Learned Multi-aperture Color-coded Optics for Snapshot Hyperspectral Imaging - Supplemental Document -

ZHENG SHI*, Princeton University, USA XIONG DUN*, Tongji University, China HAOYU WEI, The University of Hong Kong, China SHIYU DONG, ZHANSHAN WANG, XINBIN CHENG, Tongji University, China FELIX HEIDE, Princeton University, USA YIFAN PENG, The University of Hong Kong, China

ACM Reference Format:

Zheng Shi, Xiong Dun, Haoyu Wei, Shiyu Dong, Zhanshan Wang, Xinbin Cheng, Felix Heide, and Yifan Peng. 2024. Learned Multi-aperture Colorcoded Optics for Snapshot Hyperspectral Imaging - Supplemental Document - . *ACM Trans. Graph.* 43, 6, Article 208 (December 2024), 9 pages. https: //doi.org/10.1145/3687976

In this Supplemental Document, we present additional results and details on the proposed method in the main manuscript. We will release lens designs, color filter designs, network checkpoints, and all code needed to reproduce the results presented in the manuscript.

Contents

Cor	ntents	1
1	Fabrication Details	1
2	Additional Calibration Details	1
3	Neural Network Details	2
4	Additional Simulation Results	2
5	Additional Experimental Results	5
Ref	erences	9

1 FABRICATION DETAILS

DOE fabrication: The DOE array is fabricated using grayscale lithography and a molding process [Ikoma et al. 2021], closely following the methodology from prior work [Dun et al. 2020]. The

0730-0301/2024/12-ART208

https://doi.org/10.1145/3687976

process begins by spin-coating a positive photoresist (AZ-1512, MicroChemicals) onto a titanium-coated glass substrate. The photoresist is patterned using a direct-write optical grayscale lithography machine (MicroWriter ML3, Durham Magneto Optics). After patterning, the photoresist is developed with a base developer (MF-319, Microposit), creating a mold for replication onto polydimethylsiloxane (PDMS, SYLGARD184, Dow). The PDMS, mixed in a 10:1 ratio and degassed in a vacuum, is cured at room temperature for three days with the master mold, resulting in an elastomeric mold.

Next, four sub-circular apertures with a 3 mm diameter and a focal length of 20 mm are fabricated using a light-blocking chromium trilayer (200 nm) on a 3 mm thick float glass substrate (30-773, Edmund Optics) through the lift-off process. A drop of optically-clear UVcurable resin (NOA61, Norland Products) is applied between the PDMS mold and the glass substrate. After aligning the pattern with the circular apertures, the resin is cured using a mercury vapor lamp. The PDMS mold is then gently peeled off, leaving a patterned layer of clear resin on the glass substrate. Fig.S1 shows a microscope image of the fabricated DOE.

The four optimized sub-lens patterns are imprinted on the NOA 61 material filmed over the one-piece glass substrate, with a Cr layer serving as the sub-aperture array to block unwanted light.

Color filters fabrication: The color filters are fabricated through an iterative coating process involving alternating layers of SiO₂ and TiO₂, as detailed in [Dong et al. 2018]. Each filter consists of 20 layers, and the thickness of these layers dictates the filter's spectral response. In order to achieve the learned spectral responses, each of the four aperture-specific filters requires a tailored thinfilm structure – specifically, a unique set of 20 layer thicknesses, as detailed in Tab. S1. To maintain precision, the thickness of each dielectric layer (SiO₂ and TiO₂) is indirectly controlled by quartz crystal monitoring during the coating process. As such, a single continuous coating process yields only one type of filter, and the complete set is produced through four separate coating cycles and subsequently diced for attachment to the corresponding DOEs.

2 ADDITIONAL CALIBRATION DETAILS

PSF Calibration: The point spread function (PSF) of the fabricated DOE array is calibrated using an iHR 320 monochromator paired with a white light point source. A 100 mm Thorlabs asphere lens is used to collimate the light emitted from the monochromator. To

^{*}denotes equal contribution. Tongji University is the first affiliation. Corresponding authors: evanpeng@hku.hk, dunx@tongji.edu.cn.

