Aperture Diffraction for Compact Snapshot Spectral Imaging
—— Supplementary Material ——

1. Overview

In the supplementary material, we first provide a comprehensive account of imaging model in section 2, elucidating the variances in the PSF based on different parameters. Then We expound on the test set employed in the simulation experiments of this paper in section 3. Additionally, detailed discussion and comparison of simulation experiments at lower exposures are presented in Section 4. Ultimately, additional results of ADIS's reconstruction from the real acquisitions are illustrated in section 5.

2. Detailed process of imaging model

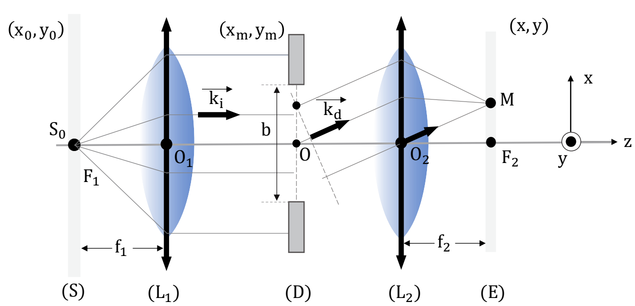

Figure 1. Simplified schematic of the ADIS's profile

Imaging forward model. We now consider a multi-slit aperture comprising $N \times N$ parallel rectangular apertures, with each rectangular aperture having a width of a, a length of b, and a center-to-center distance between adjacent slits of d. The Fraunhofer diffraction formula, a fundamental calculation method in optics, is utilized to characterize the diffraction phenomenon when light passes through an aperture. When light passes through a finite-sized aperture, it generates a series of interference and diffraction patterns within the far-field region. The Huygens-Fresnel principle states that each point on a wave surface can be treated as a new secondary wave source, and the wave surface can be considered as a superposition of spherical waves emitted by an infinite number of point sources.

Linear systems possess an essential characteristic known as the principle of superposition, which asserts that the resulting output of a linear system, when multiple input signals are applied, is the linear superposition of these input signals. This principle is applicable to the phenomenon of aperture mask diffraction, wherein each aperture behaves as a point source. The waves emanating from every point source combine constructively and destructively to produce the output wave of the entire aperture. This process of wave superposition is linear, implying that the spatial distribution of the output wave corresponds to the superposition of waves generated by all point sources as the number of apertures increases. Consequently, each rectangular aperture in the aperture mask can be viewed as a point source, and the waves produced by all point sources can be added coherently to yield the diffraction pattern across the entire aperture mask.

The initial form of Fraunhofer diffraction is:

$$E_p = c \int\limits_A e^{ikr} dA \tag{1}$$

Considering a single square hole mask with a width of a and a length of b, we can write its imaging distribution on the Fourier surface as:

$$E_p = c \int_0^b \int_0^a e^{ik(r_0 + x\sin\phi + y\sin\theta)} dx dy \tag{2}$$

Therefore, for a multi-slit mask comprising $N \times N$ parallel rectangular apertures, we can write the diffraction formula as follows:

$$\begin{aligned} E_p = c e^{ikr_0} &\left[\int_0^b e^{iky\sin\theta_1} dy + \cdots + \int_{(N-1)d}^{(N-1)d+b} e^{iky\sin\theta_1} dy \right] \\ &\times \left[\int_0^a e^{ikx\sin\theta_2} dx + \cdots + \int_{(N-1)d}^{(N-1)d+a} e^{ikx\sin\theta_2} dx \right] \end{aligned} \tag{3}$$

Calculated to get:

$$\begin{aligned} E_p = c e^{ikr_0} &\frac{e^{ikb\sin\theta_1} - 1}{ik\sin\theta_1} \times \frac{1 - e^{ikNd\sin\theta_1}}{1 - e^{ikd\sin\theta_1}} \\ &\times \frac{e^{ika\sin\theta_2} - 1}{ik\sin\theta_2} \times \frac{1 - e^{ikNd\sin\theta_2}}{1 - e^{ikd\sin\theta_2}} \end{aligned} \tag{4}$$

