
Single-shot Hyperspectral Imaging via Deep Chromatic Aberration Deconvolution

Corey Zheng

College of Biomedical Engineering, Georgia Tech
czheng45@gatech.edu

Austin Barton

Georgia Tech
abarton40@gatech.edu

Mohammad Taher

Georgia Tech
mtaher3@gatech.edu

Abdulaziz Memesh

Georgia Tech
a.memesh@gatech.edu

Abstract

This paper addresses the challenge of chromatic aberration in snapshot hyperspectral imaging, where the introduction of a third spectral dimension complicates encoding information into a two-dimensional detector plane. While traditional methods rely on scan-based approaches, our proposed method aims to enhance the quality of hyperspectral images by mitigating distortions inherent in snapshot acquisitions by leveraging a blind deconvolution approach with a U-Net neural network architecture for single-shot hyperspectral imaging. Our approach allows real-time correction without the need for complex scanning mechanisms nor knowledge of point spread functions. Through experimental validation, we demonstrate the efficacy of our method in preserving image structure and spectral composition information, contributing to improved imaging throughput and simplified hardware requirements. Our best performing model is capable of restoring spectral information among test data, even in pixel locations with highly varying wavelength intensities, while deblurring and restoring an approximation of the latent sharp image. Our paper represents a simple yet effective approach to circumventing issues in snapshot hyperspectral imaging, providing a practical solution for applications in medical imaging, agriculture, materials identification, and geological surveillance.

1 Introduction

Hyperspectral imaging is a highly important imaging modality, with the ability to capture the unique wavelength emission and reflectance spectra of objects within the field of view. For every spatial pixel within a hyperspectral image, the captured spectrum is a superposition of all unique spectra produced by molecules within the spatial region, and thus is a measure of the chemical composition at each spatial location within the field of view (FOV). This higher-dimensional spectral information has therefore enabled major strides in medical imaging [1], agriculture [2], materials identification [3], geological surveillance [4], and more.

However, hyperspectral images have been difficult to acquire, as their three-dimensional nature (two spatial dimensions and a spectral dimension) cannot be fully represented on electronic image sensors which feature two-dimensional arrays of pixels. Traditional hyperspectral imaging approaches are scan-based, scanning through either one spatial dimension such as pushbroom approaches, or one spectral dimension by utilizing a series of filters [5]. Such approaches limit the imaging throughput speed and also require more complex imaging setups that involving parts such as scanning galvo mirrors, acousto-optical deflectors, filters, or acousto-optical filters.

Thus, a major research interest is *snapshot hyperspectral* imaging, in which the full hyperspectral volume can be captured in a single image frame, greatly increasing the acquisition throughput and eliminating the need for complex scanning components. Early snapshot hyperspectral imaging began by mapping slices of the hyperspectral volume to different spatial positions on a two-dimensional camera detector. Approaches in this category include Computed Tomography Imaging Spectrometers (CTIS) [6] and Integral Field Spectrography (IFS) [7]. These mapping approaches necessitate a trade-off due to their subdivision of the imaging sensor, requiring a reduction of either the field of view (FOV) or resolution in order to fit both the spatial image and spectral components onto a sensor with the same number of pixels.

Motivated by this inherent compromise, various computational based approaches aimed at reconstructing the hyperspectral information without subdivisions of the imaging field have been introduced. In particular, *chromatic aberration* approaches that take advantage of a uniquely varying wavelength dependent point-spread function (PSF) have been used to encode spectral information into the spatial blur of an image [8, 9, 10] in a similar vein as depth-from-defocus approaches [11].

Following this research area, we present our work: Single-shot Hyperspectral Imaging via Deep Chromatic Aberration Deconvolution. We explore using deep U-net networks to perform hyperspectral volume reconstruction from a single chromatically-aberrated monochromatic image. In particular, our contributions compared to related works are (1) independence of costly and customized diffraction masks that require lithography approaches, lowering the barrier to entry and presenting a more accessible snapshot hyperspectral imaging solution, (2) requires only a single monochromatic image at an infinity focus, rather than a set of images at different foci for improved imaging throughput, and (3) usage of a deep network rather than iterative reconstruction process, improving forward model processing speed. We envision these advancements will guide the development of lower-cost, simpler snapshot hyperspectral imaging systems for wide applications.

2 Related Works

In this section we review approaches towards snapshot hyperspectral imaging.

2.1 CTIS

Snapshot hyperspectral imaging began with the introduction of Computed Tomography Imaging Spectrometers (CTIS) [6]. In this approach, an additional dispersive element is introduced into the imaging system, essentially splitting an incoming image into multiple tiled copies, with each copy exhibiting a different amount of radial chromatic blur. This follows the same concept as lightfield imaging [12, 13] but utilizes diffraction orders produced from chromaticity rather than depth. In a sense, such an approach could also be considered a class of *chromatic aberration* approaches. However, it is differentiated from other chromatic aberration approaches by key disadvantages: (1) The authors utilize an iterative reconstruction algorithm (MART) in order to recover the final image, but this algorithm is time consuming and requires measurement of multiple calibration matrices. (2) The usage of a dispersive element to create multiple tiled copies (e.g. 7×7) results in a reduction in the field of view of the image due to the subdivision of the camera sensor area.

