# Spectral DefocusCam: Compressive Hyperspectral Imaging from Defocus Measurements

**Georgia Channing***
University of Tennessee, Knoxville
Tickle College of Engineering
gchannin@vols.utk.edu

**Abstract**

Hyperspectral imaging is used for a wide range of tasks from medical diagnostics to crop monitoring, but traditional imagers are prohibitively expensive for widespread use. I propose a tunable lens with varying amounts of defocus paired with a 31-channel spectral filter array mounted on a CMOS camera. These images are then fed into a reconstruction network to recover the full 31-channel hyperspectral volume from a few encoded images with different amounts of defocus.

## Introduction

Hyperspectral imaging is a technique that collects and analyses light from across the electromagnetic spectrum rather than categorizing pixels into red, green, and blue (RGB) values. Hyperspectral imaging is useful for applications ranging from medical diagnostics to agricultural crop monitoring; however, traditional scanning hyperspectral imagers are prohibitively slow and expensive for widespread adoption. Snapshot hyperspectral cameras aim to capture a hyperspectral volume in a single encoded image. In this project, I study the design of a new hyperspectral camera that is compact and inexpensive but able to capture high resolution hyperspectral volumes. I propose using a tunable lens that can rapidly change focus paired with a 31-channel spectral filter array mounted on a CMOS camera. By rapidly taking a burst of several images with varying defocus, I hope to encode both high resolution spatial and spectral information. These images will then be fed into a reconstruction network that recovers the full 31-channel hyperspectral volume from a few encoded images (Monakhova et al. 2020).

## Methods

Though a reconstruction method could be developed and tested on a hand-picked prior, using machine learning to simulate the camera's behavior allows us to generate a better prior on the scene and craft a better reconstruction algorithm (Baek et al. 2021). To simulate the camera's behavior and generate our prior, I use a forward model, the outputs of which I use to train the reconstruction network.

*With the support of Dr. Laura Waller, Kristina Monakhova, and Kyrollos Yanny of the UC Berkeley Computational Imaging Lab and the National Science Foundation.


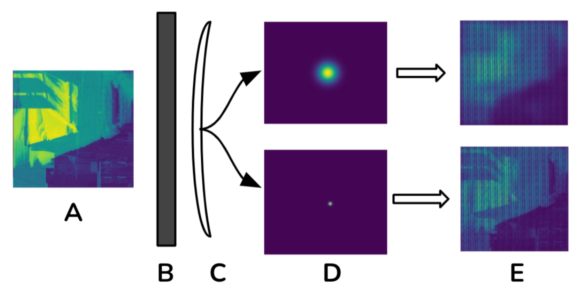


Figure 1: A) Ground truth. B) Sensor. C) Lens. D) Top: Blurry PSF, Botton: Sharp PSF. E) Simulated data.

### Forward Model

A forward model, generally, is used to simulate an outcome or, as in this case, to produce data from some input. Here, the forward model serves as a simulation of the physical camera's behavior by taking our assumed ground truth, the original hyperspectral images, and converting them into single-dimensional color channel images with varying amounts of defocus. The forward model convolves the two point spread functions with each image. In each projection, the 31 color channel image was projected into single-dimensional color channel space. The two resulting tensors were stacked for a final input with dimensions of (256, 256, 2).

### Reconstruction

The reconstruction model takes the simulated images, depicted in Figure 1, and reconstructs them into a 31 color channel hyperspectral volume.

**Model Architecture & Loss** Two model architectures were considered: a two-dimensional U-Net and a three-dimensional U-Net. A U-Net is comprised of a compressive path and an expansive path. The compressive path follows a typical CNN architecture with downsampling, while the expansive path consists of upsampling the feature channels. The distinction between a two-dimensional U-Net and a three-dimensional U-Net is the dimensions of the convolutions.

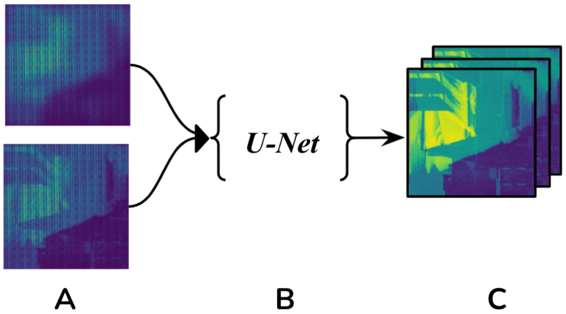


Figure 2: A) Ground truth. B) Reconstruction model. C) Hyperspectral output.

