MetaSpectra+: A Compact Broadband Metasurface Camera for Snapshot Hyperspectral+ Imaging

Supplementary Material

5. Implementation Details of the DDPM

The DDPM used for hyperspectral reconstruction closely follows the architecture of Hazineh *et al*. [28]. During training, we randomly crop four uniform-sized patches—one from each sub-image $I_{1:V}$—and use them as the model input. At inference time, we partition the sub-images into L non-overlapping patches $\{P^k_{1:V}\}_{k=1}^{L}$ of the same size. At each diffusion step $t \in [0, T]$, the DDPM employs a U-Net to estimate a hyperspectral datacube corresponding to each group of four patches with index k:

$$\epsilon^{k,t} = \text{U-Net}\left(\boldsymbol{s}^{k,t}, \{P^k_{1:V}\}\right), \tag{17}$$

$$H^{k,t} = \frac{\boldsymbol{s}^{k,t} - \sqrt{1-\upsilon^t}\epsilon^{k,t}}{\sqrt{1-\upsilon^t}}, \tag{18}$$

where $\boldsymbol{s}^{k,t}$ denotes the state of the k–th set of patches at time step t, $\epsilon^{k,t}$ is the corresponding estimated noise, and υ^t is the cumulative product of noise-retention factors up to time step t. The predicted datacubes $\{H^{k,t}\}_{k=1}^{L}$ are stitched together to form the full-scene hyperspectral estimate:

$$H^t(\boldsymbol{x}) = a^{k,t}H^{k,t}(\boldsymbol{x}) + b^{k,t}, \quad \text{where } P^k_{1:V} \ni \boldsymbol{x}. \tag{19}$$

The per-patch scale and offset parameters $(a^{k,t}, b^{k,t})$ ensure consistency across neighboring patches and are obtained with the guidance of measured sub-images:

$$\tilde{a}^{k,t}, \tilde{b}^{k,t} = \underset{\{a^{k,t}, b^{k,t}\}}{\arg\min} \left\| \hat{I}_{1:V}(H^t) - I_{1:V} \right\|^2, \tag{20}$$

where $\hat{I}_{1:V}$ denotes the re-rendered sub-images generated from the predicted hyperspectral datacube H^t via Eq. (14). Finally, the latent state $\boldsymbol{s}^{k,t}$ is updated via normalized gradient descent:

$$\hat{\boldsymbol{s}}^{k,t} = \boldsymbol{s}^{k,t} - \gamma(t)\frac{\nabla_{\boldsymbol{s}^{k,t}}\mathcal{L}}{\|\nabla_{\boldsymbol{s}^{k,t}}\mathcal{L}\|}, \tag{21}$$

where $\mathcal{L}$ is the loss in Eq. (20) evaluated with the converged $(\tilde{a}^{k,t}, \tilde{b}^{k,t})$. We use a decaying step size $\gamma(t) = \sqrt{t/T}$, which we found empirically to improve reconstruction stability and accuracy. This guidance process (Eqs. (18) to (21)) can be done iteratively to refine scale and offset parameters. The updated state $\hat{\boldsymbol{s}}^{k,t}$ is then used for the next time step.

We train the DDPM on an NVIDIA GeForce RTX 5090 GPU with 32 GB of memory. The sampling process uses 50 denoising diffusion steps and 20 guidance iterations. Table 2 lists all training hyperparameters used in our implementation.

Table 2. Hyperparameters of DDPM.