2.2 Bioinspired Chromatic Vision Blur

Unlike mammalian eyes which have separate neuroreceptors with sensitivities tailored towards different wavelengths (primarily red, blue, green), cephalopod eyes have broadband neuroreceptors sensitive to all wavelengths. Yet, these animals can still discriminate colors through recognition of the form of the chromatic blur they produce. Inspired by this, Zhan et al. utilized a chromatic aberration system coupled with an SVD-based iterative reconstruction algorithm to restore hyperspectral stacks from monochromatic blurred images [8]. However, the authors had to utilize 3 images of the same object at different distances in order to achieve their results, meaning a compromise between imaging speed, FOV, or resolution is necessary to apply this approach in the field, as three separate images must be acquired for a single reduction. Furthermore, the data the authors demonstrated their results on was extremely well-behaved - the authors only showed reconstruction on images that contained single-peak gaussian-esq spectra, rather than more realistic scenarios, indicating that there may have been potential trouble with separability in their approach via classical iterative deconvolution.

2.3 Diffracted Rotation and End-to-End Optimization

Utilizing dispersion engineering, it is possible to create custom optical elements with controllable chromatic blurring patterns. Jeon et al. [10] formulated an approach, designing a custom diffractive optical element (DOE) that creates a triple-helix point-spread function which rotates based on the wavelength. Thus, the authors were able to increase the separability of the chromatically-aberrated images by eliminating the radial symmetry of their blur kernels. They combined this with a learned unrolled network architecture based on U-net, combined with an iterative algorithm to reconstruct hyperspectral stacks. The author also later followed on this work with a joint hyperspectral-and-depth reconstruction from a single monochromatic image using end-to-end optical optimization [9]. In this approach, end-to-end optimization was performed by implementing a differentiable optics simulation prior to a deep learning U-net and optimizing both systems jointly [14, 15], resulting in a co-optimized PSF encoding of spectral and depth information that could be best reconstructed by the deep U-net. Both of these works showed promising results, however, in both cases, a complex DOE was fabricated via lithographic means [16], which restricts accessibility.

3 Methods

3.1 Image formation model

As a basis for the optical physics of the system, our image formation model utilizes the following assumptions:

1. The point spread function (PSF) is *spatially invariant*.
2. The camera operates in the *hyperfocal range* where all objects are in-focus [17]:

$$PSF(\lambda, z) = PSF(\lambda) \quad (1)$$

3. The camera has uniform sensitivity across the entire light spectrum.

Assumption (1) is frequently utilized in optical imaging fields, as with it an image can be conveniently described as the convolution of an underlying object with the PSF [18]:

$$I(\lambda) = g(\lambda) * PSF(\lambda) \quad (2)$$

Where $I(\lambda)$ is the captured image at some wavelength and $g(\lambda)$ is the underlying unblurred object at some wavelength. Given a monochromatic camera that is equally sensitive to all wavelengths (assumption 3), the object as imaged by our chromatic aberration optical system can be simulated as:

$$I = \int_{\lambda} g(\lambda) * PSF(\lambda) \quad (3)$$

This set of assumptions is often employed in optical imaging to simplify the computational process and has been similarly applied in a related work [10]. We simulated an optical imaging system with chromatic aberration and generated a chromatic PSF for this range shown in Figure 1. Additional information about the parameters of the optical system are presented in Appendix A.1.

3.2 Dataset

We are utilizing a combination of 5 hyperspectral datasets [19, 20, 21, 22, 23], which results in a total of 394 unique hyperspectral images covering objects including human faces, architecture, household objects, landscapes, and more. In general, they're all images of objects or scenes that can be captured in daily life. Some of them are nicely taken pictures, some of them are not. Each datapoint (hyperspectral image) is a 3D tensor of size *channels* \times *width* \times *height*, where each channel represents a band (small range) of the spectrum.

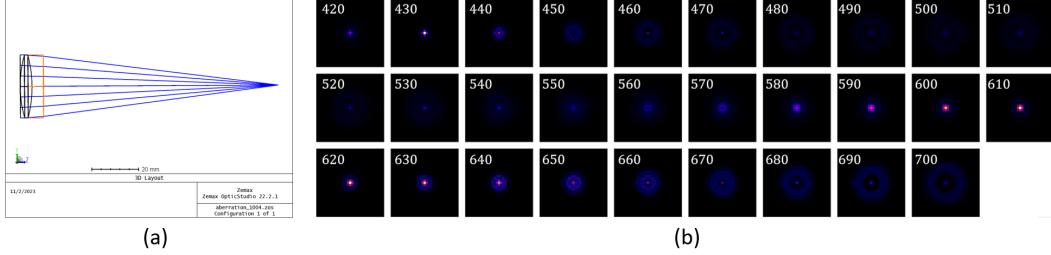


Figure 1: Optical imaging system. **(a)** Ray tracing simulation of chromatic aberration imaging system with singular doublet lens. **(b)** Chromatic PSF in every wavelength band generated from simulated optical system.

