Binarized Spectral Compressive Imaging

Anonymous NeurIPS supplementary submission

The source code and pre-trained models will be made publicly available for further research.

In this supplementary material, we share more details that are not in our main paper, including:

(a) Mathematical model of CASSI in Sec. 1

(b) More visual results in Sec. 2

(c) Limitation of our work in Sec. 3

(d) Broader impact in Sec. 4

(e) Code submission and reproducibility in Sec. 5

1 Mathematical Model of CASSI

The mathematical model of coded aperture snapshot spectral imaging (CASSI) is illustrated in Fig. 1. The 3D HSI cube is denoted as $\mathbf{F} \in \mathbb{R}^{H\times W\times N_\lambda}$, where H, W, and N_λ refer to the height, width, and total number of the wavelengths, respectively.

Firstly, $\mathbf{F}$ is modulated along the channel dimension by a pre-defined coded aperture (*i.e.*, a physical mask) $\mathbf{M}^* \in \mathbb{R}^{H\times W}$, which can be formulated as

$$\mathbf{F}'(:,:,n_\lambda) = \mathbf{F}(:,:,n_\lambda) \odot \mathbf{M}^*, \tag{1}$$

where $\mathbf{F}' \in \mathbb{R}^{H\times W\times N_\lambda}$ indicates the modulated HSIs, $n_\lambda \in [1,\dots,N_\lambda]$ indexes the spectral wavelengths, and $\odot$ denotes the element-wise multiplication.

Secondly, the modulated cube passes through the disperser that scatters the light to different spatial locations according to the wavelengths. This process makes $\mathbf{F}'$ change its shape and become tilted. Therefore, $\mathbf{F}'$ could be considered as sheared along the y-axis. The tilted HSI cube is represented as $\mathbf{F}'' \in \mathbb{R}^{H\times(W+d(N_\lambda-1))\times N_\lambda}$, where d denotes the step of spatial shifting. We assume that λ_c implies the reference wavelength and $\mathbf{F}''[:,:,n_{\lambda_c}]$ is the anchor image without being sheared along the y-axis. Subsequently, the dispersion process can be formulated as

$$\mathbf{F}''(u,v,n_\lambda) = \mathbf{F}'(x, y + d(\lambda_n - \lambda_c), n_\lambda), \tag{2}$$

where (u,v) represents the coordinate system on the detector array, λ_n implies the wavelength of the n_λ-th spectral channel, and $d(\lambda_n - \lambda_c)$ denotes the spatial shifting offset of the n_λ-th channel on $\mathbf{F}''$.

Thirdly, the 3D data cube is compressed into a 2D measurement by integrating the spectral signals across all the wavelengths. Suppose that the sensor integrates the whole light ranging from $\lambda_{\min}$ to $\lambda_{\max}$. Then the compressed measurement $y(u,v)$ can be formulated as

$$y(u,v) = \int_{\lambda_{\min}}^{\lambda_{\max}} f''(u,v,n_\lambda)\mathrm{d}\lambda, \tag{3}$$

where f'' indicates the continuous representation of $\mathbf{F}''$. Furthermore, to discretize Eq. (3), we denote the 2D compressed measurement as $\mathbf{Y} \in \mathbb{R}^{H\times(W+d(N_\lambda-1))}$. Then Eq. (3) can be reformulated as

$$\mathbf{Y} = \sum\nolimits_{n_\lambda=1}^{N_\lambda} \mathbf{F}''(:,:,n_\lambda) + \mathbf{E}, \tag{4}$$

where $\mathbf{E} \in \mathbb{R}^{H\times(W+d(N_\lambda-1))}$ implies the random imaging noise generated by the detector.

