Spectral DefocusCam: super-resolved hyperspectral imaging through defocus

Christian Foley[1,*], Eric Markley[2], Kyrollos Yanny[2], Laura Waller[1], and Kristina Monakhova[3]

2025 IEEE International Conference on Computational Photography (ICCP) | 979-8-3315-5513-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICCP64821.2025.11143845

Abstract—On-chip spectral filter arrays enable compact, snapshot hyperspectral cameras that could be useful for applications ranging from agriculture to medical imaging and robotics. The spectral range and resolution of these cameras are determined by the number of spectral filters on the sensor, which are usually grouped together in repeating super-pixels, similar to the structure of a Bayer filter. Many tasks require dozens to hundreds of spectral channels, but this usually comes at the cost of a proportional decrease in spatial resolution and light throughput. We propose a multi-shot approach that leverages lens defocus to encode light before it is filtered by an on-chip spectral filter. Using this approach, we capture a burst of spatio-spectral encoded measurements at different focus planes, then solve an inverse problem to recover a high-resolution spatio-spectral datacube. Since most lens-based cameras already have a focusing mechanism, our approach is practical and could be broadly used by spectral filter array-based hyperspectral cameras. We provide analysis of our proposed system and comparisons against alternative methods. Finally, we demonstrate a prototype system and present experimental results of challenging dense, daylight scenes, demonstrating greater than 3× sub-super-pixel super-resolution.

Index Terms—Computational Photography, Hyperspectral Imaging

1 Introduction

While human vision and most consumer cameras only capture three bands of light (red, green, and blue), hyperspectral cameras measure many wavelengths. This additional spectral resolution can provide new insights that are invisible to the human eye; for example, enabling precision crop monitoring in agriculture [1], localizing tumors during surgery [2], and improving food inspection [3]. Since hyperspectral imaging captures multidimensional data (2D spatial + 1D spectral) and imaging sensors are two-dimensional, there is a trade-off between spatial, spectral, and temporal resolution. Over the years, several scanning and snapshot methods have been developed to acquire high-resolution hyperspectral datacubes [4]. However, they generally require bulky, expensive setups with moving components and/or complex optical elements, such as prisms, gratings, spatial-light modulators, and multiple lenses. Because of this, such methods are limited to applications where cost and form factor are not crucial considerations [5].

Recently, advances in silicon-level on-chip spectral filter arrays have enabled the mass manufacturing of spectrally-sensitive CMOS cameras. These devices enable compact hyperspectral imaging and could be brought to the consumer market as handheld devices. However, filter array based hyperspectral methods face a trade-off between spectral and spatial resolution. Spectral filters are often bunched into repeating super-pixels, similar to Bayer filters for color imaging [6], but with dozens to hundreds of filters instead of three. This leads to a significant loss in resolution. For example, a super-pixel with 64 spectral channels arranged in an 8×8 grid will have an 8× loss in spatial resolution in each dimension. Prior work addresses this loss of resolution using a diffuser and compressive reconstruction [7]. However, this approach requires sparsity in the sample and has limited dynamic range, resulting in poor performance for dense daylight scenes.

We propose a new opportunistic method to improve spatial-spectral resolution in filter array hyperspectral cameras based on lens defocus. Most handheld consumer cameras can easily and rapidly defocus their lens. Defocusing the lens increases the size of the PSF, allowing us to control how many spectral filters the light from each point in the scene passes through. Small PSFs (in focus measurements) give good spatial resolution but incomplete spectral information, whereas large PSFs (defocused measurements) spread light across many spectral filters, giving more complete spectral information, but worse spatial resolution. Thus, we capture a burst of measurements with varying levels of defocus and then computationally recover a hyperspectral data cube with good resolution in both spatial and spectral dimensions. We compare our method against other filter-based snapshot hyperspectral imaging methods, showing improved image quality and spatial-spectral resolution. In addition, we demonstrate a hardware prototype for hyperspectral imaging of challenging, dense, bright scenes.

Our proposed system, as depicted in Fig. 1, is:

1) **Compact and scalable**: The spectral filter array can be directly manufactured on a CMOS imaging sensor, similar to a Bayer filter but with more filters.
2) **Practical**: We do not need specialized hardware besides the on-chip spectral filters.

[1] *Department of Electrical Engineering & Computer Sciences, University of California, Berkeley, CA, 94720, USA*
[2] *UCB/UCSF Joint Graduate Program in Bioengineering, University of California, Berkeley, CA, 94720, USA*
[3] *Department of Computer Science, Cornell University, Ithaca, New York, 14850*
* *Email: cffoley@berkeley.edu*
* *Project website: https://github.com/Christianfoley/SpectralDefocusCam*

979-8-3315-5513-9/25/$31.00 ©2025 IEEE
Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

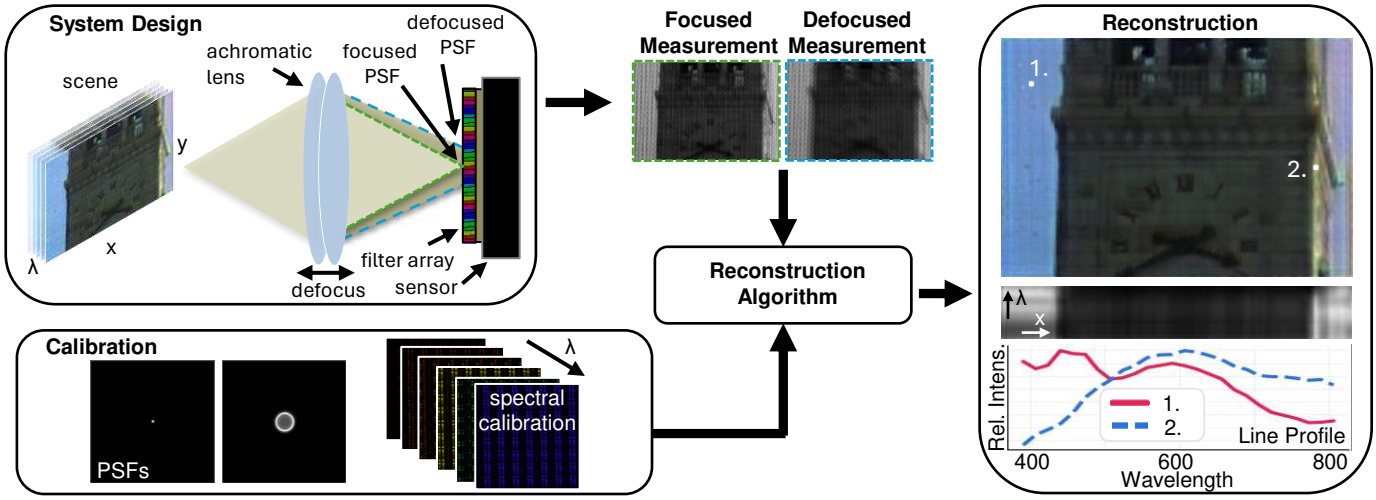

Fig. 1. Spectral DefocusCam consists of a spectral filter array placed at the sensor plane of a lensed camera. The lens is defocused to change how light is encoded onto the filter array. After capturing multiple encoded measurements with varying defocus, we solve an underdetermined inverse problem to recover a high-resolution hyperspectral datacube. A one-time calibration procedure captures PSFs for different focus levels and calibrates the spectral filter placement on the sensor.

3) **Provides greater than 3× super-resolution**: We demonstrate more than 3× improvement in resolution beyond the super-pixel size in experimental scenes.
4) **Works for dense daylit scenes**: We experimentally demonstrate performance on challenging dense scenes.

2 Related Work

Compressive hyperspectral cameras: A number of compressive hyperspectral imagers have been developed over the years. Some are based on coded apertures and utilize prisms, diffraction gratings, and multiple relay lenses to encode hyperspectral information [8], [9], [10], [11]. While these methods provide high-quality hyperspectral reconstructions, they are often bulky and confined to lab benches. Other methods use dispersion through a prism [12] or diffraction through a diffractive optical element [13] [14] to encode hyperspectral information. These methods can be compact but have complex design trade-offs that make it difficult to control and adapt the spectral resolution for different applications. In contrast to these methods, our system is compact and has a controllable spectral resolution determined by the spectral filters.

Spectral filter arrays: Some hyperspectral cameras utilize arrays of spectral filters across the sensor to obtain a hyperspectral datacube. Like Bayer filters for color imaging, spectral filters are often arranged in super-pixels that repeat across a sensor, Fig. 2(a). These filters can be narrowband [15], [16] or broadband, with a custom-designed spectral response [17], [18], [19], [20]. Fabry-Perot spectral filters can be monolithically integrated on top of a CMOS sensor at the wafer level, making these compact and potentially scalable [16]. Unlike dispersive hyperspectral cameras, filter-array based hyperspectral cameras offer the distinct advantage of spectral band selection, making them easily customizable to downstream applications.

For narrowband filters, the spectral resolution and range are determined by the number of filters and their bandwidths. As mentioned before, the spatial resolution decreases as the number of filters increases, creating a trade-off between spatial resolution and spectral sampling. Prior work has shown that a randomizing optic, like a diffuser, can overcome this limit and provide sub-super-pixel resolution [7]. However, because of the non-compact point spread function (PSF) of the randomizing optic, this method struggles to recover dense and dim scenes. Our work takes inspiration from this method to achieve sub-super-pixel resolution using a more general purpose optic – a lens – and is less sensitive to noise and scene density due to its compact PSFs.