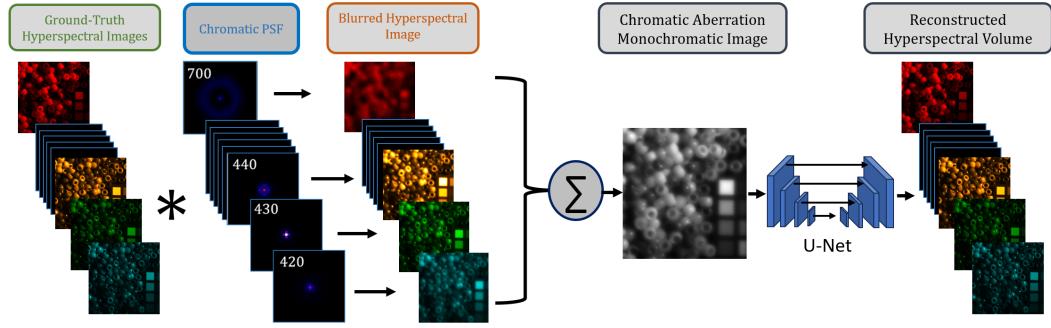


Figure 2: Main model flowchart. Chromatic aberration data is synthesized from ground truth data combined with simulated chromatic PSFs as an input to the deep U-net model for hyperspectral volume reconstruction.

3.3 Data Preprocessing

The model flowchart is depicted in Figure 2. For model training, we synthesize the input monochromatic image by simulating the sensor image of the chromatically aberrated imaging system described above. All data from the 5 hyperspectral datasets were cropped to a common spectral range of 420–710nm, with steps of 10nm bands. Images were resized to a height of 512 pixels, maintaining the same aspect ratio, in order to compensate for variation in image resolutions and maintain feature density to minimize the changes of sampling low-feature empty patches. To synthesize blurred images simulating a chromatically-aberrated optical imaging system, each spectral band was convolved with the corresponding PSF and summed as described in the image formation model, resulting in a monochromatic blurred image representing the image captured by our chromatic aberration imaging system. Finally, we randomly (with uniform probability) sampled 16 128×128 pixel patches from each monochromatic image and extracted the corresponding 29×128×128 pixel hyperspectral volume to form input and ground truth pairs, respectively. The resultant dataset consists of 6304 patches, with no augmentations (additional discussion provided in Section 4.2.4).

3.4 U-Net Architecture

To solve this task, we're using the U-Net, a deep convolutional neural network model developed for biomedical image segmentation by Ronneberger et al. [24]. A U-Net model essentially acts as an Encoder-Decoder Convolutional Neural Network. There's the encoding path that shrinks the image and increases the number of channels to bottleneck the input, then the decoding path which expands the image and decreases the number of channels. It's similar to an autoencoder, but with skip connections from the encoded representations concatenated to the decoded outputs at each upsampling step.

The model inputs a 1×128×128 pixel chromatically aberrated monochromatic image synthesized as described previously, and outputs a 29×128×128 hyperspectral volume in which each channel

Layer (type)	Output Shape	Param #	Layer (type)	Output Shape	Param #
EncoderBlock-1	[N, 29, 64, 64]	8004	EncoderBlock-1	[N, 29, 64, 64]	8004
EncoderBlock-2	[N, 128, 32, 32]	181632	EncoderBlock-2	[N, 128, 32, 32]	181632
EncoderBlock-3	[N, 256, 16, 16]	886272	EncoderBlock-3	[N, 256, 16, 16]	886272
EncoderBlock-4	[N, 512, 8, 8]	3542016	BottleNeck	[N, 512, 16, 16]	3542016
BottleNeck	[N, 1024, 8, 8]	14161920			
DecoderBlock-1	[N, 512, 16, 16]	9178624	DecoderBlock-1	[N, 256, 32, 32]	2295552
DecoderBlock-2	[N, 256, 32, 32]	2295552	DecoderBlock-2	[N, 128, 64, 64]	574336
DecoderBlock-3	[N, 128, 64, 64]	574336	DecoderBlock-3	[N, 29, 128, 128]	37758
DecoderBlock-4	[N, 29, 128, 128]	37758			
<hr/>					
Total params: 30,866,114			Total params: 7,525,570		
Trainable params: 30,866,114			Trainable params: 7,525,570		
Non-trainable params: 0			Non-trainable params: 0		
<hr/>					
Input size (MB): 0.06			Input size (MB): 0.06		
Forward/backward pass size (MB): 1118543.84			Forward/backward pass size (MB): 1085788.59		
Params size (MB): 117.74			Params size (MB): 28.71		
Estimated Total Size (MB): 1118661.65			Estimated Total Size (MB): 1085817.36		

Figure 3: Model architecture summary. (a) HyperUnet, HyperUnet_{LR}, HyperUnet_{SSIM}. (b) HyperUnet_{Reduced}

represents a 10nm wavelength band from 420-710nm. We implement two different architectures, depicted in Figure 3.

U-Nets were originally designed for more precise medical image segmentation, as it exceeded existing models' performance [25]. They were created to achieve precise localization and detail preservation within image tasks. Localization maintains that every outputted pixel can be trained to match a target value that restores the hyperspectral data, and detail preservation ensures that the model doesn't lose the original image's structure.

Ronneberger et al. [24] found that their U-Net outperformed the existing method of patchwise classification via CNNs and drastically improved the computational costs in achieving their task. U-Nets can also be adapted for complicated image tasks. For example, Oktay et al. [26] created Attention U-Nets that could identify organs such as a pancreas, further solidfying the localization capabilites of U-Nets. U-Nets have also been used in CT scans, X-rays, and microscopy. [27].

Lastly, CNNs, which U-Nets are based on, have far greater computational efficiency than existing methods [28]. In addition, the bottle-necking procedure further improves efficiency by reducing image sizes while maintaining accuracy. Therefore, we decided to use U-Nets due to their high performance and low computational complexity.