Item	Value
Patch size	128×128
Batch size	64
Input channels	38
Output channels	26
Time steps	1000
Time embedding dimension	1024
U-Net block channels	[64, 128, 256, 512, 1024]
Use attention	All blocks
Attention head dimension	32
Classifier-free guidance	False
β schedule	Linear, 1e-4 $\rightarrow$ 2e-2
Learning rate schedule	Cosine, 1e-4 $\rightarrow$ 1e-6
Min-SNR weighting	True
K_{min}-SNR	5.0
Loss type	L1 (on noise)
Optimizer	AdamW
Epochs	15000

6. Metasurface Design and Fabrication

6.1. Fabrication Procedure

The fabrication process for our metasurfaces is identical to the process described in Brookshire *et al*. [7], which can be performed in a typical university cleanroom. It begins with coating a fused silica substrate with a 775 nm thick silicon nitride layer deposited via plasma-enhanced chemical vapor deposition. Then, it spin coats a 200 nm ARP6200 resist layer onto the surface, followed by depositing a 20-nm aluminum conductive layer. The designed pattern is printed onto the sample using e-beam lithography with an electron dose of 700 uC/cm^2. After exposure, the Aluminium layer is removed, followed by a 90 s development step. Finally, we deposit a 30 nm Al_2O_3 layer and perform lift-off, forming a hard mask for subsequent etching. The designed nanostructures are transferred into the silicon nitride layer using plasma dry etching.

In addition to university cleanrooms, there are also several companies across different countries that offer custom metasurface fabrication services, ranging from single-sample prototyping to wafer-level production (e.g., $\sim$ 100 devices per wafer), with costs spanning from several thousand to hundreds of thousands of USD.

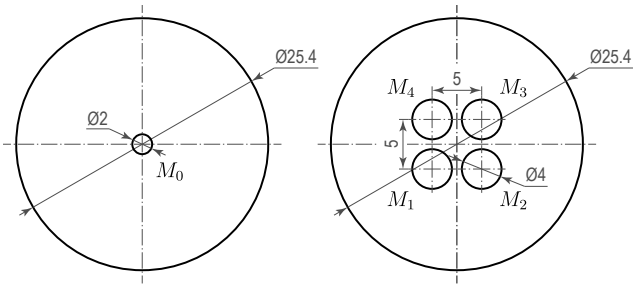

Figure 6. Schematics of the beamsplitting metasurface M_0 (left) and the dispersion-control metasurfaces $M_{1:4}$ (right). Units: mm.

6.2. Details of Metasurface Design

Deflection Vectors. As shown in Fig. 6, the beamsplitting metasurface and the dispersion-control metasurfaces have clear apertures of 2 mm and 4 mm, respectively, separated by $s = 4$ mm. The dispersion-control metasurfaces $M_{1:4}$ are arranged in a 2×2 configuration with a center-to-center spacing of 5 mm. The designed deflection vectors, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$, are set to:

$$\begin{aligned} \boldsymbol{\alpha}_i &= \frac{\lambda_{c,i}}{\lambda_{c,2}} \left[(-1)^{\lfloor i/2 \rfloor + 1}, (-1)^{\lfloor (i-1)/2 \rfloor + 1}\right]^\top \cdot \boldsymbol{\alpha}_2, \\ \boldsymbol{\beta}_i &= -\boldsymbol{\alpha}_i + \begin{cases} \left[0.017, (-1)^{i-1} 0.017\right]^\top & i = 1, 2 \\ \mathbf{0}, & i = 3, 4 \end{cases}, \\ \boldsymbol{\alpha}_2 &= [0.385, -0.385]^\top. \end{aligned} \tag{22}$$

These parameters are selected to ensure the front-parallel incident beam with wavelength 550 nm to be deflected to the center of each dispersion-control metasurface. The dispersion of channels 3 and 4 is eliminated as $\boldsymbol{\alpha}_i = \boldsymbol{\beta}_i, i = 3, 4$, while channels 1 and 2 retain a small amount of dispersion whose amount is decided by simulation.