Multi-shot hyperspectral imaging: Traditional hyperspectral imagers scan the scene to obtain hyperspectral information, trading temporal resolution for spatio-spectral resolution. Common methods include pushbroom imagers, which scan spatial positions [21], and filter wheels, which scan one wavelength band at a time [22]. Often, these methods fully sample the hyperspectral datacube of interest, but this comes at the loss of temporal resolution due to slow scanning speeds. Compressive hyperspectral imagers can have good spatial and temporal resolution but must solve an underdetermined inverse problem that relies on sparsity or structure of the scene. Several methods propose leveraging multiple distinct measurements to improve spatial and spectral reconstruction fidelity for cases where the temporal resolution requirements are relaxed or multiple sensors are feasible. For example, using different aperture codes [23] or multiple cameras to capture both panchromatic and hyperspectral images at the same time [24], [25]. Similarly, our method utilizes multiple images to improve the conditioning of the hyperspectral recovery problem.

Depth from defocus: Another line of work utilizes focal stacks to infer depth from images [26], [27], [28], [29], [30], [31]. Inspired by this work, we propose leveraging defocus to encode additional information and improve the conditioning of our compressive hyperspectral recovery inverse problem.

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

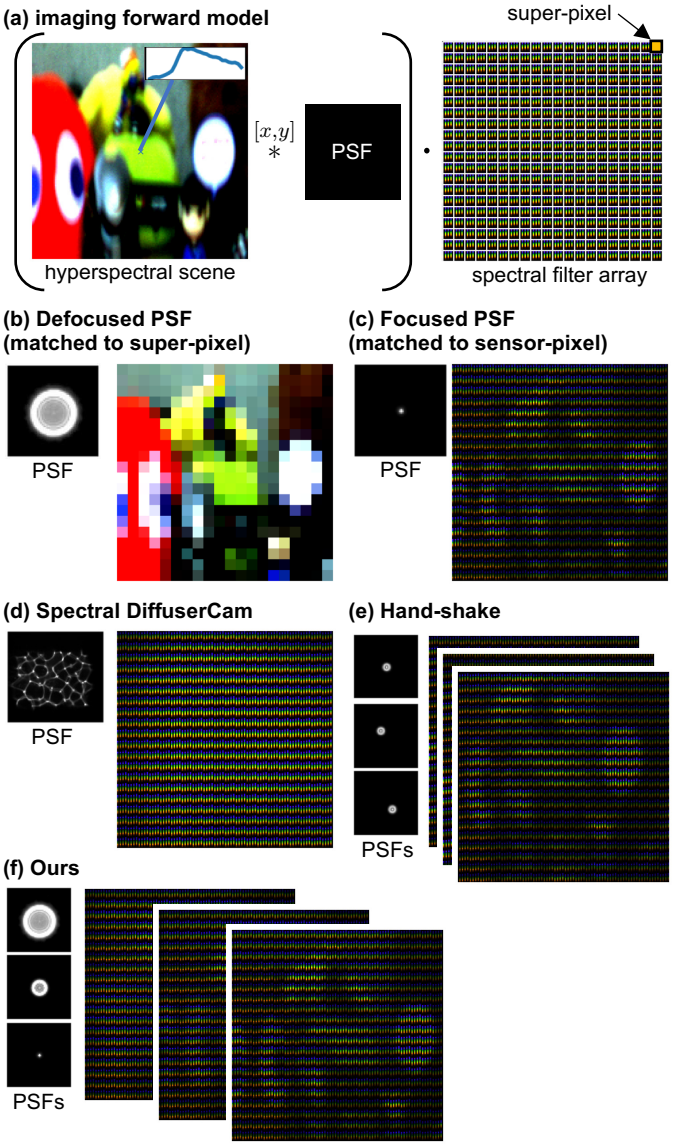

Fig. 2. Imaging forward models, comparing our method to alternative filter-based, compressive hyperspectral imaging techniques. (a) A hyperspectral scene is first convolved with a point spread function (PSF) and then filtered by the on-chip spectral filter array, consisting of repeating super-pixels, each with an 8×8 grid of narrowband spectral filters. (b) In one approach, a defocused PSF matched to the super-pixel size results in single-shot hyperspectral imaging with 8× reduced spatial resolution. (c) A focused PSF matched to the filter or sensor-pixel size results in high resolution, but holes in the measurement. (d) Spectral DiffuserCam uses a caustic PSF pattern to spread the light, but maintain high-frequency features. Compressed sensing enables recovery of the full hyperspectral cube with minimal loss of spatial resolution, provided sparsity priors are applicable. (e) A handshake method could be used to capture additional shifted measurements to improve the reconstruction. (f) We propose using multiple measurements with different focus to improve the conditioning of hyperspectral recovery with complementary information: complete low spatial-resolution measurements and incomplete high spatial-resolution measurements.

3 Methods

To motivate our method, let us first examine two single-shot filter-based hyperspectral imagers: one with a defocused measurement, Fig. 2(b), and the other with a focused measurement, Fig. 2(c). In both cases, the dense spatial-spectral scene is convolved with the PSF of the imager and then spectrally filtered by a spectral filter array, which consists of repeating blocks of narrowband 8×8 spectral filters, Fig. 2(a).

In the defocused measurement, light from a point in the scene will map to the entire super-pixel. This results in accurate spectral sampling but an 8× loss of spatial resolution along each dimension. In the focused measurement, light from a point in the scene maps to only a single spectral filter within a super-pixel. This results in gaps in the measurement at locations where points are not within the passband of certain filters. Both methods are underdetermined and cannot recover the high-resolution spatial-spectral image without using strong priors and solving an underdetermined inverse problem. For the defocused measurement, this inverse problem is akin to super-resolution to recover high-frequency features from a low-resolution image [32]. For the focused measurement, this problem resembles image inpainting [33] to fill the holes in the high-resolution spatial-spectral image.

Previous snapshot hyperspectral imaging methods propose using a diffuser as a dispersing optical element to create a PSF of large spatial extent with sharp, high spatial resolution features, Fig. 2(d) [7]. This maps each point in the scene to many spectral filters across the sensor, but with a high-frequency structure that enables the recovery of spatio-spectral features using compressive sensing. However, since preserving a high-frequency structure demands that the spatial extent of the PSF be much larger than a single super-pixel, these methods require sparsity constraints and suffer from poor signal-to-noise ratio (SNR) due to significant multiplexing and deconvolution artifacts. This works well for very sparse scenes, but results in reduced performance for dense and bright scenes. We propose to leverage multiple, easy-to-acquire measurements to improve the conditioning of the underdetermined inverse problem and more effectively recover high-resolution spatial-spectral images. Two ways to increase the number of unique measurements are from handshake, Fig. 2(e), and lens defocus, Fig. 2(f).

Handshake leverages natural hand tremors to obtain multiple unique measurements. This method has been applied to smartphone cameras to achieve multi-frame super-resolution [34]. In our system, handshake could enable each point in the scene to be measured by multiple spectral filters, improving the conditioning of the problem. This method is promising because it requires no hardware modifications but, in practice, relies on an additional challenging frame registration step, informed by gyroscopic measurements and locally aligned sensor measurements from each frame in an image burst. Frame registration is sensitive to noise and excessive local movement, making it especially challenging for hyperspectral imaging given the limited light throughput and large size of the spectral filters, which create significant gaps in the measurements.

Our method instead leverages defocus to acquire additional unique measurements and improve the conditioning of the inverse problem. Modern autofocusing implementations can rapidly acquire defocus stacks. Using algorithms to combine defocused and focused measurements, we can leverage the high spatial resolution of the focused measurement and the high spectral resolution of the defocused measurement to solve a better-determined inverse problem

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

and recover high-frequency spatio-spectral features within a scene. Furthermore, since the PSF is compact, our method has substantially higher SNR and light efficiency than methods that use a non-compact randomizing mask and can be used for dense, natural scenes. We compare our method against the focused, defocused, Spectral DiffuserCam, and handshake methods in Sec. 5 and 5.5.

3.1 Optical Forward Model

We capture n total measurements by positioning our sensor at n distances from the lens, where the ith measurement, $\mathbf{b}_i$, is given by the forward model:

$$\mathbf{b}_i = \sum_{\lambda=0}^{K-1} \mathbf{F}_\lambda[x,y] \cdot \text{crop}\Big(\mathbf{h}_i[x,y] \overset{[x,y]}{*} \mathbf{v}[x,y,\lambda]\Big) \quad (1)$$

$$= \mathbf{A}_i \mathbf{v}, \quad (2)$$

where $\mathbf{F}_\lambda[x,y]$ is the filter function, $\mathbf{h}_i[x,y]$ is the PSF sampled from our lens at the ith offset of defocus position, $\mathbf{v}[x,y,\lambda]$ is the scene, and crop$(*)$ is a cropping to the area of the sensor covered by the finite filter. The scene is first convolved with the lens' PSF, $\mathbf{h}_i[x,y]$, and then point-wise multiplied by the filter function of the filter array.

3.2 Reconstruction

To reconstruct the original data cube, we formulate the following inverse problem:

$$\hat{\mathbf{v}} = \underset{\mathbf{v}\geq 0}{\text{argmin}} ||\mathbf{A}\mathbf{v} - \mathbf{b}||_2^2, \quad (3)$$

where $\mathbf{A}$ represents the linear forward model and $\mathbf{b}$ represents the measurement. To accommodate multiple levels of focus and the resulting measurements, we express $\mathbf{A}$ as $[\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_n]$ and $\mathbf{b}$ as $[\mathbf{b}_1\mathbf{b}_2\cdots\mathbf{b}_n]$ where n is the number of levels of defocus. To solve this inverse problem, we propose to use both a classic iterative method with a simple prior as well as a deep-learned reconstruction method. The classic method lets us analyze the performance and posedness of our system without reliance on a black-box reconstruction method, whereas the deep-learned method provides advantages in expressivity, robustness to noise and reconstruction speed given the overhead of a training step.