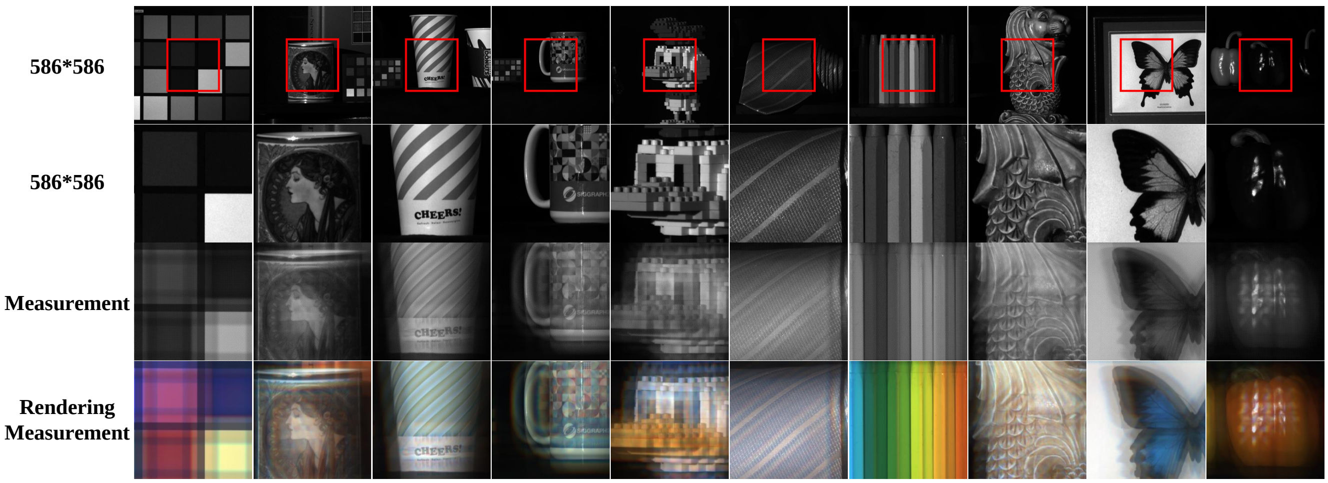

Figure 2. Illustrations of the test set and the rendering of the generated measurements

To simplify the parameters, let:

$$\beta_1 = \frac{1}{2}kb\sin\theta_1, \beta_2 = \frac{1}{2}ka\sin\theta_2,$$
$$\gamma_1 = \frac{1}{2}kd\sin\theta_1, \gamma_2 = \frac{1}{2}kd\sin\theta_2 \tag{5}$$

$$E_p = ce^{ikr_0}ab\frac{e^{2i\beta_1}-1}{2i\beta_1} \times \frac{1-e^{2iN\gamma_1}}{1-e^{2i\gamma_1}} \times \frac{e^{2i\beta_2}-1}{2i\beta_1} \times \frac{1-e^{2iN\gamma_2}}{1-e^{2i\gamma_2}} \tag{6}$$

Then we get:

$$E_p = E_0\frac{\sin\beta_1}{\beta_1}\frac{\sin N\gamma_1}{\sin\gamma_1}\frac{\sin\beta_2}{\beta_2}\frac{\sin N\gamma_2}{\sin\gamma_2} \tag{7}$$

Since focusing is performed under paraxial conditions, the formula $\sin\theta_1 \approx \tan\theta_1 = \frac{x_m}{f_2}$, $\sin\theta_2 \approx \tan\theta_2 = \frac{y_m}{f_2}$:

$$I(x_m, y_m, \lambda) = I_0 \cdot D(x_m, y_m, \lambda) \cdot P(x_m, y_m, \lambda) \tag{8}$$

$$D(x_m, y_m, \lambda) = \sin c^2(\frac{\pi b}{\lambda f_2}x_m)\sin c^2(\frac{\pi a}{\lambda f_2}y_m) \tag{9}$$

$$P(x_m, y_m, \lambda) = \left[\frac{\sin(N\frac{\pi d}{\lambda f_2}x_m)}{\sin(\frac{\pi d}{\lambda f_2}x_m)}\right]^2 \times \left[\frac{\sin(N\frac{\pi d}{\lambda f_2}y_m)}{\sin(\frac{\pi d}{\lambda f_2}y_m)}\right]^2 \tag{10}$$