Interleaving Strategy. Regular interleaving, which periodically arranges multiple phase patterns on a uniform nanocell grid, has been widely used in metasurface design to realize multi-functionality [24, 36]. However, this strategy becomes problematic when the uninterleaved phase profile corresponds to large deflection angles. To illustrate this effect, consider regularly interleaving the four deflection patterns $M_{0,1:4}$ in a 2×2 mosaic pattern:

$$M_0(\boldsymbol{x}; \lambda) = \sum_{i=1}^{4} M_{0,i}(\boldsymbol{x}, \lambda) \operatorname{comb}\left(\frac{\boldsymbol{x} - \Delta \boldsymbol{x}_i}{2w}\right), \tag{23}$$

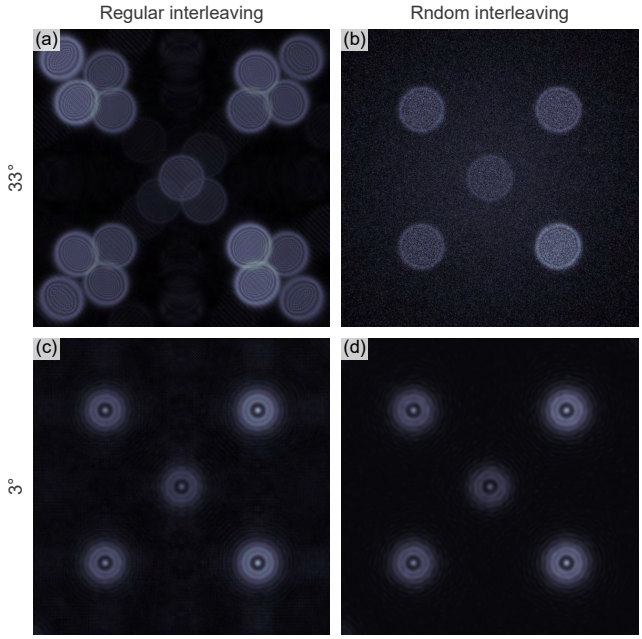

Figure 7. Simulated wavefront amplitude incident on the dispersion-control metasurface when using (a,c) regular interleaving and (b,d) random interleaving in the beamsplitting metasurface M_0 at different deflection angles. Random interleaving suppresses higher-order diffraction artifacts at the cost of increased background residual light. The central bright spot arises from zeroth-order diffraction when the system is operating at non-designed wavelength.

where comb$(\cdot)$ is a 2D Dirac comb function and $\Delta \boldsymbol{x}_i$ controls the sampling position

$$\Delta \boldsymbol{x}_i = \begin{cases} (0, 0) & i = 1 \\ (w, 0) & i = 2 \\ (w, w) & i = 3 \\ (0, w) & i = 4 \end{cases}. \tag{24}$$

Figure 7a shows the simulated field distributions of 550 nm at the plane immediately before the dispersion-control metasurface under different deflection angles and interleaving strategies. It clearly shows that regular interleaving leads to residual diffraction orders when the deflection angle becomes large, which will become repeated patterns on the image plane.

Random interleaving avoids these periodic replicas by replacing the structured sampling with an irregular arrangement, but it also introduces a weak background due to random mixing of the four patterns [7]. In our optical design, this trade-off makes random interleaving safer for multiplexing under large deflection angles. This can be evidenced in Fig. 7b, which shows four clear deflection patterns without residual diffraction orders under various deflection angles using random interleaving.

Table 3. List of parts.

No.	Item	Vendor & Stock Number	Quantity	Description
1	Objective lens	Thorlabs, AC254-400-A	1	⌀ 25.4 mm, 400 mm focal length, achromatic
2	Beam-splitting metasurface	SNOChip	1	⌀ 2 mm, 1-inch substrate
3	Dispersion-control metasurface	SNOChip	1	⌀ 4 mm, 2×2, 1-inch substrate,
4	Eyepiece	Edmund Optics, 63-704	4	⌀ 4 mm, 4 mm focal length, achromatic
5	Neutral density (ND) filter	Edmund Optics, 62-662	1	⌀ 50 mm, optical density (OD) 0.3
6	Neutral density (ND) filter	Edmund Optics, 62-665	1	⌀ 50 mm, optical density (OD) 0.9
7	Linear Polarizer	Edmund Optics, 26-919	1	50 mm×50 mm×0.11 mm
8	Photosensor	Basler, dmA3536-9gc	1	3536 × 3536 resolution, 2 μm pixel pitch, RGB