Authors' addresses: Zheng Shi, zhengshi@princeton.edu, Princeton University, USA; Xiong Dun, dunx@tongji.edu.cn, Tongji University, China; Haoyu Wei, haoyuwei@c onnect.hku.hk, The University of Hong Kong, China; Shiyu Dong, Zhanshan Wang, Xinbin Cheng, dongsy@tongji.edu.cn, wangzs@tongji.edu.cn, chengxb@tongji.edu.cn, Tongji University, China; Felix Heide, fheide@princeton.edu, Princeton University, USA; Yifan Peng, evanpeng@hku.hk, The University of Hong Kong, China.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). © 2024 Copyright held by the owner/author(s).

ensure proper focus on the light source, we slightly adjust the gap between the DOE array and the sensor. The captured PSFs exhibit residual artifacts, which we correct by applying a rotational symmetry approximation to the images. Fig. S2 displays the measured PSFs (top row), the processed PSFs (middle row), and the designed PSFs



Fig. S1. Microscope image of the fabrication DOE that consists of 4 sub-lens patterns. The scale bar at the bottom-right: is 0.4 mm.

Table S1.	Thickness	Specifications for	r Designed Colo	or Filters

Number	Material	Filter 1	Filter 2	Filter 3	Filter 4
1	TiO2	169.35nm	60.94nm	60.94nm 58.77nm	
2	SiO2	115.49nm	109.77nm	74.82nm	137.92nm
3	TiO2	69.59nm	93.29nm	133.98nm	112.53nm
4	SiO2	98.96nm	103.29nm	79.92nm	105.01nm
5	TiO2	78.14nm	200.00nm	45.55nm	129.92nm
6	SiO2	92.71nm	108.31nm	113.27nm	139.42nm
7	TiO2	67.72nm	200.00nm	112.40nm	180.61nm
8	SiO2	98.53nm	88.29nm 113.00nm		114.83nm
9	9 TiO2		159.12nm	56.76nm	20.00nm
10	SiO2	83.37nm	64.78nm	111.96nm	89.39nm
11	TiO2	54.55nm	113.57nm	170.24nm	54.41nm
12	SiO2	116.98nm	87.51nm	137.54nm	83.72nm
13	TiO2	64.05nm	135.81nm	63.09nm	47.06nm
14	SiO2	69.90nm	80.32nm	80.43nm	70.39nm
15	TiO2	143.80nm	200.00nm	106.27nm	38.41nm
16	SiO2	58.55nm	99.11nm	96.47nm	81.78nm
17	TiO2	73.59nm	80.10nm	148.76nm	51.97nm
18	SiO2	108.18nm	101.68nm	70.82nm	86.84nm
10	TiO2	36.73nm	196.19nm	112.68nm	134.97nm
20	SiO2	20.00nm	20.00nm	174.67nm	20.03nm

ACM Trans. Graph., Vol. 43, No. 6, Article 208. Publication date: December 2024.

(bottom row) at each design wavelength, ranging from 429 to 700 nm. Image intensity levels are normalized to enhance visualization. Compared to state-of-the-art industry lithography, our fabrication is an experimental fabrication process, and, as such, a difference between the measured PSFs and the target PSFs is observable. Moving forward, this discrepancy could potentially be resolved through higher-precision electron-beam grayscale lithography [Zhu et al. 2023].

Color Filter Calibration: We calibrate the spectral response curve of the color filters to validate the fabrication. Fig. S3 displays the design, fabrication target, and measured spectral response curves for the B, G1, G2, and R channel color filters.

Noise Calibration: We model the sensor noise as pixel-wise Gaussian-Poisson noise and estimate the parameters using the calibration method described in [Foi et al. 2008]. We determine $\sigma_g = 1e^{-3}$ for Gaussian noise and $\sigma_p = 2.5e^{-5}$ for Poisson noise.

Impact of Fabrication Deviation: To address fabrication deviations, we finetuned the reconstruction model using the measured PSFs and spectral response curve for experimental validation. While this finetuning improves performance, it cannot entirely mitigate the impact of fabrication inaccuracies. Post-finetuning, the model has achieved a PSNR of 29.49dB on the CAVE dataset. We note that, while it still surpasses the existing baselines by over 2dB, this result is 3.34dB lower than that achieved by our proposed design.

3 NEURAL NETWORK DETAILS

As discussed in the main text, the proposed reconstruction network has a feature-extraction block for multi-scale feature extraction and two reconstruction heads for RGB or HSI output. The featureextraction block uses parallel convolution streams at varying resolutions, alongside concatenation layers and skip connections to better retain high-frequency details. The reconstruction heads, equipped with 5×5 kernels, focus on RGB or HSI reconstruction. See Tab. S2 for detailed specifications of the network. This table compensates for the image reconstruction unit as visualized in Fig. 1 of the main manuscript. Refer to Fig. S6 for the impact of different architectural components, including the multi-scale feature extraction, skip connections, and separate reconstruction heads.