3.5 Objective Function

We want to minimize the difference of the spectral information between the output and ground truth images as well as restore perceptual quality. Therefore, the objective functions we experiment with are Structural Similarity Index Measure (SSIM) and Mean Squared Error (MSE). SSIM is a metric for image comparisons that compares the overall shapes and patterns, while MSE focuses on the raw pixel value differences [29]. MSE can also be represented as the peak signal-to-noise ratio (PSNR), which accounts for differences in data type representations.

An underlying issue that MSE has in the context of image restoration is the fact that many perceptually different restored images can correspond to one MSE value. Because our problem is taking hyperspectral images that are corrupted by chromatic aberration and restoring them to their corresponding latent sharp images, it is possible that deblurred images outputted by our model may have low MSE but perceptually poor quality. Because of this, we experiment with a combination of MSE and SSIM as an objective function. Details of the mathematical formulation and conceptual explanations are found in A.2 and [29]. Note SSIM indicates high structural similarity for higher values. Hence, we formulate our SSIM loss function as $1 - \text{SSIM}$.

Table 1: Experimental Hyperparameters

Experiment	Loss Function	Optimizer	Learning Rate
HyperUnet	MSE	Adam	0.01
HyperUnet _{LR}	MSE	Adam	[0.1, 0.01, 0.001] at epochs [0,200,600]
HyperUnet _{SSIM}	MSE + SSIM	Adam	[0.1, 0.01, 0.001] at epochs [0, 200, 600]
HyperUnet _{Reduced}	MSE	Adam	[0.1, 0.01, 0.001] at epochs [0, 200, 600]

Table 2: Loss comparison between our models and published hyperspectral imaging approaches

Model	PSNR(dB)	SSIM
Baek et al.[32]	27.96	0.75
Jeon et al.[10]	28.81	0.81
Baek et al.[9]	29.31	0.81
HyperUnet	25.029	0.8681
HyperUnet _{LR}	24.945	0.8991
HyperUnet _{SSIM}	14.887	0.5187
HyperUnet _{Reduced}	22.292	0.7870
Richardson-Lucy Deconvolution	14.713	0.8615

3.6 Baseline Comparison - Richardson-Lucy Deconvolution

Richardson-Lucy (RL) deconvolution is a classical iterative procedure for recovering an unblurred image from a blurred image with a known PSF imaging system [30, 31]. It can be considered a baseline approach for any sort of image deblurring. In the case that RL deconvolution is used to project a two-dimensional image back into a 3-dimensional underlying image by means of a 3D PSF (ie, depth varying PSF), RL deconvolution follows the form [13, 12]:

$$g(z)^{k+1} = g(z)^k \cdot \left(\frac{O}{\sum_z \text{PSF}(z) * g(z)^k} * \text{PSF}'(z) \right) \text{ for all } z \quad (4)$$

Where the g^k is the underlying 3D volume and each slice of g can be computed according to the RL update rule above. Here, the forward projection has been modified, projecting a 3D volume g into the 2D image space via superposition of the convolution of each slice with its corresponding slice of the 3D PSF. In the case of hyperspectral images, the z dimension can simply be replaced with the wavelength dimension λ . We consider RL deconvolution using 15 iterations as a baseline approach towards solving hyperspectral deconvolution and compare the results of our deep network against it. Additional details of implementation and optimization of RL deconvolution hyperparameters is discussed in the Appendix A.3.

3.7 Experimental Setup

We implemented the U-Net architectures in Python using the PyTorch deep learning platform. Models were trained for 1000 epochs, and the epoch which resulted in the lowest validation set loss was considered to be the best model and saved for final analysis. Training required approximately 8-9 hours on a workstation with a NVIDIA Quadro A4000 with 16GB VRAM and an Intel(R) Xeon(R) W-2145 CPU @ 3.70GHz with 512GB RAM. We conducted experiments involving varied model hyperparameters and architectures, as shown in Table 1. In-depth architectures are presented in Figure 3.

4 Results

4.1 Main Findings

The loss of each network variation is presented in Table 2. Of our models, HyperUnet and HyperUnet_{LR} resulted in the best performance according to different metrics (PSNR, SSIM), respectively. Compared to other state-of-the art snapshot hyperspectral imaging methods, the absolute error (PSNR) was worse, but interestingly, our two top models had a lower SSIM loss compared to these approaches. This may indicate that custom dispersive elements utilized in previous approaches to create complex PSFs results in additional image aberrations that, though minimize the MSE, also result in a loss of the overall structure of the reconstructed volumes. In all cases, our deep models outperformed Richardson-Lucy deconvolution (Appendix A.3, Supplementary Figure 8), which demonstrated an inability to perform spectral discrimination with the loss of multiple color channels.

Further analysis of test set sample reconstructions from both HyperUnet and HyperUnet_{LR} revealed that, although HyperUnet_{LR} has similar MSE and PSNR results to HyperUnet, HyperUnet_{LR} failed to accurately recover crucial spectral information (Figure 6) in comparison to HyperUnet (Figure 5). While HyperUnet_{LR} obtained predictions that minimized SSIM_{Loss} more by a considerable margin, the qualitative difference in deblurring is almost negligible and the spectral recovery was significantly worse than HyperUnet. Thus, HyperUnet was selected as the best network of the different experiments.

