at the end. To conduct next intra-group fusion, the *channel split* is ultilized to generate three parts before *squeeze* operation. By integrating the features across groups, the inter-group fusion module can adaptively borrows complementary information, which helps to recover the more missing details.

2.3. Feature Context Fusion

As mentioned earlier, the hyperspectral image has a remarkable characteristic that high similarity exists in adjacent bands [18]. If the features produced in previous band are transferred to the reconstruction task of the current band, we can also get complementary information. Hence, feature context fusion (FCF) is introduced into our network, and two weights w_1 and w_2 are set to adaptively yield fusion data, i.e.,

$$C[w_1 * f_{FCF}(F_{DF}), w_2 * F_{PF}], \tag{5}$$

where C represents concatenation operation, and F_{PF} is the intermediate features generated in the previous band. f_{FCF} denotes FCF function. Intuitively, this approach is similar to the structure of RNN [21]. The structure helps the network to modulate the discriminative representations from other bands. Meanwhile, the ability to obtain complementary information is further improved.

3. EXPERIMENT

In this section, the implementation details and evaluation metrics of the algorithm are first provided. Then, the effectiveness of the model is analyzed by comparison and ablation.

3.1. Implementation Details and Evaluation Metrics

Two public datasets, CAVE [22] and Harvard [23], are adopted to evaluate the proposed method. We augment the data by randomly selecting 24 patches. Each patch is scaled by 1, 0.75, and 0.5, respectively, and these patches are rotated by 90° and horizontally flipped. Then, these patches are downsampled as LR hyperspectral images with the size of $L \times 32 \times 32$ by bicubic interpolation. In our work, we optimize SR network by minimizing $L1$ loss. The ADAM optimizer with β_1 = 0.9 and β_2 = 0.999 is utilized to train our model. The batch size is set to 64. The learning rate is initialized as 10^{-4} for all layers, which decreases by a half at every 30 epochs. To qualitatively assess the proposed method, three evaluations are used, including Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity (SSIM), and Spectral Angle Mapper (SAM).

3.2. Comparisons with the State-of-the-art Methods

In this section, we compare the proposed method on two benchmark datasets with three SR algorithms, namely GDRRN

Table 1. Quantitative evaluation of state-of-the-art algorithms by average PSNR/SSIM/SAM for different scales.

Dataset	Scale	GDRRN PSNR / SSIM / SAM	3D-FCNN PSNR / SSIM / SAM	MCNet PSNR / SSIM / SAM	Ours PSNR / SSIM / SAM
CAVE	×2	41.67 / 0.9651 / 3.84	43.15 / 0.9686 / 2.31	45.10 / 0.9738 / 2.24	**45.46 / 0.9742 / 2.22**
	×3	38.83 / 0.9401 / 4.54	40.22 / 0.9453 / 2.93	41.03 / 0.9526 / 2.81	**41.47 / 0.9528 / 2.78**
	×4	36.96 / 0.9166 / 5.17	37.63 / 0.9195 / 3.36	39.03 / 0.9319 / 3.24	**39.36 / 0.9324 / 3.22**
Harvard	×2	44.21 / 0.9775 / 2.28	44.45 / 0.9778 / 1.89	46.26 / 0.9827 / 1.88	**46.38 / 0.9834 / 1.87**
	×3	40.91 / 0.9523 / 2.62	40.59 / 0.9480 / 2.24	42.68 / 0.9627 / 2.21	**42.81 / 0.9635 / 2.16**
	×4	38.60 / 0.9259 / 2.79	38.14 / 0.9188 / 2.36	40.08 / 0.9367 / 2.41	**40.21 / 0.9372 / 2.37**

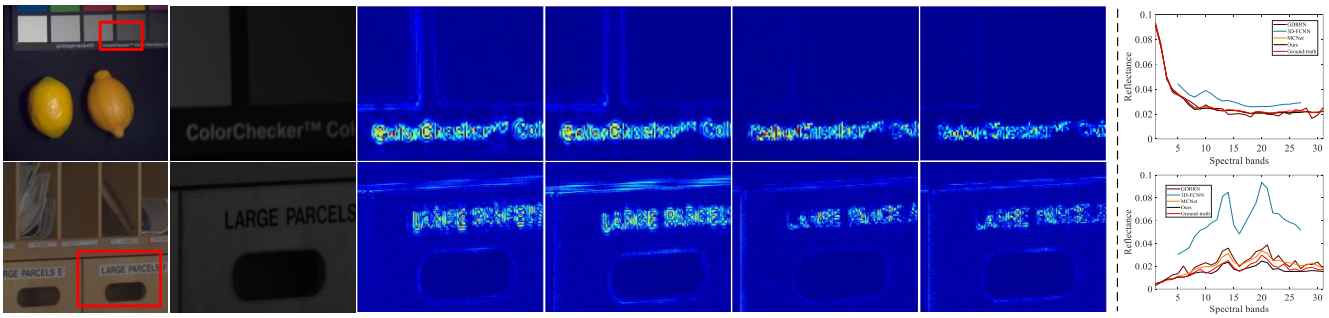

Fig. 3. Visual comparisons with different algorithms for ×4 SR on two datasets.

Table 2. Ablation study about the modules.

Module	Different combinations of modules				
Inter_GF	√	√	×	×	√
Intra_GF	×	×	√	√	√
FCF	×	√	×	√	√
PSNR	44.38	44.41	45.34	45.36	45.46
SSIM	0.9732	0.9731	0.9740	0.9741	0.9742
SAM	2.26	2.25	2.23	2.23	2.22

[7], 3D-FCNN [9], and MCNet [6]. Table 1 exhibits quantitative comparisons for different scales. Specifically, GDRRN reveals the worst performance, which is caused by the introduction of SAM in loss function. Compared with GDRRN, the overall results of 3D-FCNN and MCNet using 3D convolution are superior. It reveals that spectral information helps to improve performance. MCNet provides the second best results, which benefits from the effective mining of spatial and spectral features by mixed 2D/3D convolution. Since the output of some modules of MCNet is not used effectively, the performance is slightly worse. Among these competitors, our method attains excellent performance than other algorithms.