3.2.1 Iterative Reconstruction

For the classic, iterative method, we use the fast iterative shrinking-threshold algorithm (FISTA) [35] with 3D Total-Variation (3DTV) regularization [36] to solve the inverse problem. Since this inverse problem is underdetermined, we rely on a prior to constrain the set of solutions. We leverage anisotropic 3DTV, which promotes sparse image gradients, as our imaging prior:

$$\hat{\mathbf{v}} = \underset{\mathbf{v}\geq 0}{\text{argmin}} ||\mathbf{A}\mathbf{v} - \mathbf{b}||_2^2 + \tau||\nabla_{\mathbf{xy}\lambda}\mathbf{v}||_1, \quad (4)$$

where $||\nabla_{\mathbf{xy}\lambda}(\cdot)||_1$ is total variation along each dimension and τ is a tuning parameter along each dimension. We solve this problem using 800 iterations of FISTA.

3.2.2 Deep Reconstruction

Alternatively, we can use a physics-informed deep-learning approach to solve the inverse problem posed in Eq. (3) [37], [38], [39]. A deep-learning reconstruction provides advantages not only in expressivity and robustness to noise but also in reconstruction speed; unlike iterative approaches, it requires only a single forward pass to reconstruct the full hyperspectral data cube.

Our proposed model consists of first calculating the adjoint of the measurement and then feeding the adjoint through a modified 3D U-Net.

The adjoint of each compressed 2D measurement is calculated using the following equation:

$$\tilde{\mathbf{v}}_i = \mathbf{h}_i^* \overset{[x,y]}{*} \text{pad}(\mathbf{F}_\lambda[x,y] \overset{[\lambda]}{\cdot} \mathbf{b}_i) \quad (5)$$

$$= \mathbf{A}_i^H \mathbf{b}_i \quad (6)$$

where $\mathbf{A}_i^H$ is the adjoint of our forward model, * is the complex conjugate, and pad$(*)$ is a centered zero-padding to prevent wrap-around effects from circular convolution. This step effectively projects each compressed 2D ($[x,y]$) measurement into 3D ($[x,y,\lambda]$). Using the knowledge of the physical system to pre-process the inputs to the neural network decreases its computational load and leads to faster training and better reconstruction.

To further improve training stability and decrease time to convergence, we condition the latent features of each convolutional block in the U-Net's contracting path on the most focused and defocused PSF. To ensure matching between the spatial dimensions of the PSFs and the latent features of the network, we downsample the PSFs through sequential average pooling to the appropriate spatial dimensions. Further details and study on the network architecture can be found in the supplementary material.

Our U-Net aims to map the computed adjoints, $\tilde{\mathbf{v}}_i$, to the ground truth hyperspectral cube, $\mathbf{v}$. The network is trained in a supervised manner using simulated data with the following loss function:

$$\mathcal{L}_{MSE}(\Theta) = \frac{1}{N}\sum_{k=1}^{N} \|\mathbf{v}_k - \Theta(\tilde{\mathbf{v}}_k, \mathbf{h})\|_2^2, \quad (7)$$

where N is the number of training samples and Θ is the physics-informed conditional U-Net. To accommodate multiple measurements, we express the kth adjoint, $\tilde{\mathbf{v}}_k$, as $[\tilde{\mathbf{v}}_1\tilde{\mathbf{v}}_2\cdots\tilde{\mathbf{v}}_n]_k$ and the PSFs, $\mathbf{h}$, as $[\mathbf{h}_1\mathbf{h}_2\cdots\mathbf{h}_n]$, where n is the number of levels of defocus.

4 Implementation Details

4.1 Hardware Details

In Fig 3 we provide images of our experimental hardware both in-lab and outside while capturing outdoor scenes. Our system consists of a single optical lens and a thin-film spectral filter array donated by Viavi Solutions (Santa Rosa, CA) [40] that is adhered directly to a CMOS sensor using an optical adhesive. The filter array is composed of repeating super-pixels, each of which consists of 8×8 $20\,\text{nm}$ narrowband spectral filters with a height and width of $10\,\mu\text{m}$. We use a 2-inch bi-convex achromatic lens with a

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

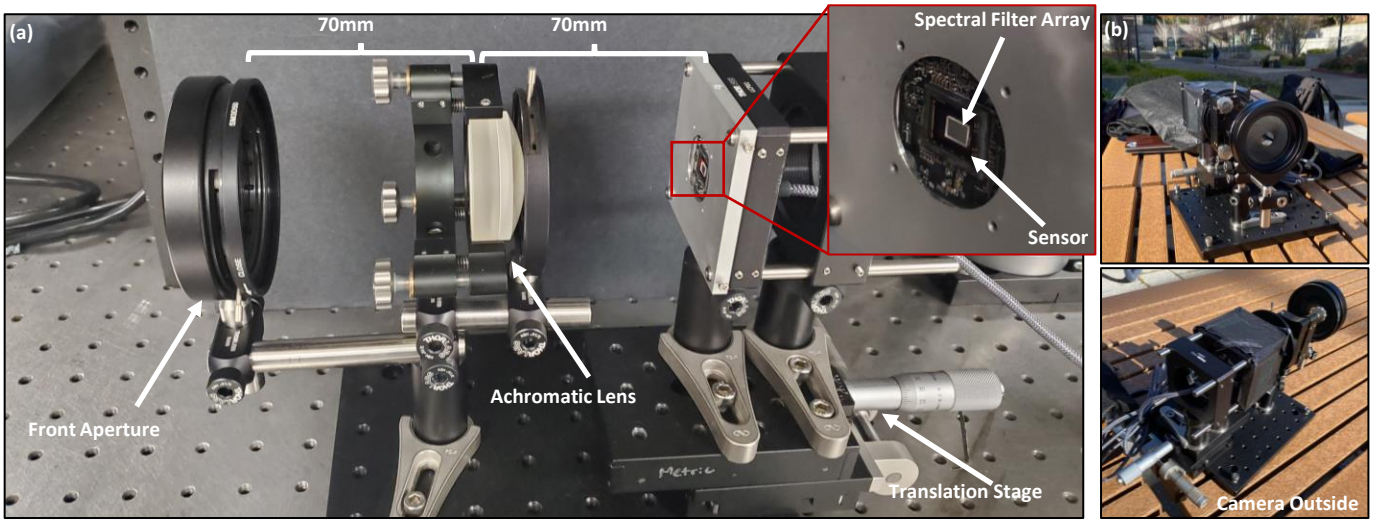

Fig. 3. Photos of experimental prototype. (a) Labeled photo of the experimental prototype in the lab, mounted on a manual translation stage, and aligned by hand. (b) Experimental prototype capturing outdoor scene pictured in Fig. 1. While outside, we use an additional dark hood to limit stray light to the sensor.

focal length of $70\,\mathrm{mm}$ and a board-level DMM37UX178-ML camera from Imaging Source with a monochromatic CMOS Sony IMX 178 sensor. Our sensor consists of 3072×2048 single-photo-diode pixels, each with a height and width of $2.4\,\mu\mathrm{m}$. To defocus our system, we mount the camera sensor on a z-translation stage and take measurements at increments of $0.5\,\mathrm{mm}$. We place an aperture with a diameter of $12\,\mathrm{mm}$ one focal length ($70\,\mathrm{mm}$) in front of the lens to limit the field of view, and another directly behind the lens to limit high-angle aberrated light.

We calibrate the PSF of our system using a point source placed $2\,\mathrm{m}$ away from the front aperture-stop. Using the above systems specifications, we calculate the diffraction limit of our system at $2\,\mathrm{m}$ as:

$$\delta = 1.22\frac{\lambda}{D} \times 2\,\mathrm{m} \approx 18.7\,\mu\mathrm{m} \tag{8}$$

when λ is the minimum wavelength of our system, $390\,\mathrm{nm}$, and D is the size of our lens. As a result, we select a $5\,\mu\mathrm{m}$ pinhole to mask our point source.

Since our spectral filter and sensor pixels have different sizes and are not perfectly aligned, a thorough spectral calibration is needed to determine the spectral response at each pixel of our camera sensor. Spectral calibration is performed without the lens in place using a monochromatic beam ($8\,\mathrm{nm}$ width) produced by an Oriel Cornerstone $1/3$m scanning monochromator. We record the sensor response behind the mask at increments of $2\,\mathrm{nm}$, ranging from $390\,\mathrm{nm}$ to $870\,\mathrm{nm}$. After calibration, we condense the mask using a moving average with a stride of 8, resulting in 30 channels from $390\,\mathrm{nm}$ to $870\,\mathrm{nm}$ with a spectral step size and bandwidth of $16\,\mathrm{nm}$.

4.2 Software Details

We perform reconstructions on a single RTX A6000 GPU with 48GB of VRAM. Images are cropped to a 1260×1860 patch covered entirely by the spectral filter, then downsampled by a factor of three to fit our largest models in memory. FISTA reconstructions are run for 800 iterations with anisotropic TV parameters of 1.30, 0.858, and 0.13 in the y, x, and λ directions, respectively. The algorithm usually converges in 300-400 iterations, taking 5-10 minutes.