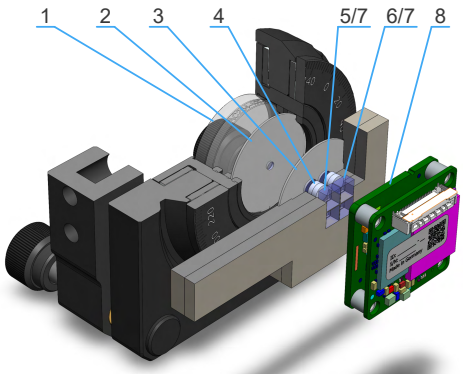

Figure 8. CAD model of the assembly. Numbers correspond to the items in the list of parts in Tab. 3.

7. Prototype

Assembling Instruction. The prototype is constructed from a combination of off-the-shelf components and custom 3D-printed parts. A full list of components is provided in Tab. 3 and the CAD model of the assembly is visualized in Fig. 8. The objective lens and the beam-splitting metasurface are mounted inside a rotation mount. The remaining optical elements are secured in a 3D-printed holder, where we design a circular recessed pocket and four holes to fit and align the dispersion-control metasurface and subsequent lenses. We cut off-the-shelf neutral density (ND) filters and polarizers to fit the filter holders using a water-jet blade. The 3D printed holder and photosensor are mounted on a three-axis translation stage to allow fine positional adjustment during installation.

8. Additional Information

Figure 9 shows the object being measured in Fig. 3c. It is a printed, multi-colored pattern mounted in a front-parallel orientation.

The three scenes that only contain front-parallel textured planes used exclusively for fine-tuning the computational model are shown in Fig. 10. Textures appearing in the fine-tuning scenes do not appear in the testing scenes to avoid information leakage.

Figure 9. Picture of the target being measured in Fig. 3c.

Figure 10. Sub-images I_3 of the real-world scenes for fine-tuning.

9. Additional Results

Snapshot Hyperspectral Imaging. Figure 11 shows additional hyperspectral reconstruction results using synthetic data generated from the KAUST dataset. MetaSpectra+ consistently outperforms previous hyperspectral imaging solutions in terms of both metrics and visual quality. We also validate the proposed system's performance on additional real-world experiments. Figure 12a demonstrates low reconstruction errors and faithful spectral recovery across real-captured, diverse scenes, demonstrating the robustness and effectiveness of our system.

Snapshot Hyperspectral+ Imaging. Figure 12b-c provides additional real-world experimental results for HDR + hyperspectral and polarization + hyperspectral imaging.