We also show visual comparisons on two typical scenes for 10-th band. As depicted in Fig. 3, the band of the hyperspectral image is grey. Thus, to show clearly, the absolute error map between reconstructed and ground-truth image is introduced. The figure shows that our method generates shallow edges or no edges in some regions, while other algorithms exhibit obvious texture information. Moreover, we randomly select a pixel of each image to analyze the reconstructed spectral distortion. Obviously, the curve of our method is approximately consistent with that of ground-truth in most cases.

3.3. Ablation Study

We investigate the influence of different combinations of modules on the performance. Table 2 shows the ablation study about modules for ×2 SR on CAVE dataset. We can observe that each module is an indispensable part for studying model. Concretely, only inter_GP or intra_GP exists in the network, there are significant differences in performance between the two. It is because intra_GP has a deeper network structure, which is beneficial to information mining. When the FCF is attached into the network, although the difference is not obvious, its performance is also slightly improved. Finally, all modules are added into the model. We can notice that all results in three evaluation metrics are superior to any other combinations. From these analyses, it reveals that each module contributes to network learning and optimization.

4. CONCLUSION

In this paper, we develop a new structure for hyperspectral image SR via adjacent spectral fusion strategy, claiming the following contributions: 1) the adjacent bands with high similarity to the reference band are divided into three groups, which guides the adjacent bands to explore content; 2) an alternative fusion mechanism is designed, which helps to obtain complementary information from neighboring bands. Experiments demonstrate that the proposed method achieves satisfactory results in both spatial reconstruction and spectral fidelity. In the future, we will improve our network by searching the optimal combination of intra-group and inter-group fusion through network architecture search.

5. REFERENCES

[1] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse dirichlet-net for hyperspectral image super-resolution," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2511–2520.

[2] B. Wen, U. S. Kamilov, D. Liu, H. Mansour, and P. T. Boufounos, "DeepCASD: An end-to-end approach for multi-spectral image super-resolution," in *IEEE Int. Conf. Acoust. Speech Signal Process Proc.*, 2018, pp. 6503–6507.

[3] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 1082–1096, 2020.

[4] Q. Li, Q. Wang, and X. Li, "Exploring the relationship between 2D/3D convolution for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sensing*, 2020.

[5] N. Akhtar, F. Shafait, and A. S. Mian, "Hierarchical beta process with gaussian process prior for hyperspectral image super resolution," in *European Conference on Computer Vision*, 2016, pp. 103–120.

[6] Q. Li, Q. Wang, and X. Li, "Mixed 2D/3D convolutional network for hyperspectral image super-resolution," *Remote Sens.*, vol. 12, no. 10, pp. 1660, 2020.

[7] Y. Li, L. Zhang, C. Ding, W. Wei, and Y. Zhang, "Single hyperspectral image super-resolution with grouped deep recursive residual network," in *IEEE Int. Conf. Multimed. Big Data*, 2018, pp. 1–4.

[8] Q. Huang, W. Li, T. Hu, and R. Tao, "Hyperspectral image super-resolution using generative adversarial network and residual learning," in *IEEE Int. Conf. Acoust. Speech Signal Process Proc.*, 2019, pp. 3012–3016.

[9] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3D full convolutional neural network," *Remote Sens.*, vol. 9, pp. 1139, 2017.

[10] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2472–2481.

[11] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2359–2368.

[12] Q. Wang, Q. Li, and X. Li, "Spatial-spectral residual network for hyperspectral image super-resolution.," *arXiv: 2001.04609*, 2020.

[13] P. Arun, K. M. Buddhiraju, A. Porwal, and C. Chanussot, "CNN-based super-resolution of hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, 2020.

[14] J. Yang, Y. Zhao, J. C. Chan, and L. Xiao, "A multi-scale wavelet 3D-CNN for hyperspectral image super-resolution," *Remote Sens.*, vol. 11, no. 13, pp. 1557, 2019.

[15] J. Li, R. Cui, Y. Li, B. Li, Q. Du, and C. Ge, "Multitemporal hyperspectral image super-resolution through 3D generative adversarial network," in *Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2019, pp. 1–4.

[16] S. Xie, C. Sun, J. Huang, Z. Tu, and K. Murphy, "Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification," in *Eur. Conf. Comput. Vis.*, 2018, pp. 318–335.

[17] J. Li, R. Cui, B. Li, Y. Li, S. Mei, and Q. Du, "Dual 1D-2D spatial-spectral cnn for hyperspectral image super-resolution," in *Int. Geosci. Remote Sen.s Symp.*, 2019, pp. 3113–3116.

[18] Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sensing*, 2020.

[19] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940–4950, 2019.

[20] J. Hu, X. Jia, Y. Li, G. He, and M. Zhao, "Hyperspectral image super-resolution via intrafusion network," *IEEE Trans. Geosci. Remote Sensing*, vol. 58, no. 10, pp. 7459–7471, 2020.

[21] T. Mikolov, S. Kombrink, L. Burget, J. Černockỳ, and S. Khudanpur, "Extensions of recurrent neural network language model," in *IEEE Int. Conf. Acoust. Speech Signal Process Proc.*, 2011, pp. 5528–5531.

[22] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, 2010.

[23] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 193–200.