Learned models are trained using a combination of publicly available benchmark datasets: the Geissen fruit internals dataset [41] (42 images of varying size), the BGU iCVL Hyperspectral Image Dataset [42] (197 images of size 1392×1300), the Harvard Real-World Hyperspectral Images Database [43] (81 images of size 1040×1392), and the public benchmark dataset Pavia University (2 images of size 610×340). To support a batch size of 4 during training, we patch our dataset into 5653 unique 256×256 pixel patches, split between training, test, and validation partitions with ratios of 0.70, 0.15, and 0.15, respectively. For all learned reconstructions, we train for 100 epochs on mean squared error (MSE) loss. We use the ADAM optimizer [44] with a learning rate of 1×10^{-4}, weight decay of 1×10^{-3}, and an exponential learning rate schedule with a decay factor of 0.94. Trainings generally converge in 50-70 epochs (about 7-8 hours wallclock).

5 System Analysis

5.1 Robustness to Noise

A key consideration for our system is robustness to noise in both the calibration data and raw measurements. To examine this, we evaluate the reconstruction quality of both FISTA and the learned reconstruction network on simulated data with multiple sources of noise in Fig. 4(b).

To simulate calibration noise, we sample from a normal distribution $\mathbf{N}1 \sim \mathcal{N}(0,\sigma_1)$ and add this noise to our spectral filter calibration function $\mathbf{F}\lambda[x,y]$. Using this noisy calibration, we simulate each measurement $\mathbf{b}_i$ with our forward model A_i.

We then introduce signal-dependent shot noise using a stochastic operator $\mathbf{N}_2(\mu,\cdot)$, which scales the max-normalized measurement by μ, samples from a Poisson

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

distribution with rate equal to the scaled pixel value, and then divides by μ. This operator is defined as:

$$\mathbf{N}_2(\mu, \mathbf{s}) = \frac{\text{Poisson}(\mu \cdot \mathbf{s})}{\mu} \tag{9}$$

Finally, we add a signal-independent noise term $\mathbf{N}_3 \sim \mathcal{N}(0, \sigma_2)$ to simulate read noise.The resulting model of noisy measurement, $\tilde{\mathbf{b}}_i$, can be expressed as:

$$\tilde{\mathbf{F}}_\lambda[x, y] = \mathbf{F}_\lambda[x, y] + \mathbf{N}_1 \tag{10}$$

$$\tilde{\mathbf{b}}_i = \mathbf{N}_2(\mu, \mathbf{A}(\tilde{\mathbf{F}})_i \mathbf{v}) + \mathbf{N}_3, \tag{11}$$

We use this noise model to run reconstructions on a subset of our 848 patch test set. We vary the read noise parameters σ_1 and σ_2 logarithmically from 1×10^{-6} to 1×10^{-1}. Similarly, we vary the photon count parameter μ logarithmically from 1×10^{6} to 1×10^{1} to simulate scenes of varying intensity levels. Higher values of μ correspond to brighter scenes with more incident photons and thus better SNR under the shot noise model, while lower μ values represent dimmer conditions resulting in lower SNR measurements.

We compute the noisy reconstruction with both FISTA and the learned reconstruction model. For each set of reconstructed hyperspectral cubes, we compute mean squared error (MSE), cosine similarity (COSSIM), peak signal-to-noise ratio (PSNR), and structural similarity index metric (SSIM) against a subset of our 848 patch test set and plot the average of each metric in Fig 4(c).

FISTA reconstructions see a dramatic loss in quality when the calibration and read noise (σ_1, σ_2) are greater than 1×10^{-3}. Conversely, learned reconstructions show lower sensitivity to noise, demonstrating insensitivity to noise in the spectral calibration through σ_1 of 1×10^{-1}, and read noise through σ_2 of 1×10^{-2} (Fig 4(c) columns 1 and 3). These results suggest that while both reconstruction algorithms demonstrate moderate robustness to noise, learned models may be more robust to high amounts of system miscalibration and read noise experienced in experimental conditions. This may be explained by their learned spectro-spatial priors of natural image structure.

We find similar trends in reconstruction quality on noisy experimental data. To best represent such experimental conditions, we perform all subsequent simulation studies under the effect of simulated shot noise of μ of 1×10^{4} and simulated calibration and read noise σ_1, σ_2 of 1×10^{-3}.

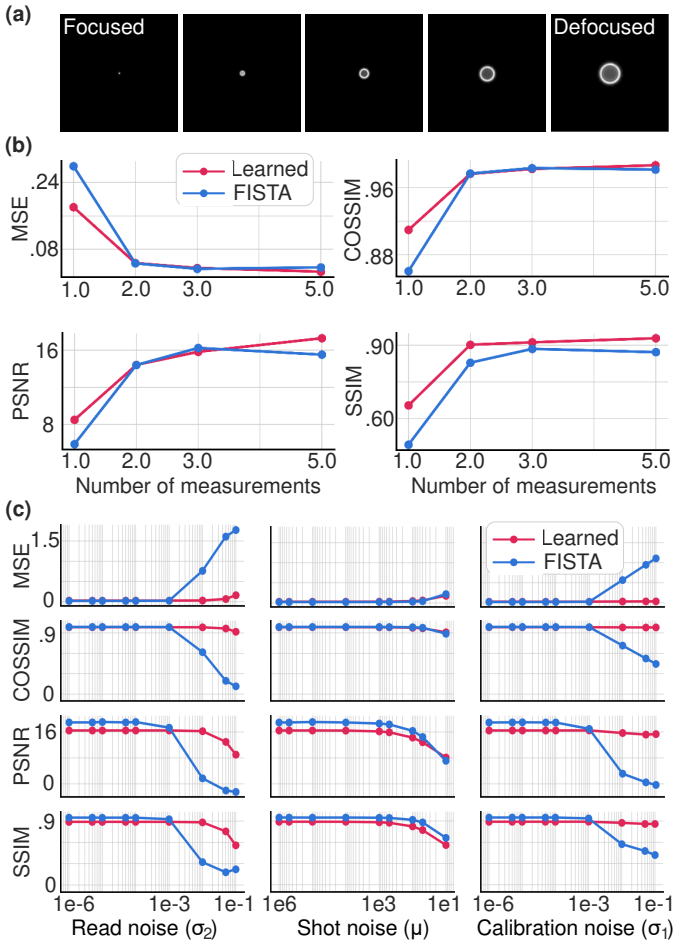

Fig. 4. Ablation studies varying the system conditions against reconstruction quality in simulation. (a) PSFs from each defocus level. (b) We vary the number of measurements fed into each reconstruction algorithm. Intermediate levels of defocus improve performance for both FISTA and the learned network. Reconstruction quality for both models plateaus after 3 measurements, with the learned model showing minor improvement with 5 measurements, and FISTA reconstruction quality declining slightly. (c) We study the impact of read, shot, and calibration noise on reconstruction quality, varying additive read noise standard deviation σ_1, maximum measurement photon count μ, and calibration noise standard deviation σ_2, respectively. While both reconstruction algorithms tolerate moderate amounts of noise, FISTA reconstructions demonstrate slightly higher sensitivity to read and shot noise, and substantially higher sensitivity to read and calibration noise than learned reconstructions. The reader should note that the x-axis for shot noise has larger μ values on the left and lower on the right.

5.2 Levels of Defocus

We vary the number of defocused measurements used in reconstruction from one to five to evaluate the impact on reconstruction quality. In the case of a single defocus level, we include the *most* defocused measurement, whose PSF is approximately the size of 1 filter super-pixel. For comparisons involving more than one defocused measurement, we include both the most in-focus and the most defocused measurements. Any additional measurements are selected such that their levels of defocus are uniformly distributed between these two extremes. We compute our benchmark metrics for each set of reconstructed hyperspectral cubes and plot the average of each metric in Fig. 4(a). Reconstruction quality increases as the number of measurements increases, which shows that the problem becomes better conditioned as a function of the number of measurements, as expected. After 3 levels of defocus reconstruction quality plateaus, suggesting that information recovered from further intermediate levels of defocus is minimal.

5.3 Resolution Analysis

Here, we discuss our system's spatial, spectral, and temporal resolution.

5.3.1 *Spatial resolution*

Since we use multiple measurements with unique PSFs, there is no traditional metric, such as the full-width half-max

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

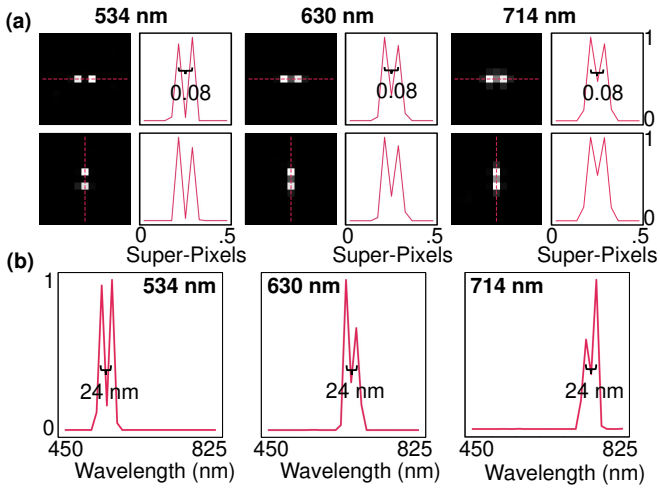

Fig. 5. Simulated spatial and spectral resolution reconstructions demonstrate uniform resolution across the bandwidth of interest. All reconstructions are performed in the presence of noise parameters described in Sec 5.1. (a) Two-point spatial resolution is 0.08 super-pixels, or 12.5× super-pixel super-resolution, in both the horizontal and vertical directions. (b) Spectral resolution is 24 nm across the system bandwidth.

of a PSF's autocorrelation peak, that accurately captures the two-point resolution of our system. Therefore, we calculate the two-point resolution through simulations. Specifically, we simulate two delta functions of a single wavelength, propagate the volume through our system's forward model to generate the simulated measurement, and then reconstruct the image using FISTA with a native sparsity prior. We incrementally increase the spatial distance between the two delta functions by one pixel until they become resolvable. To account for the spectral filter's non-uniformity, we perform these calculations in both the vertical and horizontal directions at $534\,\mathrm{nm}$, $630\,\mathrm{nm}$, and $714\,\mathrm{nm}$, to cover the system's expected bandwidth. Our findings, as shown in Fig. 5(a), reveal that we can resolve simulated points separated by 0.08 super-pixels, or $12.5\times$ super-pixel super-resolution. This corresponds to pixel-limited resolution on our sensor.