4 ADDITIONAL SIMULATION RESULTS

Error Map Visualization of the Multi-Aperture Configuration Analysis. Beyond the qualitative reconstruction presented in Fig. 3 of the main manuscript, we provide also a higher-resolution error map for enhanced clarity, as shown in Fig. S4. The absolute reconstruction error for RGB and each spectral channel is illustrated as a grayscale image, with the intensity scaled by a factor of 1.5 to for better readability.

Additional Ablation Study on Color Filter Initialization. The proposed method employed a standard RGGB Bayer filter for color filter initialization, allowing the end-to-end optimization process to learn shifts in the primary peak locations of the identical G curves to enhance spectral detail capture. We further evaluated two alternative

Learned Multi-aperture Color-coded Optics for Snapshot Hyperspectral Imaging - Supplemental Document - • 208:3



Fig. S2. Measured (top set) and processed (bottom set) PSFs across wavelengths ranging from 429 nm to 700 nm (left to right). For each category, four rows of center-cropped PSFs are visualized, corresponding to the B, G1, G2, and R channels, respectively. Image intensity levels are normalized for visualization purposes.



Fig. S3. **Spectral Response Curve**. The design, fabrication target, and measured spectral response curves for the B, G1, G2, and R channel color filters are shown.

color filter initialization strategies: (1) four identical monocular sensor response curve; and (2) four evenly distributed Gaussian curves. We provide qualitative and quantitative comparisons in Fig. S5 and Tab. S3, respectively. In both cases, the optimization led to the development of four color curves with distinct primary peaks. However, Gaussian initialization favored high imaging quality at four narrow bandwidths with almost zero color response for the intervening wavelengths, relying on the reconstruction network to infer spectral information in the gaps, very similar to a conventional RGB-to-HS setup. As a result, it achieves a slightly higher RGB reconstruction Table S2. **Network Architecture Description**. Specifically, "conv-k(*a*)-s(*b*)-LRelu" represents a convolution layer with an $a \times a$ kernel window, using the stride *b*, followed by a Leaky Relu ($\alpha = 0.02$) activation function. We use "convT" to denote transposed convolution and "concat" to denote concatenation.

Input	Layer Type	Output (# Channels)	
sensor_capture	conv-k5-s1-LRelu conv-k3-s1-LRelu	x_00 (32)	
x_00	conv-k3-s1-LRelu conv-k3-s1-LRelu	x_01 (32)	
x_00	conv-k3-s2-LRelu conv-k3-s1-LRelu conv-k3-s1-LRelu	x_11 (64)	
x_11	convT-k2-s2-Relu	x_11u (32)	
concat(x_00, x_11u)	conv-k3-s1-LRelu conv-k3-s1-LRelu	x_02 (32)	
x_01	conv-k3-s2-LRelu	x_01d (64)	
concat(x_11, x_01d)	conv-k3-s1-LRelu conv-k3-s1-LRelu	x_12 (64)	
x_01	conv-k3-s4-LRelu	x_01dd (128)	
x_11	conv-k3-s2-LRelu	x_11d (128)	
concat(x_11d, x_01dd)	conv-k3-s1-LRelu conv-k3-s1-LRelu	x_22 (128)	
x_12	convT-k2-s2-Relu	x_12u (32)	
x_22	convT-k4-s4-Relu	x_22uu (32)	
concat(x_02, x_12u, x_22uu)	conv-k3-s1-LRelu conv-k3-s1-LRelu	x_03 (32)	
x_02	conv-k3-s2-LRelu	x_02d (64)	
x_22	convT-k2-s2-Relu	x_22u (64)	
concat(x_11, x_02d, x_12, x_22u)	conv-k3-s1-LRelu conv-k3-s1-LRelu	x_13 (64)	
x_13	convT-k2-s2-Relu	x_13u (32)	
concat(x_00, x_03, x_13u)	conv-k3-s1-LRelu conv-k3-s1-LRelu	x_04 (32)	
x_04	conv-k5-s1-Relu	output_hs (31)	
x_04	conv-k5-s1-Relu	output_rgb (3)	

ACM Trans. Graph., Vol. 43, No. 6, Article 208. Publication date: December 2024.



Fig. S4. Error Map Visualization of the Multi-Aperture Configuration. We present the absolute reconstruction error for RGB and each spectral channel as a grayscale image, with the intensity scaled by a factor of 1.5 to enhance visibility, as the additional visualization for Fig. 3 of the main manuscript.