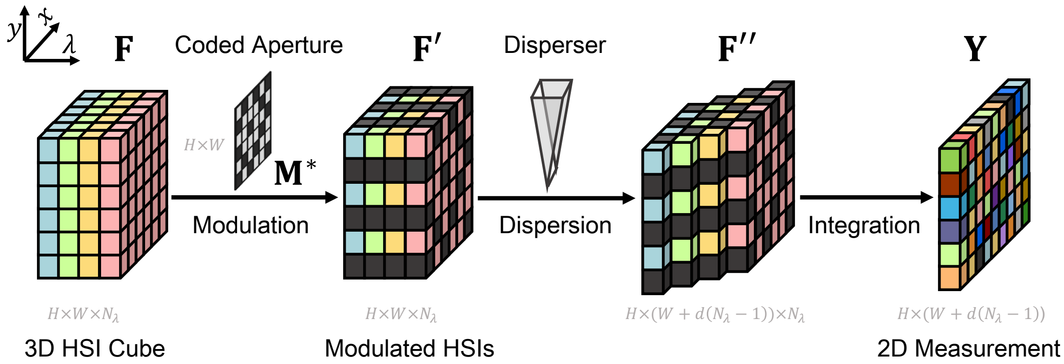

Figure 1: The schematic diagram of coded aperture snapshot spectral imaging (CASSI).

For the simplicity of model description, we denote $\mathbf{M} \in \mathbb{R}^{H\times(W+d(N_\lambda-1))\times N_\lambda}$ as the shifted version of the mask $\mathbf{M}^*$ corresponding to distinct wavelengths. Similarly, we define $\tilde{\mathbf{F}} \in \mathbb{R}^{H\times(W+d(N_\lambda-1))\times N_\lambda}$ as the shifted version of original HSI signal $\mathbf{F}$. Consequently, we have

$$\begin{aligned}\mathbf{M}(u, v, n_\lambda) &= \mathbf{M}^*(x, y + d(\lambda_n - \lambda_c)),\\ \tilde{\mathbf{F}}(u, v, n_\lambda) &= \mathbf{F}(x, y + d(\lambda_n - \lambda_c), n_\lambda).\end{aligned} \tag{5}$$

Following this, we can reformulate the measurement $\mathbf{Y}$ in Eq. (4) as

$$\mathbf{Y} = \sum_{n_\lambda=1}^{N_\lambda} \tilde{\mathbf{F}}(:, :, n_\lambda) \odot \mathbf{M}(:, :, n_\lambda) + \mathbf{E}. \tag{6}$$

Vectorization. To vectorize the matrices $\mathbf{Y}$ and $\mathbf{E}$, we set $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$ and $\mathbf{e} = \mathrm{vec}(\mathbf{E}) \in \mathbb{R}^n$, where $n = H(W + d(N_\lambda - 1))$ and $\mathrm{vec}(\cdot)$ indicates the operation of concatenating all columns of the matrix as a single vector. Define $\tilde{\mathbf{x}}^{(n_\lambda)} = \mathrm{vec}(\tilde{\mathbf{X}}(:, :, n_\lambda))$, thus the vector $\mathbf{x} = \mathrm{vec}([\tilde{\mathbf{x}}^{(1)}, \ldots, \tilde{\mathbf{x}}^{(N_\lambda)}]) \in \mathbb{R}^{nN_\lambda}$. Similarly, the sensing matrix is defined as

$$\mathbf{\Phi} = [\mathbf{D}_1, \ldots, \mathbf{D}_{N_\lambda}] \in \mathbb{R}^{n\times nN_\lambda}, \tag{7}$$

where $\mathbf{D}_{n_\lambda} = \mathrm{diag}(\mathrm{vec}(\mathbf{M}(:, :, n_\lambda)))$ is a diagonal matrix, of which the diagonal elements are $\mathrm{vec}(\mathbf{M}(:, :, n_\lambda))$. Following this, we can reformulate Eq. (6) into a vectorized version as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{e}. \tag{8}$$

Eq. (8) is similar to the compressive sensing [1, 2] task since $\mathbf{\Phi}$ is a fat matrix, which means there are more columns than rows in $\mathbf{\Phi}$. It is noticed that $\mathbf{\Phi}$ is highly sparse, with at most nN_λ nonzero elements. Although most existing compressive sensing theories can hardly work for our application due to the very special structure of $\mathbf{\Phi}$ as in Eq. (7), it has been proved that the HSI signal can still be reconstructed even if $N_\lambda > 1$ [3, 4, 5, 6].

For a given compressed measurement $\mathbf{y}$ captured by the camera and pre-designed sensing matrix $\mathbf{\Phi}$, one practical task of CASSI is to solve $\mathbf{x}$, which is also the core research problem of our work.