Where the formula $D(x_m, y_m, \lambda)$ is the diffraction factor describes the diffraction effect of each rectangular square hole. $P(x_m, y_m, \lambda)$ is the interference factor describes the effect of multi-slit interference. (x_m, y_m) denotes the spatial coordinates on the receiving screen, while f_2 denotes the distance between the diffraction array and the sensor.

Finally, we get the formula under orthogonal aperture diffraction:

$$I = I_0\left(\frac{\sin\beta_1}{\beta_1}\right)^2\left(\frac{\sin N\gamma_1}{\sin\gamma_1}\right)^2\left(\frac{\sin\beta_2}{\beta_2}\right)^2\left(\frac{\sin N\gamma_2}{\sin\gamma_2}\right)^2 \tag{11}$$

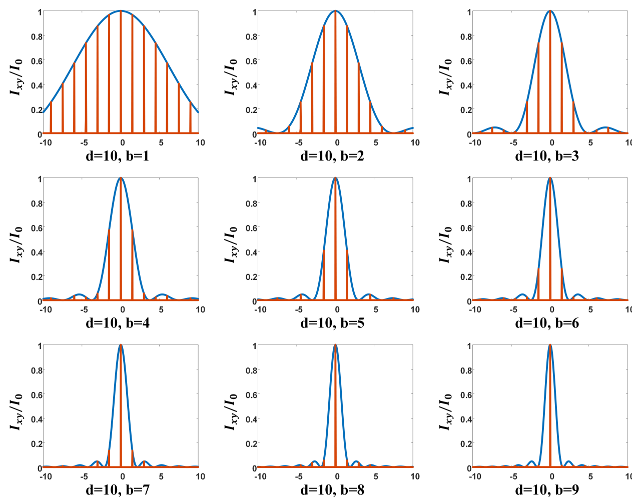

Figure 3. The intensity distribution of light on the x-axis for various mask parameters is represented by the red curve.

Mask with different b/d**.** According to Equation 8, the contrast in intensity between the zero-order diffraction and the first-order diffraction is entirely reliant on the ratio between the aperture opening and the spacing of the square holes. To simulate different ratios, we designed masks and present the simulation results under various aperture mask parameters (b/d) in Figure 3. After comparing the diffraction patterns, we selected the parameter values of $d = 10$ and $a = b = 5$ for the mask. This aperture mask exhibits a first-order diffraction intensity that is half of the zero-order diffraction and a second-order diffraction that is precisely situated in the suppressed region, creating a missing order.

3. Test set of simulation experiment

Here we show our test set of 10 scenes selected from the KAIST [11] dataset as depicted in Figure 2. The 256*256