Spectral profiles are shown in Figure 5. Extracted spectra from spatial points in the image showed remarkable agreement between the predicted and ground truth spectrum, not just for mono-peak spectra as (Beads, pixel 3), but also for varied spectra with multiple local minima (Outdoors, pixel 3). Furthermore, sample outputs from a selection of channels are shown for the same four scenes from our test dataset in Figure 4. A clear spatial deblurring effect can be observed comparing each wavelength band to its PSF-convolved blurred counterpart. Although loss of the highest-frequency details occurs in the reconstructed channels (flowers, Figure 4), we can observe a remarkable similarity between each channel and its ground truth counterpart, far beyond that recovered via the RL-deconvolution baseline (Appendix Figure 8). Overall, the hyperspectral volumes, which were reconstructed from a single-channel blurred image possess a clear visual resemblance to the ground truth, boasting SSIM scores beyond currently existing models.

4.2 Limitations

4.2.1 Training Instability and PSF Degeneracy

Instability of the network was a major factor towards model failure modes. With only variation of the hyperparameters, we found that HyperUnet would fail typically by losing several wavelength bands (Appendix Figure 7). Typical failure modes manifested in gradient explosion during training (Appendix Figure 11). However, even after implementing gradient norm clipping, we still encountered mode collapse by losing wavelength bands, indicating a particularly ill-posed separation of these channels. Such instability is likely in part due to the self-similarity of the PSF. As shown in Figure 1b, the PSF is a gaussian-esque structure with sidelobes. However, there is a clear recurrence of geometry between the ranges of 420-550nm and 560-710nm, in which the PSF sizes and sidelobes are comparable in shape. In part, this is due to the design of the optical system, which was optimized to constrain the 31 band PSFs to a 31×31 pixel region, as allowing the PSF to broaden to an arbitrary size, though reducing the geometric similarities between any two bands, will also cause the required receptive field of the network to increase, rendering it unsuitable with U-net methodologies [33]. However, in doing so, it appears to increase the difficulty of the inverse problem.

4.2.2 MSE+SSIM Objective Function Issues

The HyperUnet_{SSIM} model showed highly unstable training from adding the second loss metric (SSIM) to the objective function and resulted in mode collapse. Specifically, the network effectively shrank pixel intensities of certain bands to 0 and focused on outputting and correcting other bands (as seen in Figure 7). Table 2 shows that this model with SSIM added into the objective function with MSE ended up resulting in a higher SSIM loss despite the model’s objective function including SSIM loss.

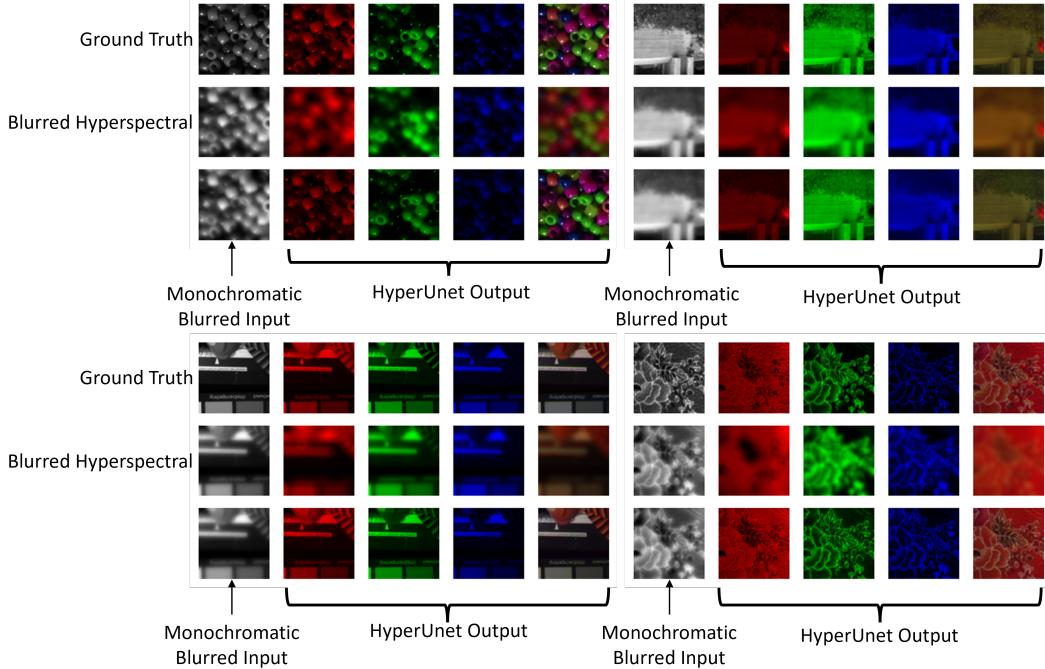


Figure 4: Comparison of Pseudo-RGB image outputs between ground truth sharp images, blurred images, and HyperUnet outputs. For each sample: top row: Ground truth, middle row: PSF convolved ground truth (blurred), bottom row: Deep model output. First column: monochromatic image averaged across all spectral bands, or model input, second column: Psuedo red channel (710nm), third column: psuedo green channel (560nm), fourth column: psuedo blue channel (420nm), fifth column: psuedo-RGB image formed from previous three channels.

4.2.3 Overfitting

Each model suffered from overfitting with the most extreme cases in the HyperUnet and HyperUnet_{LR} (loss curves for each model are shown in the Figures 5, Appendix Figure 10. We experimented with reducing the architecture size, while fixing all other hyperparameters, to combat this overfitting issue and found that although it slightly reduced the overfitting, it performed much worse. We discuss this choice for experimentation and potential implications in Section 5.