5.3.2 Spectral resolution

Similar to the spatial resolution determination, we calculate the spectral resolution through simulations. We simulate a single-point source with two distinct spectral peaks, propagate this two-peak spectral signal through our system's forward model to generate the simulated measurement data, and then reconstruct the spectrum using FISTA with an L1 sparsity prior. The wavelength separation between the two peaks is incrementally increased until they become resolvable in the reconstructed spectrum.

To account for potential variations in spectral resolution across the system's operational bandwidth, we perform these calculations at three different wavelengths: $534\,\mathrm{nm}$, $630\,\mathrm{nm}$, and $714\,\mathrm{nm}$. For each initial peak wavelength, the second peak's wavelength is progressively increased during the resolution analysis. As illustrated in Fig. 5(b), peaks separated by $24\,\mathrm{nm}$ are resolvable. Again, we are calibration-limited due to computational constraints.

5.3.3 Temporal resolution

Since our system is multi-shot, the temporal resolution is set by the time it takes to capture n defocused measurements. This capture time consists of two parts: time to refocus the measurement, t_{refocus}, and the exposure time, t_{exposure}:

$$t_{\text{total}} = n * t_{\text{exposure}} + (n-1) * t_{\text{refocus}}. \tag{12}$$

In our prototype system, we manually refocus the lens, limiting our temporal resolution. However, if this system were integrated into a modern smartphone, the time to refocus could dramatically decrease. Modern smartphones use optical image or sensor-shift stabilization, translating the lens and sensor to maintain sharp images during handheld capture. These systems operate at up to 1kHz and 5kHz [45], respectively, placing the measurement refocusing time on the order of milliseconds.

The required exposure time depends on the scene luminance and the narrowband spectral filter transmission. In practice, the refocusing time is negligible compared to the exposure time, so our total capture time is $t_{\text{total}} \approx n * t_{\text{exposure}}$. Thus, our method's temporal resolution is limited by the exposure time. In practice, we used $2\,\mathrm{ms}$ exposures for daylight, giving an effective frame rate of 167 Hz.

5.4 Condition Number Analysis

We perform a local condition number analysis [46] to gauge our system's object-dependent, multi-point resolution. A local condition number analysis probes the invertibility of a forward model, given known support.

Given N points in the scene, $\mathbf{v}$, we form a sub-matrix of $\mathbf{A}$ consisting of the columns of $\mathbf{A}$ with non-zero voxels. The reconstruction problem is considered ill-posed if this sub-matrix is ill-conditioned, which can be quantified by the condition number of the sub-matrix. We calculate the local condition number as a function of the separation distance between N points, expecting the condition number to be worse when points are close together (less than one super-pixel apart) and better when points are further away.

We compare three scenarios: a narrowband source, a broadband source, and a broadband source with pixel-aligned filters. For each scenario, we compare the 4-point (sparse) and 100-point (dense) local condition number as a function of separation distance for a single-shot focused system, defocused system, Spectral DiffuserCam [7], handshake, and our multi-shot defocus approach. Both multi-shot methods assume five defocus measurements and take the minimum condition number of the sub-matrices of $\mathbf{A}$. For all methods, we average across multiple shifts to compensate for the effect of the filter function, choosing the maximum condition number across shifts.

While Spectral DiffuserCam has a lower condition number for a narrowband source, our method has comparable or better performance for a broadband source, which is more representative of natural scenes, and outperforms Spectral DiffuserCam for dense scenes, Fig. 6. Furthermore, our method has a lower condition number than the focused, defocused, or handshake methods. This implies that our method is better conditioned than alternative methods, and should outperform the other methods, especially for dense scenes.

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

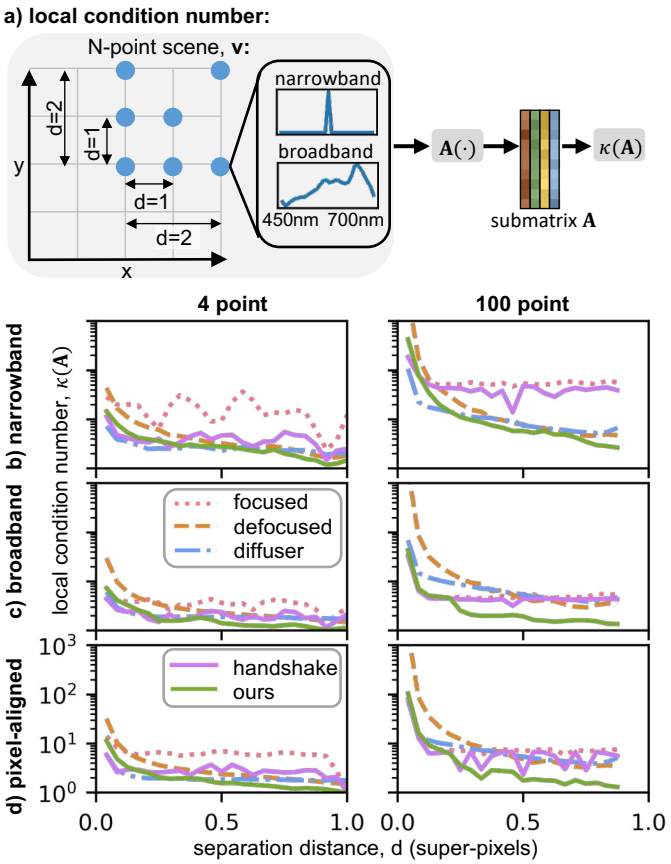

Fig. 6. Local condition number analysis provides a metric for how ill-conditioned an inverse problem is as a function of scene density and feature separation, d. (a) We compare the local condition numbers for both a sparse (4-point) scene and a dense (100-point) scene, assuming (b) narrowband point sources, (c) broadband sources, or (d) broadband sources with a pixel-aligned spectral filter. While Spectral DiffuserCam has a lower condition number for narrowband point sources, our method is comparable or better for broadband sources, which are more representative of natural scenes. Our method outperforms the other methods for dense, broadband scenes.

5.5 Filter Misalignment

Since our prototype system uses a spectral filter mask that is glued onto an imaging sensor, the alignment between filters and sensor pixels is imperfect. Often, sensor pixels fall on the border between two filters, resulting in a multi-peak response at that pixel. To probe our system's performance given ideal, pixel-aligned filters, we simulate a pixel-aligned filter and compare this to our experimentally-calibrated, misaligned filter mask in Table. 1 and Fig. 6. In simulation, pixel-aligned systems show better performance than misaligned filters. Thus, we expect performance would improve with mass-manufactured, on-chip spectral filters.

6 Results

6.1 Simulation Results

We begin with simulated results to analyze our system's viability without the challenges of experimental calibration, such as PSF calibration, optical alignment, filter angular dependence, and noise.

In Fig 7 and Table 1, we benchmark the performance of our system against other single-shot and few-shot imaging models that leverage spectral filter arrays. For each model, reconstructions are run over the entire 848 patch test set, and benchmark metrics are computed against the ground truth patch. FISTA reconstructions are run for 800 iterations with anisotropic TV. Learned reconstructions are performed by an appropriately PSF-conditioned U-Net, trained until convergence on a 3894 patch training set using MSE loss. To ensure a fair comparison, we limit the number of measurements used in Handshake reconstructions to match the number of measurements in our proposed system. For further details on the ideal number of measurements observed for each reconstruction algorithm see the supplementary material.

In the absence of noise and model mismatch, DefocusCam using FISTA outperforms both comparison methods and the deep reconstruction network. As seen in Fig. 7(b), DefocusCam with FISTA performs especially well on dense scenes with complicated spectral profiles like those shown in line profiles 3 and 4.

In addition to lower spatial resolution, handshake reconstructions suffer from heavy spectral artifacts, suggesting insufficient filter coverage. This results in a low cosine-similarity score in Tab. 1 and inconsistent performance across the test points in Fig. 7(b). Finally, using a non-compact randomizing optic in the DiffuserCam reconstructions causes it to perform poorly in dense scenes like those shown in Fig. 7.

TABLE 1
Simulated Reconstruction Quality Ablation Study

Model	MSE	COSSIM	PSNR	SSIM
Spectral DiffuserCam	0.3749	0.8125	4.923	0.4591
Handshake (5) (Fista)	0.4632	0.7683	3.723	0.4752
Focused Lens (Fista)	0.5164	0.7418	3.758	0.5976
Focused Lens (Learned)	0.1849	0.9075	7.860	0.6050
Defocused Lens (Fista)	0.1695	0.9152	8.917	0.6707
Defocused Lens (Learned)	0.2886	0.8557	5.852	0.4765
Ours (Fista)	0.0235	0.9882	17.754	0.9288
Ideal-filter case (Fista)	0.0249	0.9876	17.968	0.9277
Ours (Learned)	0.0339	0.9830	16.612	0.8842

6.2 Experimental Results

Next, we present experimental results from our system, showcasing our system's ability to reconstruct high-resolution images from dense scenes under varying ranges and lighting conditions, achieving greater than 3× sub-super-pixel super-resolution.