Two loss functions were tested: L1 and LPIPS. The L1 loss function was used as a standard baseline to be compared to the use of the relatively novel LPIPS loss. LPIPS, Learned Perceptual Image Patch Similarity, loss uses deep network activations to measure perceptual similarity (Zhang et al. 2018).

### Data

The raw data is comprised of 50 outdoor images under daylight and 27 indoor images under artificial and mixed illumination from Harvard's Real-World Hyperspectral Images Database (Chakrabarti and Zickler 2011).

In order to generate a larger training sample and to improve the robustness of the model, I augmented the original raw images. Each image was cut into 20 [256, 256] subsections while maintaining all 31 color channels. Each of those subsections was then rotated 90, 180, and 270 degrees. Of all those images, I selected a subset of 5000 images, each of size [256, 256, 31], for our learning process.

## Results

The highest performing model on images without added noise was the 2D U-Net with L1 Loss, which achieved a final test loss of 0.000988. The highest performing model on images with added Gaussian noise was the 3D U-Net with L1 Loss, which achieved a final test loss of 0.00178.

## Discussion

In future, I hope to complete testing on all loss, model architecture, and added noise combinations. I hope to compare this work to others in the field, such as Spectral DiffuserCam (Monakhova et al. 2020), before continuing to test additional model architectures and losses. While the model is able to reconstruct images well enough to be easily recognisable by the human eye, further research is required before any conclusion can be drawn about the best implementation to complement a CMOS camera.

The author would like to thank Dr. Laura Waller, Kristina Monakhova, and Kyrollos Yanny of UC Berkeley for their support of this research. The author would also like to thank the NSF for funding this research.

| Loss | 2D U-Net | 3D U-Net | Noise |
|---|---|---|---|
| L1 | 0.00099 | N/A | No Noise |
| L1 | 0.0018 | 0.0018 | Gaussian |
| LPIPS | 0.0020 | N/A | No Noise |
| LPIPS | 0.020 | 0.019 | Gaussian |

Table 1: Minimal Losses on Test Set

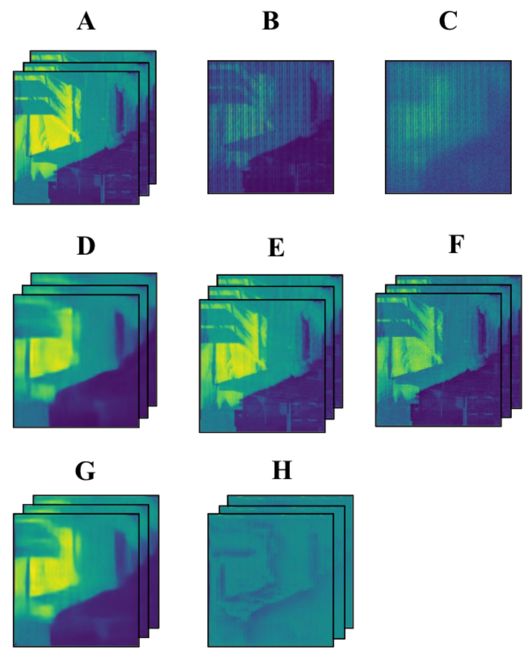


Figure 3: A) Original image slice. B) Image with sharp PSF. C) Image with sharp PSF and 0.1 Gaussian noise. D) 2D U-Net with noise and L1 loss. E) 2D U-Net without noise and L1 loss. F) 2D U-Net without noise and LPIPS. G) 3D U-Net with noise and L1 loss. H) 3D U-Net with noise and LPIPS.

## References

Baek, S.-H.; Ikoma, H.; Jeon, D. S.; Li, Y.; Heidrich, W.; Wetzstein, G.; and Kim, M. H. 2021. Single-shot Hyperspectral-Depth Imaging with Learned Diffractive Optics. arXiv:2009.00463.

Chakrabarti, A.; and Zickler, T. 2011. Statistics of real-world hyperspectral images. In *CVPR 2011*, 193–200.

Monakhova, K.; Yanny, K.; Aggarwal, N.; and Waller, L. 2020. Spectral DiffuserCam: lensless snapshot hyperspectral imaging with a spectral filter array. *Optica*, 7(10): 1298.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. arXiv:1801.03924.