4.2.4 Time and Computing Resources

Each model took approximately 8 hours to train over 1000 epochs. With the deadlines of the project, experimentation of different models and hyperparameters was limited. For this reason, augmentations were also not utilized. Because the PSFs are radially asymmetric, all augmentations would have to be applied *prior* to image convolutions, which would greatly increase processing time, and was not implemented for this work.

5 Discussion

5.1 Interpretation of Results

We demonstrate an effective blind approach to deconvolution via deep learning in our best performing model, HyperUnet. However, in general, training these models is highly unstable and sensitive to the initialization of the model. Compared to other works, our model does not utilize customized DOEs which result in complex structured PSFs, nor does it utilize stacks of input images of varying focal lengths. In this sense, our approach enables far more accessible snapshot hyperspectral imaging with a higher theoretical throughput. As a trade-off, our model appears to also be far less *separable*. Compared with customized structured PSFs, our PSFs have lower amounts of asymmetry and further

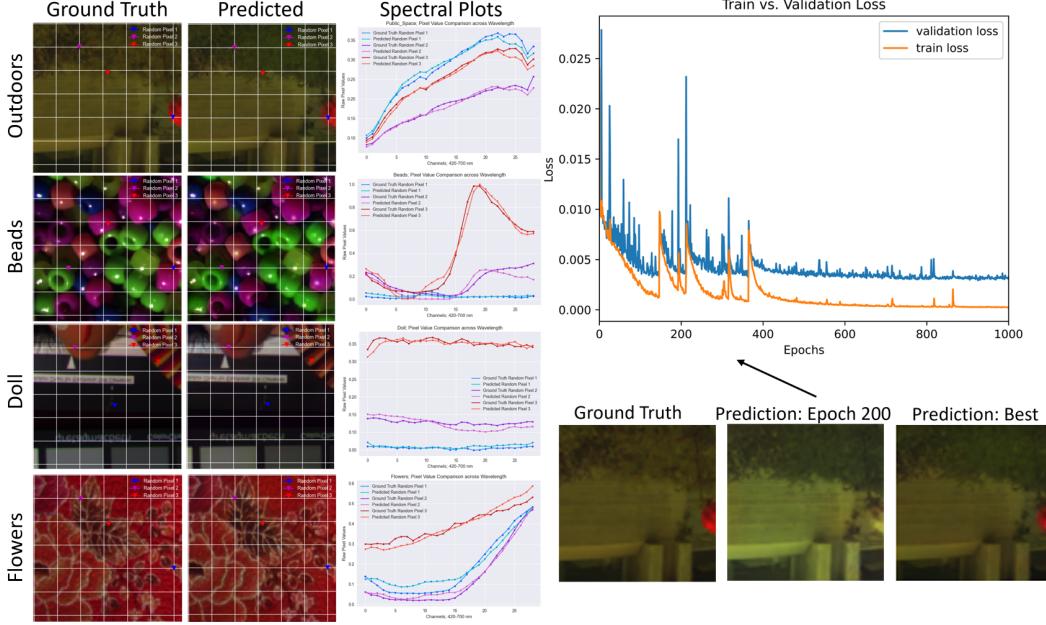


Figure 5: Results from the the fully trained HyperUnet model. Pseudo-RGB ground truth and predicted images on four different images belonging to the test dataset are shown with markings for selected pixel coordinates for spectral plots (far left, top to bottom). Spectral plots comparing ground truth and predicted wavelength intensities for corresponding pixel coordinates (middle, top to bottom). In the top right is the model’s training and validation loss curves. In the bottom right is a comparison of the model’s performance in predicted spectral information between epoch 200 and the best model obtained after epoch 900.

possess self-similarity in different wavelength bands, which may contribute to the failure modes of the model, destroying one channel in any pair of two wavelength bands that have similar PSF structures. Furthermore, unlike other works which introduced multiple focal plane images to effectively construct four-dimensional PSFs (2D spatial structure, focal depth, and wavelength) to increase separability, our work relies on a single image. Thus, in conclusion, our approach boasts higher imaging efficiency and accessibility, but in doing so, increases the ill-posedness of the inverse problem making it unstable for training. This may limit real-world implementation, when not all the optical assumptions of the system are satisfied.

5.2 Implications and Applications

Our approach allows real-time correction without the need for complex scanning mechanisms nor knowledge of point spread functions. Through experimental validation, we demonstrate the efficacy of our method in preserving image structure and spectral composition information. This method can contribute to improved imaging throughput and simplified hardware requirements.

Applications such as medical imaging, agriculture, materials identification, and geological surveillance rely on hyperspectral imaging for a variety of reasons. All of them share a common need to be able to identify chemical composition. Chemical compositions can be studied and identified through sufficient spectral information. Issues with snapshot hyperspectral imaging complicates this. However, we demonstrate effective methods in being able to recover crucial spectral information and the latent sharp image without the need for the knowledge of the point spread functions of the optical lenses.

5.3 Next Steps

Although we were able to produce a model highly capable of our deconvolution task, we failed to make improvements upon the best HyperUnet model and we expect that reproducibility on

different datasets would be quite difficult due to how sensitive the models were to the initialization of weights. Additionally, we found that our models consistently overfit the training data, even in the best performing network, demonstrating a need for further experimentation. Due to limited time and resources, we opted to experiment with a reduced architecture, the ReducedUnet, which ended up performing worse. This illustrates that it is likely that the overfitting is not due to how large the model is and opens up work for further experimentation on more sophisticated and properly tuned architectures for this task. Furthermore, as we suspect the form of the PSF to be highly involved in the failure modes of the network, investigation into asymmetric PSFs without self-similarity that can be achieved with conventional optics rather than custom DOEs will also be important towards enabling accessible snapshot hyperspectral imaging.