Table S3. Quantitative Analysis on Color Filter Initialization. We eval-
uate different color filter initializations in simulation, including Bayer RGGB
curve ('Proposed'), four evenly distributed Gaussian curve ('Gaussian'), and
four identical monochrome spectral response curve ('Monochrome').

	R	GB	HS			
	SSIM	PSNR	SSIM	PSNR	$SAM\downarrow$	
Proposed	0.96	38.01	0.92	32.82	0.21	
Gaussian	0.97	38.36	0.89	29.99	0.42	
Monochrome	0.95	37.92	0.88	29.56	0.41	

performance but noticeable reduced spectral performance, with the SAM score almost doubled compared to the proposed method. Conversely, with all four peaks initially centered around the green, the optimization struggled to shift them sufficiently, resulting in reduced coverage at the spectrum extremes.

Additional Quantitative Comparison over the KAUST Dataset. We further evaluated the proposed method and all baselines on the unseen KAUST dataset [Li et al. 2021], which includes 409 indoor



Fig. S5. **Color Filter Initialization Analysis**. We evaluate different color filter initializations in simulation, including Bayer RGGB curve ('Proposed'), four evenly distributed Gaussian curve ('Gaussian'), and four identical monochrome spectral response curve ('Monochrome').

Table S4. **Quantitative comparison over the unseen KAUST dataset.** We compare our reconstruction network architecture with state-of-the-arts, including RGB-to-Spectrum reconstruction represented by HRNet [Zhao et al. 2020] and MST++ [Cai et al. 2022], and recent compressive snapshot spectral imaging systems, represented by QDO [Li et al. 2022] and SCCD [Arguello et al. 2021]. Note, that we do not report the RGB output scores for HRNet and MST++ as it takes RGB as input, and the SCCD scores are not punished for the missing channels.

	RGB			HS		
	SSIM	PSNR	Delta E	SSIM	PSNR	SAM ↓
Proposed	0.95	34.57	2.12	0.91	27.56	0.19
MST++ [2022]	-	-	-	0.88	23.70	0.25
SCCD [2021]	0.88	26.30	4.22	0.80	22.59	0.28
HRNet [2020]	-	-	-	0.83	18.98	0.26
QDO [2022]	0.85	23.38	6.57	0.75	19.25	0.35

and outdoor scenes. The quantitative performance is summarized in Tab. S4. The results follow a similar tendency to those presented in the main manuscript, with the proposed method outperforming all baselines by more than 3.86dB in PSNR.

Additional Qualitative Comparisons. In addition to the results presented in the main manuscript, we provide additional qualitative simulation results comparing our proposed method to existing baseline methods and ablated settings in the ablation studies on the multi-aperture setup and network architecture, as shown in Fig. S6. For each method, the RGB recovery results of the full image are shown on the left, and a zoomed-in version of both the RGB and spectrum recovery results are on the right. The ground truth is displayed on the top left, while the proposed method's reconstructions are on the top right. Rows 2 to 4 on the left present results from existing snapshot hyperspectral methods, represented by the RGB-to-Spectrum method HRNet [Zhao et al. 2020] and recent compressive snapshot spectral imaging systems QDO [Li et al. 2022] and SCCD [Arguello et al. 2021]. Rows 2 to 5 show outputs of settings discussed in the multi-aperture setup ablation study. The bottom two rows on the left and the bottom row on the right are from the network architecture ablation study, elucidating the contribution of different components in our reconstruction network. Additionally, we present spectral validation plots for all approaches for two specific points, labeled 1 and 2 on the ground truth RGB images, displayed at the bottom of the figure. The snapshot baseline methods are included in the left plots, and the ablation experiments are in the right plots, with the proposed method and the ground truth appearing in both sets of plots. These additional results further validate the capability of our method in precisely reconstructing spectral details with high fidelity, and the benefit of having independent spatial and spectral modulation across each channel. This is particularly evident in improved performance at extreme spectral ends, such as 429 nm and 699 nm, as shown in the spectral validation plots.

5 ADDITIONAL EXPERIMENTAL RESULTS

We present additional experimental results using our multi-aperture color-coded HSI system in both outdoor (Fig. S8) and indoor (Fig. S7) environments, comparing the reconstructions with reference captures from the commercially available Specim IQ hyperspectral camera. Note that the indoor scene reconstructions (Fig. S7) may appear slightly hazy due to stray light from overhead lighting, as the prototype setup lacks a closed lens barrel. Additionally, a small area of the sensor was stained during the capture process, resulting in a dark patch in several captures and reconstructions. Similar to the results in the main manuscript, these additional scenes demonstrate a close alignment between the spectral profiles of scene points reconstructed by our system and those recorded by the Specim IQ camera. This consistency highlights the adaptability and efficiency of our methodology across varied environmental conditions.