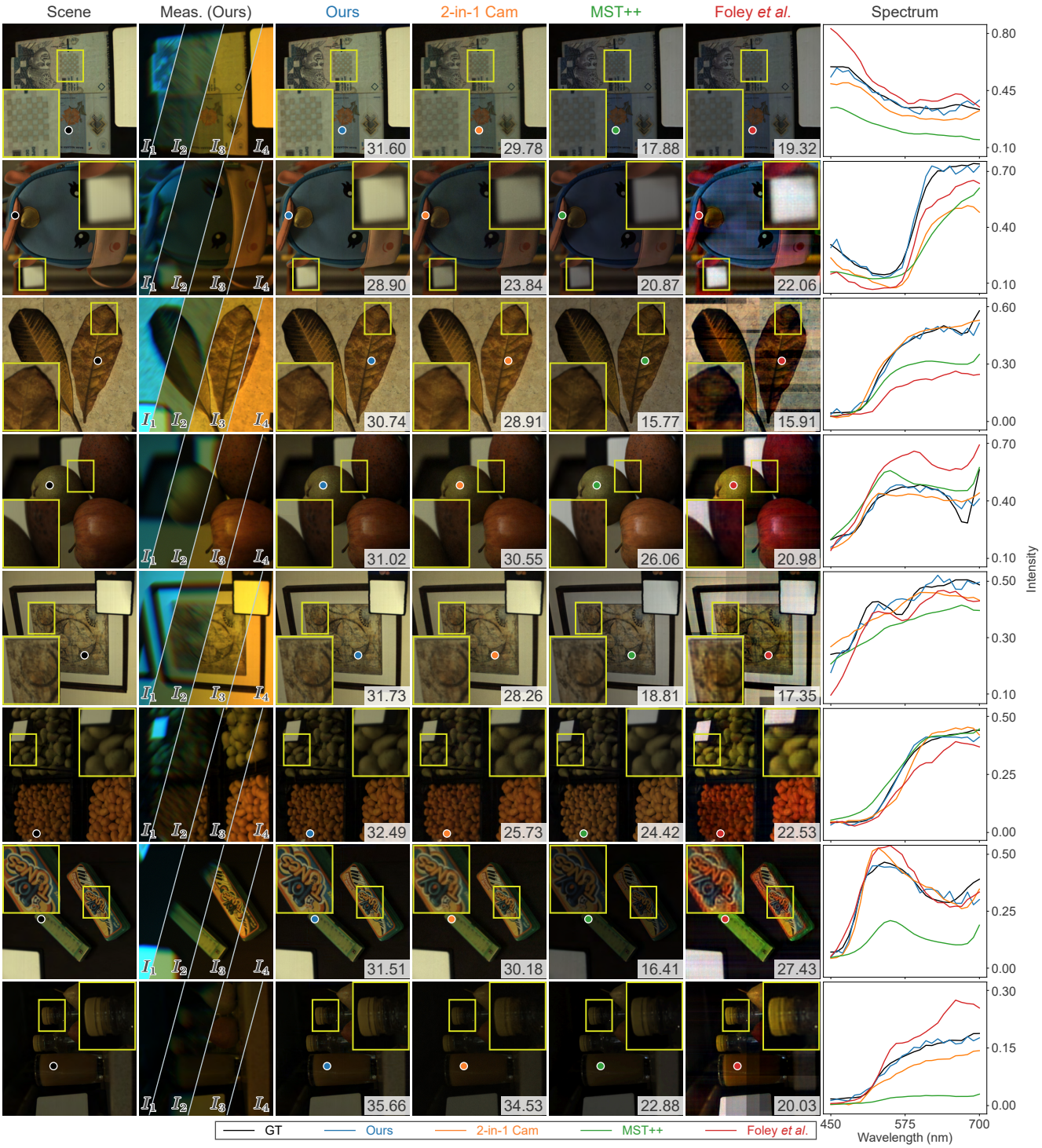

Figure 11. Additional sample hyperspectral reconstruction results on the KAUST dataset. The inset numbers are PSNR (dB) for hyperspectral reconstructions.

10. Discussion

Competing Approaches. Besides the proposed solution, we identify three alternative approaches to realize the functionality of MetaSpectra+. Here, we discuss the advantages and disadvantages of each.