6 Acknowledgements

We would like to thank Professor Danfei Xu and all of the teaching faculty for CS 7643 Deep Learning, Fall 2023 at Georgia Tech.

7 Code Repository

All code for this project was implemented in Python and is open-source. Code repository can be accessed through this link: <https://github.com/abarton51/Hyperspectral-Deep-Deconvolution>.

8 Team Contributions

Corey Zheng	Theoretical background, optical simulation & image formation model, dataset collection & dataset synthesis, RL-deconvolution baseline comparison, model training framework, model evaluation framework & metric generation, hyperparameter experimentation, network failure modes & debugging, figure generation, manuscript
Austin Barton	Figure analysis, figure generation, dataloader implementation, pixel comparison procedure & implementation, SSIM implementation, custom loss function implementation, results & analysis, manuscript
Mohammad Taher	Created the model architecture, explained usage of model, graphed & analyzed loss curves for overfitting or lack of convergence, created several model architectures to test performance, manuscript
Abdulaziz Memesh	Literature review research, proposal writing and preparation, training loop implementation, manuscript

References

- [1] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. *Journal of biomedical optics*, 19(1):010901–010901, 2014.
- [2] Bing Lu, Phuong D Dao, Jiangui Liu, Yuhong He, and Jiali Shang. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing*, 12(16):2659, 2020.
- [3] Xingchen Dong, Martin Jakobi, Shengjia Wang, Michael H Köhler, Xiaoxing Zhang, and Alexander W Koch. A review of hyperspectral imaging for nanoscale materials research. *Applied Spectroscopy Reviews*, 54(4):285–305, 2019.
- [4] Sima Peyghambari and Yun Zhang. Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review. *Journal of Applied Remote Sensing*, 15(3):031501–031501, 2021.
- [5] Liang Gao and R Theodore Smith. Optical hyperspectral imaging in microscopy and spectroscopy—a review of data acquisition. *Journal of biophotonics*, 8(6):441–456, 2015.

- [6] Bridget K. Ford, Curtis E. Volin, Sean M. Murphy, Ronald M. Lynch, and Michael R. Descour. Computed tomography-based spectral imaging for fluorescence microscopy. *Biophysical Journal*, 80(2):986–993, 2001.
- [7] Jeremy Allington-Smith. Basic principles of integral field spectroscopy. *New Astronomy Reviews*, 50(4-5):244–251, 2006.
- [8] Shuyue Zhan, Weiwen Zhou, Xu Ma, and Hui Huang. Hyperspectral imaging bioinspired by chromatic blur vision in color blind animals. In *Photonics*, volume 6, page 91. MDPI, 2019.
- [9] Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H Kim. Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2651–2660, 2021.
- [10] Daniel S Jeon, Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H Kim. Compact snapshot hyperspectral imaging with diffracted rotation. 2019.
- [11] Yue Ming, Xuyang Meng, Chunxiao Fan, and Hui Yu. Deep learning for monocular depth estimation: A review. *Neurocomputing*, 438:14–33, 2021.
- [12] Changliang Guo, Wenhao Liu, Xuanwen Hua, Haoyu Li, and Shu Jia. Fourier light-field microscopy. *Optics express*, 27(18):25573–25594, 2019.
- [13] Michael Broxton, Logan Grosenick, Samuel Yang, Noy Cohen, Aaron Andalman, Karl Deisseroth, and Marc Levoy. Wave optics theory and 3-d deconvolution for the light field microscope. *Optics express*, 21(21):25418–25439, 2013.
- [14] Ayan Chakrabarti. Learning sensor multiplexing design through back-propagation. *Advances in Neural Information Processing Systems*, 29, 2016.
- [15] Xiujuan Zou, Ruoyu Lin, Yunlai Fu, Guangxing Gong, Xuxi Zhou, Shuming Wang, Shining Zhu, and Zhenlin Wang. Advanced optical imaging based on metasurfaces. *Advanced Optical Materials*, page 2203149, 2023.
- [16] AG Poleshchuk. Fabrication and application of diffractive optical elements. In *Sixth International Symposium on Precision Engineering Measurements and Instrumentation*, volume 7544, pages 903–914. SPIE, 2010.
- [17] Rudolf Kingslake. *Optics in photography*, volume 6. SPIE press, 1992.
- [18] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company publishers, 2005.
- [19] D. Iso F. Yasuma, T. Mitsunaga and S.K. Nayar. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Technical report, Nov 2008.
- [20] A. Chakrabarti and T. Zickler. Statistics of Real-World Hyperspectral Images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 193–200, 2011.
- [21] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer, 2016.
- [22] Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2017)*, 36(6):218:1–13, 2017.
- [23] Joyce E Farrell, Feng Xiao, Peter Bert Catrysse, and Brian A Wandell. A simulation tool for evaluating digital camera image quality. In *Image Quality and System Performance*, volume 5294, pages 124–131. SPIE, 2003.
- [24] Olaf Ronneberger. U-net convolutional networks for biomedical image segmentation. *Arxiv*, page 3–3, May 2018.