Fig. S6. Additional Simulation Results. We compare our proposed method to existing baseline methods and various ablated settings from the multi-aperture setup and network architecture studies. For each method, the RGB recovery results of the full image are shown on the left, with zoomed-in RGB and spectrum recovery results on the right. Additionally, we present spectral validation plots for all approaches for two specific points, labeled 1 and 2 on the ground truth RGB images, displayed at the bottom of the figure. See Sec. 4 for details.

ACM Trans. Graph., Vol. 43, No. 6, Article 208. Publication date: December 2024.

Learned Multi-aperture Color-coded Optics for Snapshot Hyperspectral Imaging - Supplemental Document - • 208:7



Fig. S7. Additional Experimental Assessment - Indoor. We present additional experimental results for indoor environments with 400 ms integration time. For each scene, we include: (a) sensor captures comprising four sub-channel images (R, G1, G2, B); (b/e) RGB reconstructions compared to Specim IQ references; (c/f) close-up views of a cropped region across all 31 channels; and (d) spectral validation plots for four sampled points on the captured scene.



Fig. S8. Additional Experimental Assessment - Outdoor. We present additional experimental results for outdoor environments with 5.6 ms integration time. For each scene, we include: (a) sensor captures comprising four sub-channel images (R, G1, G2, B); (b/e) RGB reconstructions compared to Specim IQ references; (c/f) close-up views of a cropped region across all 31 channels; and (d) spectral validation plots for four sampled points on the captured scene.

ACM Trans. Graph., Vol. 43, No. 6, Article 208. Publication date: December 2024.

Learned Multi-aperture Color-coded Optics for Snapshot Hyperspectral Imaging - Supplemental Document - • 208:9

REFERENCES

- Henry Arguello, Samuel Pinilla, Yifan Peng, Hayato Ikoma, Jorge Bacca, and Gordon Wetzstein. 2021. Shift-variant color-coded diffractive spectral imaging system. *Optica* 8, 11 (2021), 1424–1434.
- Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. 2022. MST++: Multi-stage Spectral-wise Transformer for Efficient Spectral Reconstruction. In CVPRW.
- Siyu Dong, Hongfei Jiao, Ganghua Bao, Jinlong Zhang, Zhanshan Wang, and Xinbin Cheng. 2018. Origin and compensation of deposition errors in a broadband antireflection coating prepared using quartz crystal monitoring. *Thin Solid Films* 660 (2018), 54–58.
- Xiong Dun, Hayato Ikoma, Gordon Wetzstein, Zhanshan Wang, Xinbin Cheng, and Yifan Peng. 2020. Learned rotationally symmetric diffractive achromat for fullspectrum computational imaging. *Optica* 7, 8 (2020), 913–922.
- Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. 2008. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. IEEE transactions on image processing 17, 10 (2008), 1737–1754.

- Hayato Ikoma, Cindy M Nguyen, Christopher A Metzler, Yifan Peng, and Gordon Wetzstein. 2021. Depth from defocus with learned optics for imaging and occlusionaware depth estimation. In 2021 IEEE International Conference on Computational Photography (ICCP). IEEE, 1–12.
- Lingen Li, Lizhi Wang, Weitao Song, Lei Zhang, Zhiwei Xiong, and Hua Huang. 2022. Quantization-Aware Deep Optics for Diffractive Snapshot Hyperspectral Imaging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 19780–19789.
- Yuqi Li, Qiang Fu, and Wolfgang Heidrich. 2021. Multispectral illumination estimation using deep unrolling network. (2021), 1–8.
- Yuzhi Zhao, Lai-Man Po, Qiong Yan, Wei Liu, and Tingyu Lin. 2020. Hierarchical regression network for spectral reconstruction from RGB images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 422–423.
- Jingyuan Zhu, Siliang Zhou, Yi Ning, Xiong Dun, Siyu Dong, Zhanshan Wang, and Xinbin Cheng. 2023. Grayscale-patterned integrated multilayer-metal-dielectric microcavities for on-chip multi/hyperspectral imaging in the extended visible bandwidth. Opt. Express 31, 9 (Apr 2023), 14027–14036. https://doi.org/10.1364/OE.485869