We first compare the reconstruction quality in spatial and spectral domains between FISTA and our learned reconstruction network (see Fig. 8). To help generalization to observed experimental conditions, we augment the learned network's training data with noise from the model described in Eq. 11. While FISTA outperforms our learned method in simulation, the learned network outperforms FISTA on experimental data, generating reconstructions with more accurate spectra and fewer artifacts. This aligns with FISTA's higher sensitivity to noise observed in Fig. 4, particularly from miscalibration. Both reconstruction algorithms clearly resolve all three peaks in the USAF target lines of group 2 element 5 in less than one superpixel, or

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

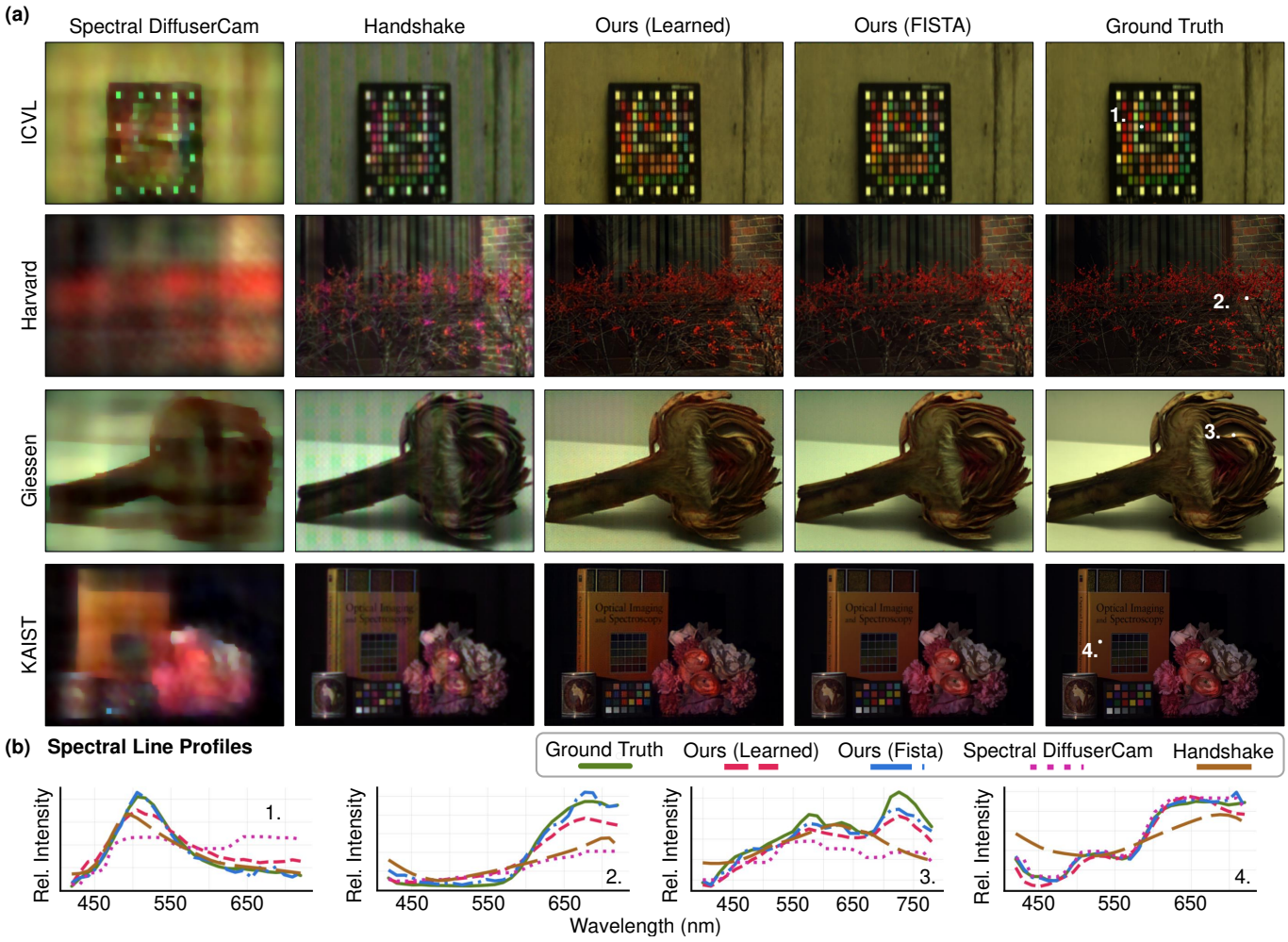

Fig. 7. Simulation results comparing our method to alternative methods for bright and dense scenes. All reconstructions are performed in the presence of noise parameters described in Sec 5.1. (a) False-color projections of reconstructed hyperspectral images from different methods: snapshot Spectral DiffuserCam, 5-shot random handshake, 3-shot DefocusCam with a learned reconstruction network, 5-shot DefocusCam with FISTA + 3DTV. Source datasets for each image [42] [43] [41] [47] are labeled on the left. (b) Spectral line profiles from each datacube averaged in 3-pixel cells around the sample location. Source points are marked in the corresponding ground truth image.

25 sensor pixels, demonstrating at least $3\times$ sub-super-pixel super-resolution on experimental data. Both algorithms fail to consistently resolve peaks beyond target lines of group 3 element 1 (Fig. 8(a) columns 2 and 4).

Next, we present reconstructions from three dense, real-life scenes in Fig. 9(a) columns 2-4, varying both range and lighting conditions. Indoor scenes are taken approximately $2\,\mathrm{m}$ from the camera, while outdoor scenes are taken at distances ranging up to $300\,\mathrm{m}$ from the camera (see example reconstruction in Fig. 1). We use point sources in the scene to estimate sensor offset and translation to compensate for shift in the image plane when imaging far-away objects.

As it demonstrates higher robustness to calibration error and noise found in experimental measurements, we choose the learned reconstruction algorithm for each of these results. In all reconstructions, we can clearly resolve sub-super-pixel features within the images and recover spectral profiles that closely track the ground truth spectra.

7 Discussion

7.1 Limitations and Future Work

One limitation of all spectral-filter-array-based cameras is light throughput. Narrowband filters in spectral-filter-array cameras limit light collection, requiring longer exposures and creating a trade-off between spectral and temporal resolution. Future work could optimize filter design based on computational constraints, expected object distributions, and application-specific spectral requirements to improve light efficiency and temporal resolution.

Due to our experimental prototype's glued-on filter array, tedious calibration with a monochromator is required. Directly fabricating the filter array on the CMOS camera would ensure pixel-aligned filters and simplify calibration. Additionally, the Fabry-Perot filters used in our spectral array suffer from angle sensitivity and flare effects [48], which we omit in our forward model. While our deep learning-based reconstruction exhibits robustness to model mismatch, a calibration approach, similar to the camera-in-the-loop used in Neural Holography [49], could further improve performance.