1. Diffraction gratings can provide multi-beam splitting,

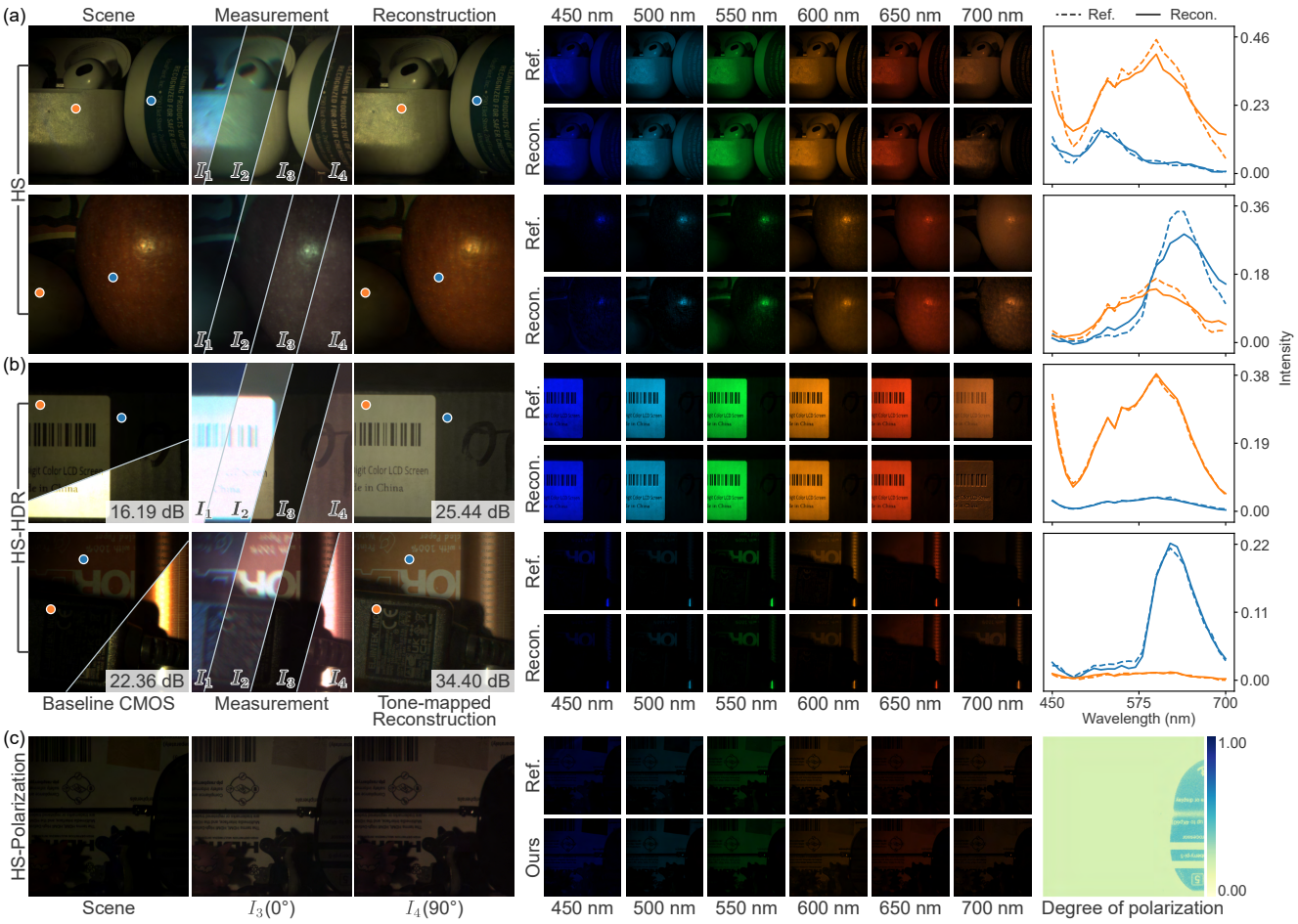

Figure 12. Additional sample real-world results of MetaSpectra+. Inset numbers in (b) represent the dynamic range (dB) of the picture.

and industrial off-the-shelf components are relatively cost-effective. However, implementing the specialized optical functionalities required in this work would necessitate custom-fabricated gratings, leading to costs comparable to metasurfaces.

2. Diffractive optical elements (DOEs) enable custom wavefront shaping and typically offer lower cost and larger aperture sizes than metasurfaces. However, realizing the multi-beam-splitting behavior of this work within a single DOE is challenging.
3. Refractive optics are generally the most cost-effective option; however, achieving comparable functionality using off-the-shelf components would significantly increase the system form factor, as the required multifunctionality would rely on cascaded beamsplitters.

Limitations. Due to a reduced diffraction efficiency caused by the random interleaving in the beam-splitting metasurface, an increased integration time is typically required. Consequently, our current prototype operates at up to 10 FPS, posing challenges for direct deployment in high-speed video applications. However, this issue can potentially be mitigated by optimizing the metasurface design, such as decreasing the beam deflection angle, or using materials with higher refractive indices, such as gallium nitride and titanium dioxide.