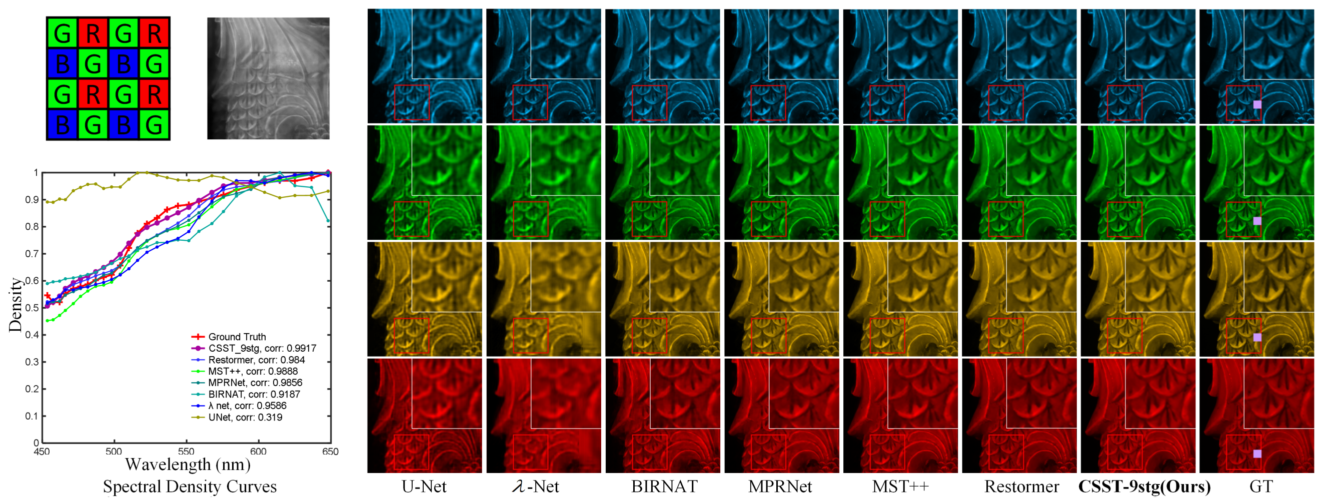

Figure 4. Qualitative comparison of reconstruction results of different algorithms at low exposure. Zoomed-in patches of the HSI in the fuchsia box are presented in the lower-left of the figure.

Algorithms	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Avg
U-Net [1]	21.43	23.89	19.87	18.82	20.72	22.08	21.16	28.51	28.40	22.31	22.72
	0.8162	0.8157	0.7554	0.5626	0.7269	0.7800	0.7894	0.8505	0.7493	0.7965	0.764
HSCN+ [2]	24.35	25.54	21.53	25.14	19.37	24.85	23.94	28.65	23.53	19.68	23.66
	0.8074	0.8078	0.7313	0.7932	0.6931	0.8063	0.7795	0.8615	0.6281	0.7324	0.764
HDNet [3]	22.42	25.62	20.72	19.49	23.38	23.68	24.43	29.21	30.43	22.50	24.19
	0.8502	0.8428	0.7964	0.6112	0.7958	0.8157	0.8193	0.8730	0.8328	0.8342	0.807
BIRNAT [4]	25.18	26.49	22.57	20.99	18.34	24.94	24.45	30.03	29.44	22.92	24.54
	0.8664	0.8533	0.7990	0.7376	0.7467	0.8297	0.8267	0.8884	0.7918	0.8365	0.818
MIRNet [5]	22.87	27.16	22.69	25.74	19.08	23.85	25.45	30.22	30.27	23.27	25.06
	0.7794	0.8375	0.7735	0.7848	0.7285	0.8149	0.8061	0.8908	0.8024	0.8139	0.803
lambda-Net [6]	29.65	27.19	24.67	24.70	24.89	25.61	26.65	31.16	33.79	23.72	27.20
	0.8943	0.8323	0.8052	0.5535	0.7772	0.7412	0.8056	0.8711	0.8998	0.8167	0.800
MPRNet [7]	29.25	29.84	25.68	29.12	26.99	27.58	26.63	32.74	33.45	26.63	28.79
	0.9167	0.9157	0.8925	0.8897	0.8799	0.8782	0.8604	**0.9723**	0.9030	0.9096	0.897
TSA-Net [8]	29.58	29.22	25.88	28.18	27.65	27.60	27.55	32.76	34.25	25.49	28.82
	0.9240	0.9028	0.8833	0.8757	0.8834	0.8792	0.8633	0.9232	0.8981	0.8826	0.892
MST++ [9]	32.52	30.76	26.24	28.68	28.01	28.24	25.81	33.31	36.24	27.45	29.70
	0.9426	0.9175	0.9076	0.8911	0.8959	0.9067	0.8793	0.9387	0.9309	0.9248	0.914
Restormer [10]	32.86	30.57	26.99	29.85	28.26	28.52	28.78	33.76	36.46	26.68	30.32
	0.9555	0.9303	0.9133	0.8969	0.9073	0.9095	0.8925	0.9436	0.9371	0.9277	0.921
CSST-9stg (Ours)	**34.18**	**33.47**	**29.20**	**30.76**	**30.79**	**30.53**	**29.36**	**35.84**	**38.55**	**28.87**	**32.16**
	0.9623	**0.9632**	**0.9477**	**0.9178**	**0.9296**	**0.9450**	**0.9056**	0.9630	**0.9610**	**0.9461**	**0.944**

Table 1. Quantitative comparison of reconstruction results of different algorithms at low exposure, PSNR (dB) and SSIM are reported.

measurements that can be generated from the 586*586*28 HSI by PSF. Meanwhile, to improve the visualization of the ADIS dispersion pattern, we perform RGB interpolation to render the measurements.