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

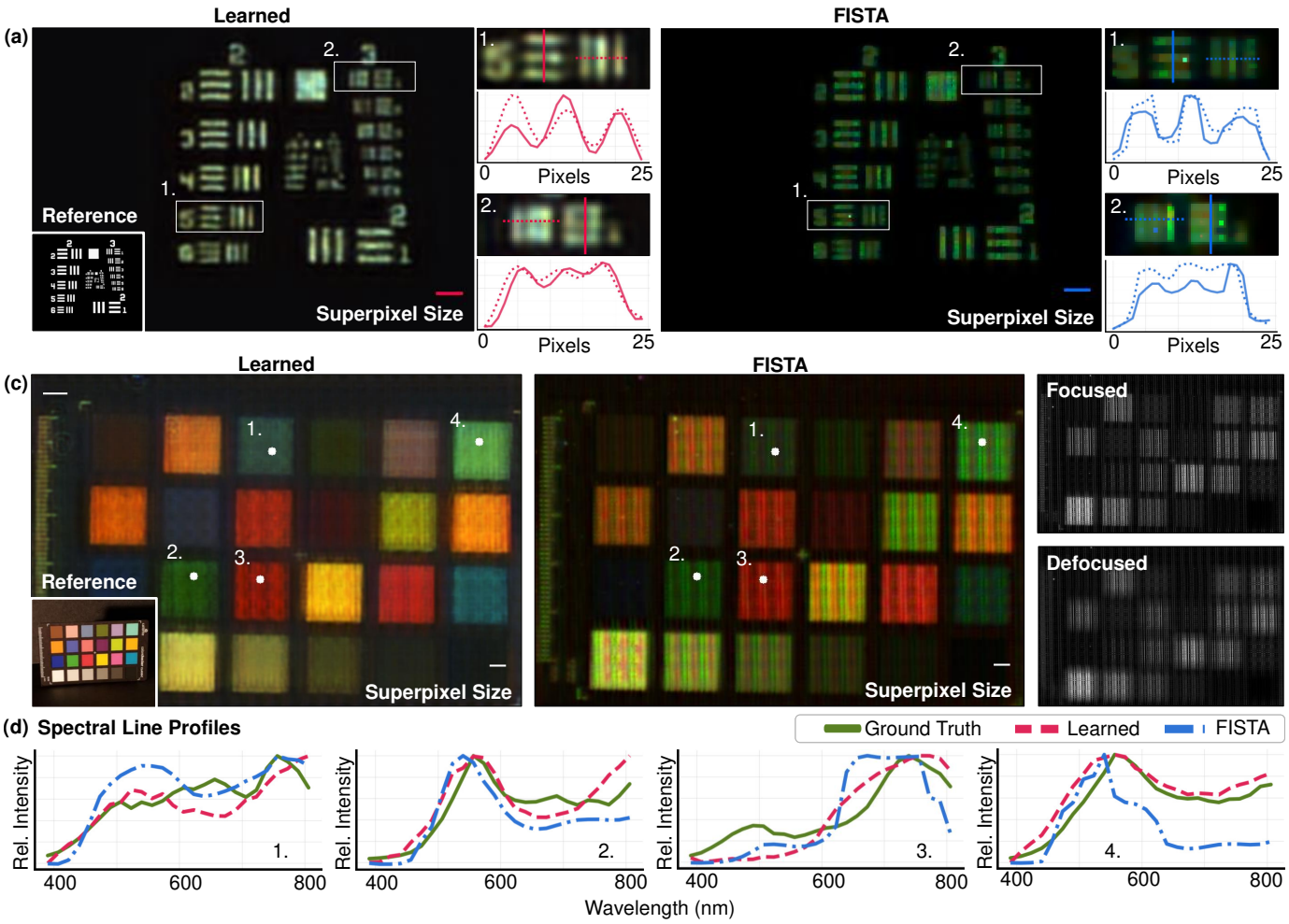

Fig. 8. Experimental results comparing FISTA and our learned reconstruction algorithm. (a) False color experimental reconstructions of a USAF resolution target negative. Spatial line profiles the length of a single superpixel (25 sensor pixels) are plotted along reconstructed horizontal and vertical target lines for both models. Both models resolve the 3 peaks in target lines 2.5, and fail to resolve peaks after target lines 3.1. (c) False color experimental reconstructions of a Calibrite color-checker palette. (d) Spectral line profiles from each reconstruction algorithm taken from points marked in the scene and compared against reference profiles recorded with an Ocean Optics HR2000+ spectrometer. Learned reconstructions consistently demonstrate more accurate recovery and fewer spectral artifacts.

Our current implementation assumes static scenes with objects sufficiently far from the camera to limit depth-dependent PSF variation, limiting applications to motion slower than $167Hz$ and scenes at a distance greater than approximately $1\,\mathrm{m}$. Incorporating temporal and 3D priors could extend capabilities to faster-moving objects and shallow depth-of-field scenarios.

7.2 Conclusion

In this work, we presented a novel hyperspectral imager that leverages multiplexing through defocus and a spectral filter array to achieve super-resolution hyperspectral imaging. Our approach is scalable to most filter-array-based cameras since we require only lens focus control to capture measurements and allows users to customize the spectral filter array for specific imaging applications. We studied the impact of varying numbers of defocused measurements and amount of noise on the system's performance, demonstrating promising robustness to imaging conditions found in real-life photography. To validate our system's capabilities, we built an experimental prototype and demonstrated greater than $3\times$ sub-super-pixel super-resolution on challenging indoor and outdoor scenes.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1752814 and by STROBE: A National Science Foundation Science & Technology Center under Grant No. DMR 1548924. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

This publication has been made possible in part by CZI grant DAF2021-225666 and grant DOI (https://doi.org/10.37921/192752jrgbnh) from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation (funder DOI 10.13039/100014989). Laura Waller is a Chan Zuckerberg Biohub investigator.

The authors would like to thank Viavi Solutions (Santa Rosa, CA) for supplying the spectral filters used in our experimental prototype, as well as Neerja Aggarwal, for

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

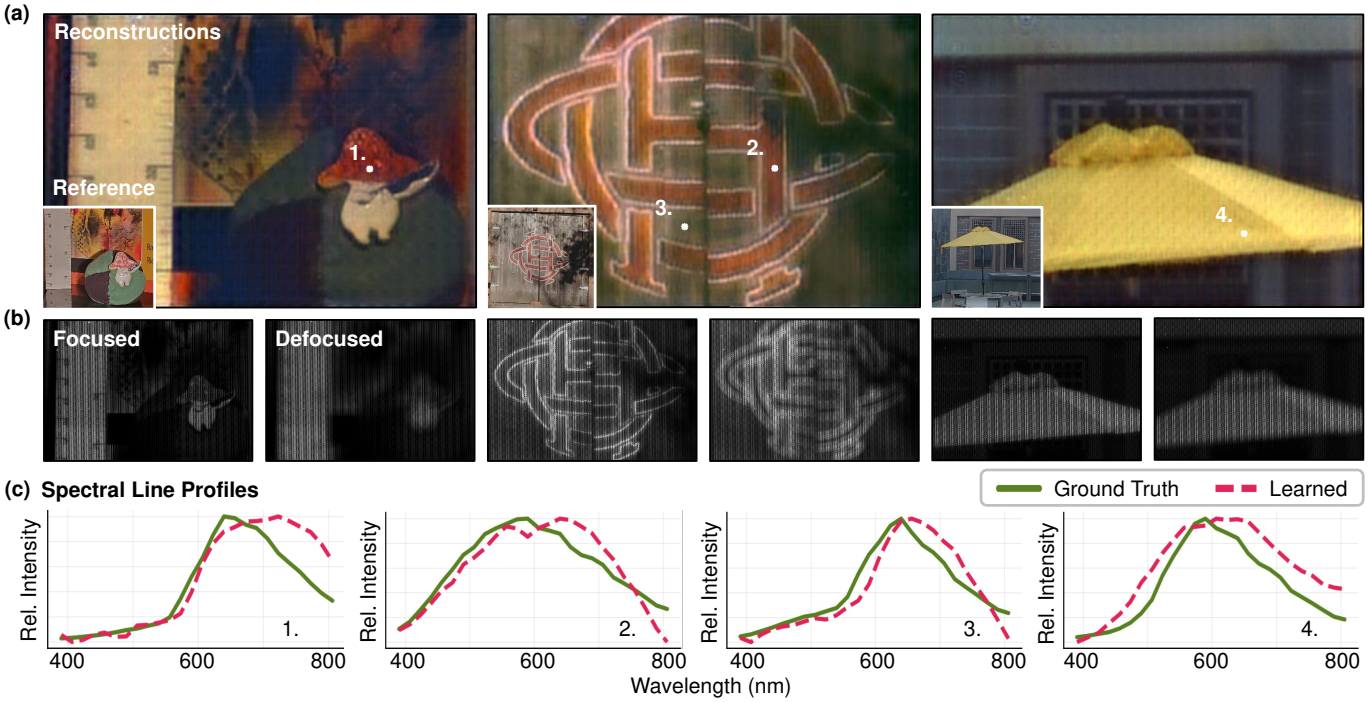

Fig. 9. Experimental results on challenging scenes. (a) Reconstructions from three scenes: and one dense indoor scene illuminated by a halogen lamp, and two outdoor scenes under natural light (left to right). Reference scenes are taken with an RGB cellphone camera. (b) Focused and defocused measurements for each scene. (c) Line spectral profiles from points in the dense scenes. Source points are marked in the corresponding reconstruction. Ground truth spectra are measured with a Thorlabs CCS200/M point spectrometer.

generous hands-on assistance, and Amit Kohli and Dekel Galor for discussions both in and out of the lab.

REFERENCES

[1] T. Adão, J. Hruška, L. Pádua, J. Bessa, E. Peres, R. Morais, and J. J. Sousa, "Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry," *Remote sensing*, vol. 9, no. 11, p. 1110, 2017.

[2] S. V. Panasyuk, S. Yang, D. V. Faller, D. Ngo, R. A. Lew, J. E. Freeman, and A. E. Rogers, "Medical hyperspectral imaging to facilitate residual tumor identification during surgery," *Cancer biology & therapy*, vol. 6, no. 3, pp. 439–446, 2007.

[3] Y.-Z. Feng and D.-W. Sun, "Application of hyperspectral imaging in food safety inspection and control: a review," *Critical reviews in food science and nutrition*, vol. 52, no. 11, pp. 1039–1058, 2012.

[4] V. Saragadam, M. DeZeeuw, R. G. Baraniuk, A. Veeraraghavan, and A. C. Sankaranarayanan, "Sassi—super-pixelated adaptive spatio-spectral imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2233–2244, 2021.

[5] N. Delavarpour, C. Koparan, J. Nowatzki, S. Bajwa, and X. Sun, "A technical study on uav characteristics for precision agriculture applications and associated practical challenges," *Remote Sensing*, vol. 13, no. 6, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/6/1204

[6] B. E. Bayer, "Color imaging array," Jul. 20 1976, uS Patent 3,971,065.

[7] K. Monakhova, K. Yanny, N. Aggarwal, and L. Waller, "Spectral diffusercam: lensless snapshot hyperspectral imaging with a spectral filter array," *Optica*, vol. 7, no. 10, pp. 1298–1307, Oct 2020. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-7-10-1298

[8] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics Express*, vol. 15, no. 21, pp. 14 013–14 027, 2007.

[9] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied Optics*, vol. 47, no. 10, pp. B44–B51, 2008.

[10] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 6, pp. 1–11, 2014.

[11] X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. J. Brady, "Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world," *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 95–108, 2016.