2 More Visual Results

Fig. 2 depicts the reconstructed HSIs with 5 out of 28 wavelengths of seven SOTA BNN-based methods (including BiConnect [7], BNN [8], Bi-Real [9], IRNet [10], ReActNet [11], BTM [12], and BBCU [13]) and our BiSRNet on simulation *Scene* 9, 6, 8, 3, and real *Scene* 1 from top to bottom. In simulation HSI restoration, other methods fail to restore high-frequency HSI contents. They tend to produce over-smoothed results sacrificing fine-grained details and structural textures, or introducing unpleasant artifacts. In contrast, our BiSRNet is more effective in producing perceptually-pleasing

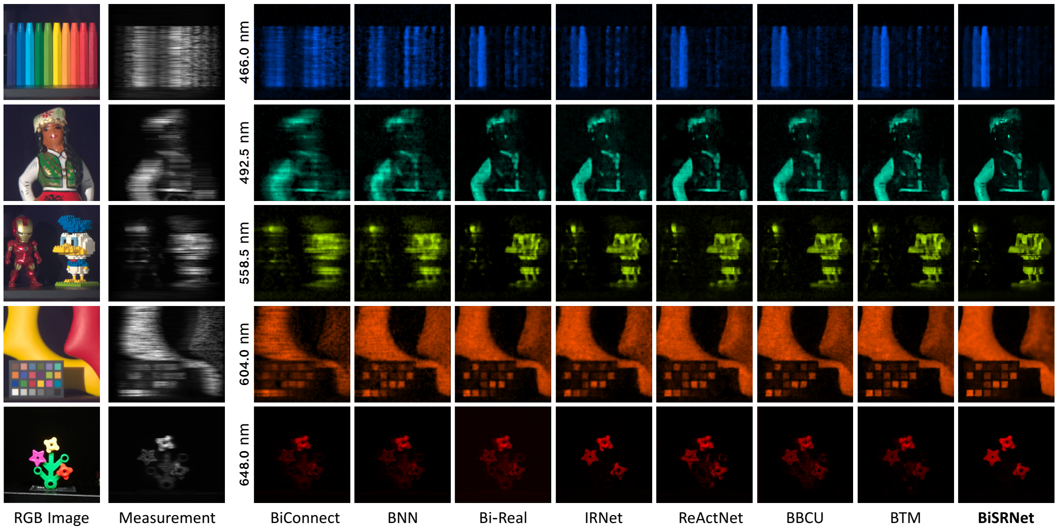

Figure 2: Reconstructed HSI results of 7 SOTA BNN-based algorithms and our BiSRNet with 5 out of 28 spectral wavelengths on simulation *Scene* 9, 6, 8, 3, and real *Scene* 1 from top to bottom. BiSRNet reconstructs more detailed contents and suppresses more real noise. Please zoom in for better visualization performance.

and sharp images, and maintaining the spatial smoothness of the homogeneous regions without introducing artifacts. Besides, in real HSI reconstruction, our BiSRNet is superior to other methods in fine-grained content reconstruction, spectral density responses, and real noise suppression. These results demonstrate the effectiveness, robustness, and generalization ability of our BiSRNet.

3 Limitation

The limitation of our work is that the model binarization sacrifices the HSI reconstruction performance. More specifically, compared to the full-precision counterpart, our BiSRNet is 4.35 (34.11 - 29.76) dB lower in PSNR and 0.099 (0.936 - 0.837) lower in SSIM. The PSNR and SSIM are reduced by 12.8% and 10.6%, respectively. However, this performance drop is smaller than that of other model binarization methods. To handle this issue, we will study how to preserve more performance while reducing the memory and computational complexity as much as possible in model binarization.

4 Broader Impact

HSI reconstruction is one of the core tasks in snapshot compressive imaging (SCI) and has been studied for decades. Compared with normal RGB images, HSIs have more spectral bands to store richer information of the desired scenes. Hence, HSIs are widely applied in many computer vision related tasks, such as medical imaging [14, 15, 16, 17], object tracking [18, 19, 20, 21], remote sensing [22, 23, 24, 25], and so on. Nowadays, billions of 3D HSIs are compressed by SCI systems. Therefore, how to reconstruct the original 3D HSI signal from the 2D compressed measurement is worth studying. Our BiSRNet is capable of reconstructing HSIs more efficiently and accurately than all existing SOTA BNN-based methods, showing great value in practical applications.