4. Simulation Experiments (Low Exposure)

Unlike the main text's simulation experiments conducted with regular exposure, here, the measurements' amplitude is scaled down to approximately one-fourth of the original to simulate varying exposure conditions. Similar to [8, 12, 13, 14, 15, 16, 17], 28 wavelengths are selected from 450nm to 650nm and derived by spectral interpolation manipulation for the HSI data.

Simulation Dataset. We adopt two datasets, i.e., CAVE-1024 [8] and KAIST [11] for simulation experiments. The CAVE-1024 consists of 205 HSIs with spatial size 1024×1024 obtained by interpolating and splicing from the CAVE [18] dataset. The KAIST dataset contains 30 HSIs of spatial size 2704×3376. 10 scenes from the KAIST dataset are selected for testing, while the CAVE-1024 dataset and another 20 scenes from the KAIST dataset are selected for training.

Implementation Details. The dispersion step of the primary diffraction is 0.5 spatial pixels, while the simulation experiment is deployed in the range of $400nm$ to $670nm$, which means that $586 \times 586 \times 28$ data cubes are needed to generate 256×256 resolution measurements for conducting experiments while preserving the tertiary diffraction. We

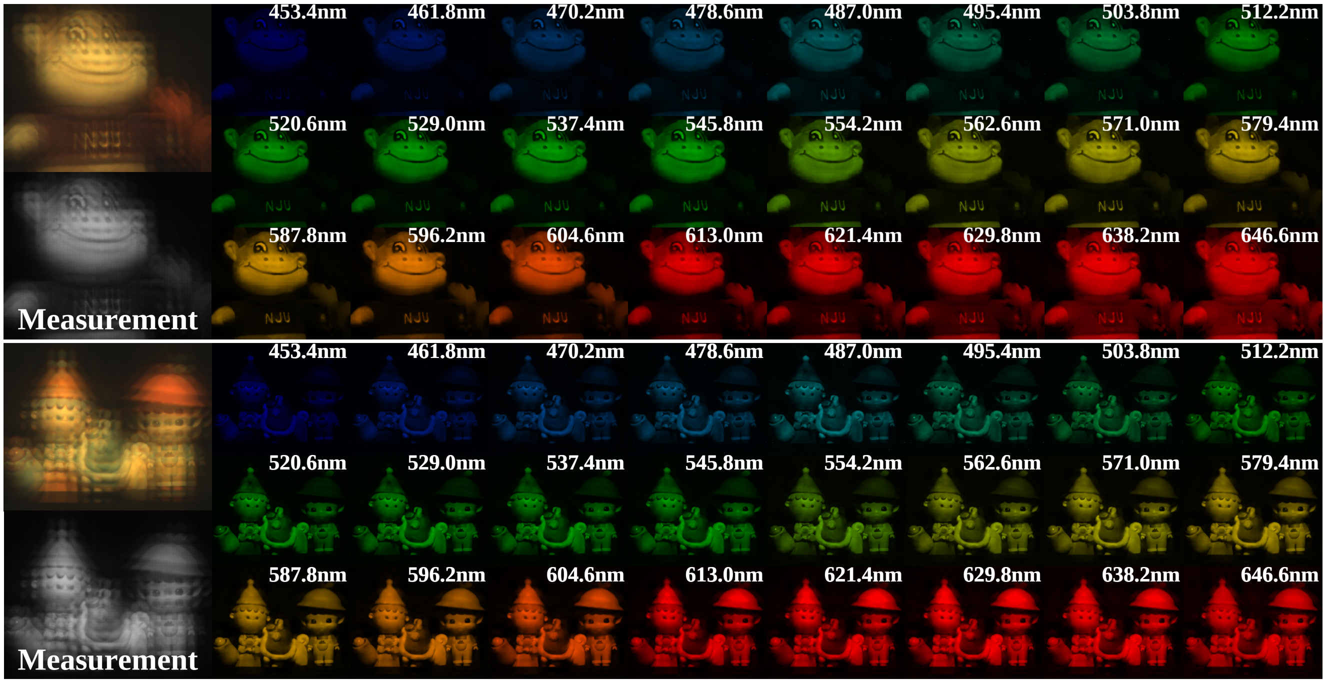

Figure 5. Additional results of ADIS's reconstruction from the real acquisitions

implement CSST by Pytorch. All CSST models are trained with Adam [19] optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) using Cosine Annealing scheme [20] for 300 epochs on an RTX 3090 GPU. The initial learning rate is 4×10^{-4}.

Quantitative Analysis. Table 1 compares the results of CSST and 10 methods including one baseline method(Unet [1]), six reconstruction methods(lambda-Net [6], HDNet [3], BIRNAT [4], TSA-Net [8], HSCNN+ [2] and MST++ [9]), three Super-resolution algorithms (Restormer [10], MPRNet [7], MIRNet[5]) on 10 simulation scenes at low exposure. CSST shows the best experimental results on the ADIS spectral reconstruction task, i.e., 32.16dB in PSNR and 0.944 in SSIM. CSST-9stg significantly outperforms two recent SOTA methods Restormer and MST++ by 1.84dBdB and 2.46dB, demonstrating stronger reconstruction performance compared to previous methods under low exposure conditions and robustness against exposure variations.

Qualitative Analysis. Figure 4 illustrates the comparative performance of our CSST and other methods in the HSI reconstruction of ADIS on the same scene at low exposure. Visual inspection of the image reveals that the CSST-9stg method provides more intricate details, sharper textures, and well-defined structures. Conversely, the previous approaches produce either overly smooth results that compromise the underlying structure or introduce color artifacts and speckled textures. Moreover, the lower left corner of the figure presents the spectral profile of the intensity-wavelength corresponding to the fuchsia square.

