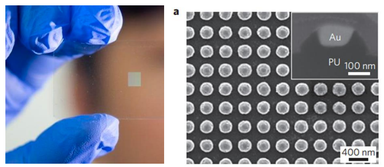
Compact computational cameras by joint metasurface and network design

Kelsey Lee
Electrical Engineering Department, Stanford University

Motivation

- **Low-latency, lightweight sensing is critical for modern computer vision**, but processing high-resolution images digitally is computationally expensive and power intensive.
- **Optical computing** enables massively parallel processing while naturally performing operations such as convolutions during image formation.
- **Metasurfaces** provide a natural platform for optical computing; they are lightweight, compact, and enable flexible phase control.

Metasurfaces



[1]

Optical Image Formation

Angular Spectrum Method:

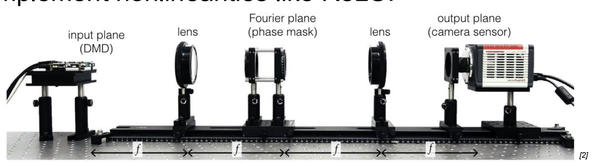
$$\mathrm{PSF}(x,y;z) = \left|\mathcal{F}^{-1}\left\{\mathcal{F}[U(x,y;0)]\,\exp\left(iz\sqrt{k^2-k_x^2-k_y^2}\right)\right\}\right|^2$$

Image formation as a spatial domain convolution:

$$I_{\mathrm{img}}(x,y) = \iint I_{\mathrm{obj}}(\xi,\eta)\,\mathrm{PSF}(x-\xi,\ y-\eta)\,d\xi\,d\eta$$

Related Work

- Chang et al. (2018) demonstrated hybrid optical classification using a 4f system with passive phase masks.
- Recent work uses metasurfaces for optical classification, including polarization multiplexing and polychromatic metasurface designs.
- A key limitation is optical linearity, since optical systems cannot easily implement nonlinearities like ReLU.



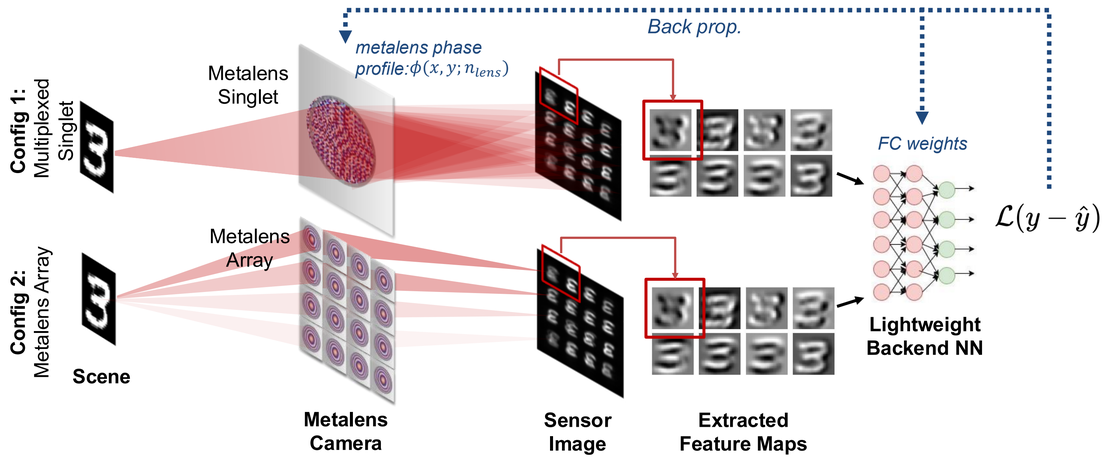
[3]

References

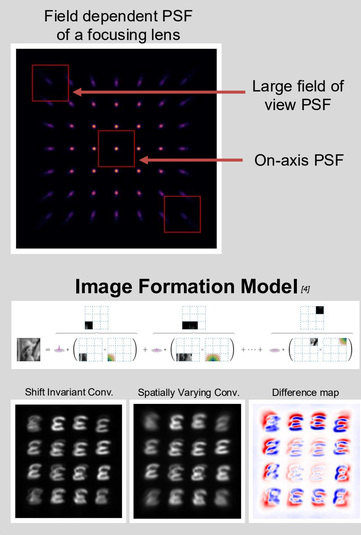
[1] Meinzer, N., Barnes, W. & Hooper, I. Plasmonic meta-atoms and metasurfaces. Nature Photon 8, 889–898 (2014).
[2] Chang, J., Sitzmann, V., Dun, X. *et al.* Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci Rep* **8**, 12324 (2018).
[3] Choi, M., & Majumdar, A. (2025). Free-space optical encoder for computer vision. *npj Nanophotonics*, *2*(1), 36.
[4] Denis, L., Thiébaut, E., Soulez, F., Becker, J. M., & Mourya, R. (2015). Fast approximations of shift-variant blur. *International Journal of Computer Vision*, *115*(3), 253-278.

Hybrid Optical-Digital Compute Pipeline

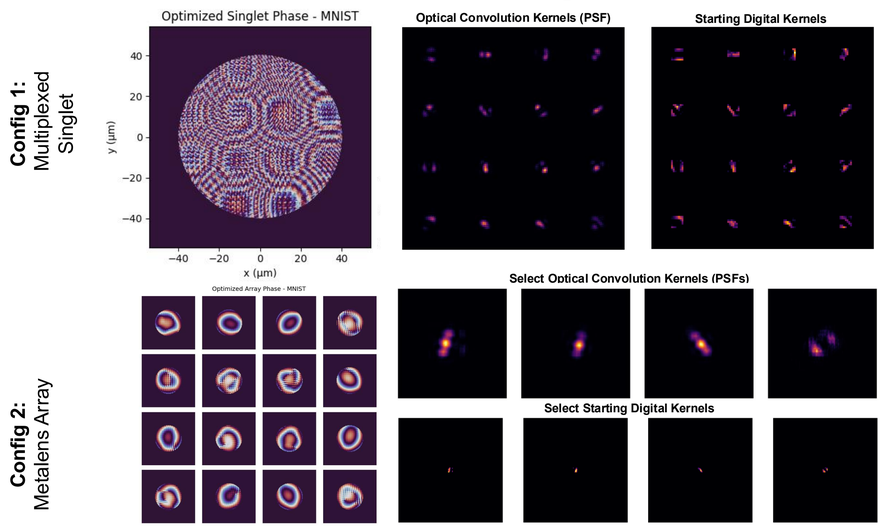
- **Optical front end:** A metalens camera forms the sensor image and optical feature maps
- **Digital backend:** The sensor image is passed to a lightweight neural network for classification



Maintaining fidelity over larger FoV: Spatially Varying PSFs



Experimental Results



	MNIST	Fashion MNIST
Classification Accuracy [%]		
Digital Backend Only	92.70	84.29
All Digital (1-Layer CNN)	98.58	90.11
Hybrid Optical-Digital (Multiplexed Singlet)	98.72	89.62
Hybrid Optical-Digital (Lens Array)	98.31	88.92
Total MAC Operations / Image		
All Digital (1-Layer CNN)	19,500,000	
Hybrid Optical-Digital	435,000	
% Computation Reduction	**97.8%**	