Until now, HSI reconstruction techniques have no negative social impact yet. Our proposed BiSRNet does not present any negative foreseeable societal consequences, either.

5 Code Submission and Reproducibility

We provide the **source code** and **pre-trained** models to reproduce the main results in Table 1, Figure 5, and Figure 5 of our paper. Please refer to the folder ‘code’ and read the file ‘README.md’ for detailed instructions. **The source code and pre-trained models will be released to the public.**

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, 2006.

[2] C. Emmanuel, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, 2006.

[3] S. Jalali and X. Yuan, "Compressive imaging via one-shot measurements," in *IEEE International Symposium on Information Theory (ISIT)*, 2018.

[4] S. Jalali and X. Yuan, "Snapshot compressed sensing: Performance bounds and algorithms," *IEEE Transactions on Information Theory*, 2019.

[5] Z. Meng, J. Ma, and X. Yuan, "End-to-end low cost compressive spectral imaging with spatial-spectral self-attention," in *ECCV*, 2020.

[6] T. Huang, W. Dong, X. Yuan, J. Wu, and G. Shi, "Deep gaussian scale mixture prior for spectral compressive imaging," in *CVPR*, 2021.

[7] M. Courbariaux, Y. Bengio, and J.-P. David, "Binaryconnect: Training deep neural networks with binary weights during propagations," in *NeurIPS*, 2015.

[8] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Binarized neural networks," in *NeurIPS*, 2016.

[9] Z. Liu, B. Wu, W. Luo, X. Yang, W. Liu, and K.-T. Cheng, "Bi-real net: Enhancing the performance of 1-bit cnns with improved representational capability and advanced training algorithm," in *ECCV*, 2018.

[10] H. Qin, R. Gong, X. Liu, M. Shen, Z. Wei, F. Yu, and J. Song, "Forward and backward information retention for accurate binary neural networks," in *CVPR*, 2020.

[11] Z. Liu, Z. Shen, M. Savvides, and K.-T. Cheng, "Reactnet: Towards precise binary neural network with generalized activation functions," in *ECCV*, 2020.

[12] X. Jiang, N. Wang, J. Xin, K. Li, X. Yang, and X. Gao, "Training binary neural network without batch normalization for image super-resolution," in *AAAI*, 2021.

[13] B. Xia, Y. Zhang, Y. Wang, Y. Tian, W. Yang, R. Timofte, and L. V. Gool, "Basic binary convolution unit for binarized image restoration network," in *ICLR*, 2023.

[14] V. Backman, M. B. Wallace, L. Perelman, J. Arendt, R. Gurjar, M. Muller, Q. Zhang, G. Zonios, E. Kline, and T. McGillican, "Detection of preinvasive cancer cells," *Nature*, 2000.

[15] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, 2014.

[16] Z. Meng, M. Qiao, J. Ma, Z. Yu, K. Xu, and X. Yuan, "Snapshot multispectral endomicroscopy," *Optics Letters*, 2020.

[17] W. R. Johnson, D. W. Wilson, W. Fink, M. Humayun, and G. Bearman, "Snapshot hyperspectral imaging in ophthalmology," *Journal of biomedical optics*, 2007.

[18] M. H. Kim, T. A. Harvey, D. S. Kittle, H. Rushmeier, J. Dorsey, R. O. Prum, and D. J. Brady, "3d imaging spectroscopy for measuring hyperspectral patterns on solid objects," *TOG*, 2012.

[19] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *TPAMI*, 2003.

[20] H. V. Nguyen, A. Banerjee, and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," in *CVPRW*, 2010.

[21] H. Jin, P. Favaro, and R. Cipolla, "Visual tracking in the presence of motion blur," in *CVPR*, 2005.

[22] M. Borengasser, W. S. Hungate, and R. Watkins, "Hyperspectral remote sensing: principles and applications," *CRC press*, 2007.

[23] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, 2004.

[24] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017.

[25] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, 1985.