[12] S.-H. Baek, I. Kim, D. Gutierrez, and M. H. Kim, "Compact single-shot hyperspectral imaging using a prism," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, pp. 1–12, 2017.

[13] M. A. Golub, A. Averbuch, M. Nathan, V. A. Zheludev, J. Hauser, S. Gurevitch, R. Malinsky, and A. Kagan, "Compressed sensing snapshot spectral imaging by a regular digital camera with an added optical diffuser," *Applied Optics*, vol. 55, no. 3, pp. 432–443, 2016.

[14] D. S. Jeon, S.-H. Baek, S. Yi, Q. Fu, X. Dun, W. Heidrich, and M. H. Kim, "Compact snapshot hyperspectral imaging with diffracted rotation," *ACM Trans. Graph.*, vol. 38, no. 4, Jul. 2019. [Online]. Available: https://doi.org/10.1145/3306346.3322946

[15] P.-J. Lapray, X. Wang, J.-B. Thomas, and P. Gouton, "Multispectral filter arrays: Recent advances and practical implementation," *Sensors*, vol. 14, no. 11, pp. 21 626–21 659, 2014.

[16] A. Lambrechts, P. Gonzalez, B. Geelen, P. Soussan, K. Tack, and M. Jayapala, "A cmos-compatible, integrated approach to hyper- and multispectral imaging," in *2014 IEEE International Electron Devices Meeting*, 2014, pp. 10.5.1–10.5.4.

[17] J. Yang, K. Cui, Y. Huang, W. Zhang, X. Feng, and F. Liu, "Deep-learning based on-chip rapid spectral imaging with high spatial resolution," *Chip*, vol. 2, no. 2, p. 100045, 2023.

[18] S. Feng, Z. Wang, X. Cheng, and X. Dun, "Superposition fabry–perot filter array for a computational hyperspectral camera," *Opt. Lett.*, vol. 48, no. 5, pp. 1156–1159, Mar 2023. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-48-5-1156

[19] Z. Wang, S. Yi, A. Chen, M. Zhou, T. S. Luk, A. James, J. Nogan, W. Ross, G. Joe, A. Shahsafi *et al.*, "Single-shot on-chip spectral sensors based on photonic crystal slabs," *Nature communications*, vol. 10, no. 1, pp. 1–6, 2019.

[20] Z. Wang and Z. Yu, "Spectral analysis based on compressive sensing in nanophotonic structures," *Optics Express*, vol. 22, no. 21, pp. 25 608–25 614, 2014.

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.

[21] P. Mouroulis, R. O. Green, and T. G. Chrien, "Design of pushbroom imaging spectrometers for optimum recovery of spectroscopic and spatial information," *Applied Optics*, vol. 39, no. 13, pp. 2210–2220, 2000.

[22] J. Brauers, N. Schulte, and T. Aach, "Multispectral filter-wheel cameras: Geometric distortion model and compensation algorithms," *IEEE transactions on image processing*, vol. 17, no. 12, pp. 2368–2380, 2008.

[23] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Applied optics*, vol. 49, no. 36, pp. 6824–6833, 2010.

[24] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," *Applied optics*, vol. 54, no. 4, pp. 848–858, 2015.

[25] L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4942–4950.

[26] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 16, no. 8, pp. 824–831, 1994.

[27] M. Subbarao and G. Surya, "Depth from defocus: A spatial domain approach," *International Journal of Computer Vision*, vol. 13, no. 3, pp. 271–294, 1994.

[28] P. Grossmann, "Depth from focus," *Pattern recognition letters*, vol. 5, no. 1, pp. 63–69, 1987.

[29] S. Suwajanakorn, C. Hernandez, and S. M. Seitz, "Depth from focus with your mobile phone," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3497–3506.

[30] Q. Guo, E. Alexander, and T. Zickler, "Focal track: Depth and accommodation with oscillating lens deformation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 966–974.

[31] D. Rajan and S. Chaudhuri, "Simultaneous estimation of super-resolved scene and depth map from low resolution defocused observations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1102–1117, 2003.

[32] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[33] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 417–424.

[34] B. Wronski, I. Garcia-Dorado, M. Ernst, D. Kelly, M. Krainin, C.-K. Liang, M. Levoy, and P. Milanfar, "Handheld multi-frame super-resolution," *ACM Transactions on Graphics (ToG)*, vol. 38, no. 4, pp. 1–18, 2019.

[35] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009. [Online]. Available: https://doi.org/10.1137/080716542

[36] U. S. Kamilov, "A parallel proximal algorithm for anisotropic total variation minimization," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 539–548, 2017.

[37] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 39–56, 2020.

[38] K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Opt. Express*, vol. 27, no. 20, pp. 28 075–28 090, Sep 2019. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-27-20-28075

[39] K. Yanny, K. Monakhova, R. W. Shuai, and L. Waller, "Deep learning for fast spatially varying deconvolution," *Optica*, vol. 9, no. 1, pp. 96–99, Jan 2022. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-9-1-96

[40] S. Saxe, L. Sun, V. Smith, D. Meysing, C. Hsiung, A. Houck, M. Von Gunten, C. Hruska, D. Martin, R. Bradley *et al.*, "Advances in miniaturized spectral sensors," in *Next-Generation Spectroscopic Technologies XI*, vol. 10657. SPIE, 2018, pp. 69–81.

[41] R. Ennis, F. Schiller, M. Toscani, and K. R. Gegenfurtner, "Hyperspectral database of fruits and vegetables," *J. Opt. Soc. Am. A*, vol. 35, no. 4, pp. B256–B266, Apr 2018. [Online]. Available: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-35-4-B256

[42] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *European Conference on Computer Vision*. Springer, 2016, pp. 19–34.

[43] A. Chakrabarti and T. Zickler, "Statistics of Real-World Hyperspectral Images," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 193–200.

[44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[45] PCMag, "The iphone 15 pro max has a tetraprism camera: What the heck is that?" https://www.pcmag.com/news/the-iphone-15-pro-max-has-a-tetraprism-camera, 2024, accessed: March 27, 2024.

[46] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "DiffuserCam: lensless single-exposure 3D imaging," *Optica*, vol. 5, no. 1, pp. 1–9, 2018.

[47] I. Choi, D. S. Jeon, G. Nam, D. Gutierrez, and M. H. Kim, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2017)*, vol. 36, no. 6, pp. 218:1–13, 2017. [Online]. Available: http://dx.doi.org/10.1145/3130800.3130810

[48] E. Sassoon, T. Treibitz, and S. Yoav, "Flare in interference-based hyperspectral cameras," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 10 173–10 181.

[49] Y. Peng, S. Choi, N. Padmanaban, J. Kim, and G. Wetzstein, "Neural holography," in *Proceedings of SIGGRAPH 2020*. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://research.nvidia.com/publication/2020-08_Neural-Holography

Christian Foley is an imaging software engineer at Insitro. His research primarily focuses on the joint design of optics and data-driven algorithms to develop novel imaging capabilities. Prior to joining Insitro, he earned his B.S. in Computer Science from Berkeley, where he worked as an associate researcher in Laura Waller's Computational Imaging Lab.

Eric Markley is a PhD student in the Joint Graduate Program in Bioengineering at UC Berkeley/UCSF advised by Dr. Laura Waller. His research is primarily on data-driven computational imaging system design. Prior to coming to the Bay Area, he earned a B.S. in Biomedical Engineering from the joint program of the UNC–Chapel Hill and NCSU. He also earned a B.S. in Applied Mathematics from UNC–Chapel Hill. While in Chapel Hill, he worked on super-resolution ultrasound under the direction Dr. Paul Dayton.

Kyrollos Yanny received his PhD in Computational Imaging from UC Berkeley and UC San Francisco advised by Laura Waller and funded by NSF GRFP. His work combines optical design, convex optimization, and deep learning to achieve capabilities that are not possible with conventional imaging setups. He graduated from UCLA with a B.S. in Bioengineering in 2016 where he worked with Aydogan Ozcan on designing digital holographic microscopes for water-monitoring.

Laura Waller leads the Computational Imaging Lab, which develops new methods for optical imaging, with optics and computational algorithms designed jointly. She holds the Ted Van Duzer Endowed Professorship and is a Senior Fellow at the Berkeley Institute of Data Science (BIDS), with affiliations in Bioengineering and Applied Sciences & Technology. Laura was a Postdoctoral Researcher and Lecturer of Physics at Princeton University from 2010-2012 and received BS, MEng and PhD degrees from MIT in 2004, 2005 and 2010, respectively. She is a Moore Foundation Data-Driven Investigator, Bakar fellow, Distinguished Graduate Student Mentoring awardee, NSF CAREER awardee, Chan-Zuckerberg Biohub Investigator, SPIE Early Career Achievement Awardee and Packard Fellow.

Kristina Monakhova is an Assistant Professor in the Department of Computer Science at Cornell University. She works on co-designing optics and algorithms to create better, smaller, and more capable cameras and microscopes. Before joining Cornell, Kristina was a postdoctoral fellow at MIT, working with Sixian You and George Barbastathis, and was supported by the MIT Postdoctoral Fellowship for Engineering Excellence. She received her Ph.D. in Electrical Engineering and Computer Sciences from UC Berkeley, where she worked with Laura Waller. During her Ph.D., she was affiliated with the Berkeley Artificial Intelligence Research (BAIR) Lab and was supported by the NSF GRFP fellowship. Her dissertation focused on Physics-Informed Machine Learning for Computational Imaging.

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 01,2025 at 19:47:19 UTC from IEEE Xplore. Restrictions apply.