5. Additional real reconstruction results

Here we further show the reconstruction outcomes of various scenes captured by ADIS in the figure 5. These results exhibit distinct textures and well-structured edges, thereby corroborating the efficacy of ADIS in snapshot sub-super pixel resolution spectral imaging.

References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 3, 4

[2] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947, 2018. 3, 4

[3] Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Hdnet: High-resolution dual-domain learning for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17542–17551, 2022. 3, 4

[4] Ziheng Cheng, Bo Chen, Ruiying Lu, Zhengjue Wang, Hao Zhang, Ziyi Meng, and Xin Yuan. Recurrent neural networks for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 3, 4

[5] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020. 3, 4

[6] 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. 3, 4

[7] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 3, 4

[8] 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. 3, 4

[9] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 745–755, 2022. 3, 4

[10] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 3, 4

[11] Inchang Choi, MH Kim, D Gutierrez, DS Jeon, and G Nam. High-quality hyperspectral reconstruction using a spectral prior. Technical report, 2017. 2, 3

[12] Xin Yuan, David J Brady, and Aggelos K Katsaggelos. Snapshot compressive imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 38(2):65–88, 2021. 3

[13] Ziyi Meng, Shirin Jalali, and Xin Yuan. Gap-net for snapshot compressive imaging. *arXiv preprint arXiv:2012.08364*, 2020. 3

[14] 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. 3

[15] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17502–17511, 2022. 3

[16] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Coarse-to-fine sparse transformer for hyperspectral image reconstruction. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII*, pages 686–704. Springer, 2022. 3

[17] Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc Van Gool. Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. *arXiv preprint arXiv:2205.10102*, 2022. 3

[18] Jong-Il Park, Moon-Hyun Lee, Michael D Grossberg, and Shree K Nayar. Multispectral imaging using multiplexed illumination. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007. 3

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[20] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 4