- [25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18, pages 234–241. Springer, 2015.
- [26] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Matthias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
- [27] Nahian Siddique, Sidike Paheding, Colin P Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *Ieee Access*, 9:82031–82057, 2021.
- [28] Xiaofei Yang, Yunming Ye, Xutao Li, Raymond YK Lau, Xiaofeng Zhang, and Xiaohui Huang. Hyperspectral image classification with deep learning models. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9):5408–5423, 2018.
- [29] D. Brunet, E. R. Vrscay, and Z Wang. On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, 21(4):1488–1499, Oct 2011.
- [30] Leon B Lucy. An iterative technique for the rectification of observed distributions. *Astronomical Journal, Vol. 79, p. 745 (1974)*, 79:745, 1974.
- [31] William Hadley Richardson. Bayesian-based iterative method of image restoration. *JoSA*, 62(1):55–59, 1972.
- [32] Seung-Hwan Baek, Incheol Kim, Diego Gutierrez, and Min H Kim. Compact single-shot hyperspectral imaging using a prism. *ACM Transactions on Graphics (TOG)*, 36(6):1–12, 2017.
- [33] Diptodip Deb, Zhenfei Jiao, Ruth Sims, Alex Chen, Michael Broxton, Misha B Ahrens, Kaspar Podgorski, and Srinivas C Turaga. Fouriernets enable the design of highly non-local optical encoders for computational imaging. *Advances in Neural Information Processing Systems*, 35:25224–25236, 2022.

A Supplementary Material

A.1 Optical System

The optical system is a simple single lens design (Figure 1a). The aperture is F/4, with a doublet lens with a 100mm focal distance at 525nm. Doublet lens is comprised of a positive asphere of N-BK7 crown glass, with a spherical concave flint component using N-F2 glass. The system was designed in Ansys Zemax OpticStudio.

A.2 Structural Similarity Index Measure

SSIM achieves accurate image comparisons without allowing the overall structure of the image to change, so the network can't recreate an image that's completely different. This metric's goal is to accurately measure the perceptual quality of an image using a combination similarity measurements between each image's luminance, contrast, and structure. Luminance is described by a function of each image's the mean pixel value, contrast is described by a function of each image's standard deviation, and structure is a function of the centered and normalized pixel values. Something important to note is that the range of SSIM is $[-1, 1]$ with larger output values corresponding to greater similarity. Therefore, we formulate our "SSIM Loss" function as

$$\text{SSIM}_{\text{Loss}} = 1 - \text{SSIM} \quad (5)$$

which has a range of $[0, 2]$, with high similarity corresponding to lower values. However, this may cause the model to prioritize improving structural similarity to the point where it fails to restore spectral information. On the other hand, MSE can severely penalize any pixel differences, which may ensure that we don't lose as much information, specifically spectral information.

We experimented with a combination of both of these since our problem requires both image deblurring and spectral information recovering.

A.3 Richardson-Lucy Deconvolution Details

Originally, Richardson-Lucy (RL) deconvolution was an iterative procedure for two-dimensional image deblurring with a known PSF [30][31], and widely applied as a standard method towards image deblurring. The RL deconvolution scheme is as follows for a two-dimensional image and two-dimensional known PSF:

$$g^{k+1} = g^k \cdot \left(\frac{O}{\text{PSF} * g^k} * \text{PSF}' \right) \quad (6)$$

Where O is the captured blurred image, g^k is the estimate of the unblurred image at iteration k , PSF is the point-spread function of the system, and PSF' is the point spread function of the system rotated by 180 degrees.

In the case that RL deconvolution is used to project a two-dimensional image back into a 3-dimensional underlying image by means of a 3D PSF (ie, depth varying PSF), RL deconvolution follows a slightly modified originating from optical tomography [13][12], maintaining the structure of forward (into the blurred image space) and backwards (into the underlying space) projections, presented in Equation 4.

We coded a hyperspectral RL deconvolution algorithm. As shown in Supplementary Figure 9, our modified 3D RL deconvolution performs similarly to classical 2D RL deconvolution on a standard blurred 2D image. Because RL deconvolution does not have a set stopping criteria and may have differing iteration count needs depending on the type of images present, we applied our 3D RL deconvolution to restore 100 random samples of hyperspectral data and characterized the baseline using one of our evaluation metrics, MSE, with respect to the number of RL iterations. We determined that 15 was the best number of iterations, producing the lowest loss, and processed the entire hyperspectral dataset using RL deconvolution for 15 iterations.

The results are shown in Figure 8. Immediately we note that the RL deconvolution was unable to properly extract the spectrum of each pixel, as all wavelengths past 550nm had no intensity

contribution (Figure 8, Beads RL Profile 1), resulting in the picked wavelengths for the pseudo-RGB image to have low intensities and appearing dark. This is likely due to the similarity of the PSF in this range: as shown in Figure 1b, the PSF shows self-similarity across the range 420-550nm and 560-710nm range. This also results in the loss of some wavelength-specific textures, such as the background roughness in the 710nm slice of the Painting. However, the RL deconvolution is still able to perform mild deblurring from the monochrome image, as shown in the 560nm slice of the Painting as expected. As such, we note that the baseline method, RL-deconvolution, performs relatively poorly in the hyperspectral regime. This likely explains why previous approaches towards hyperspectral imaging instead relied on deep networks [10], and those that did utilize classical filtering approaches had highly structured spectra with only single peaks [8].

A.4 Additional Figures

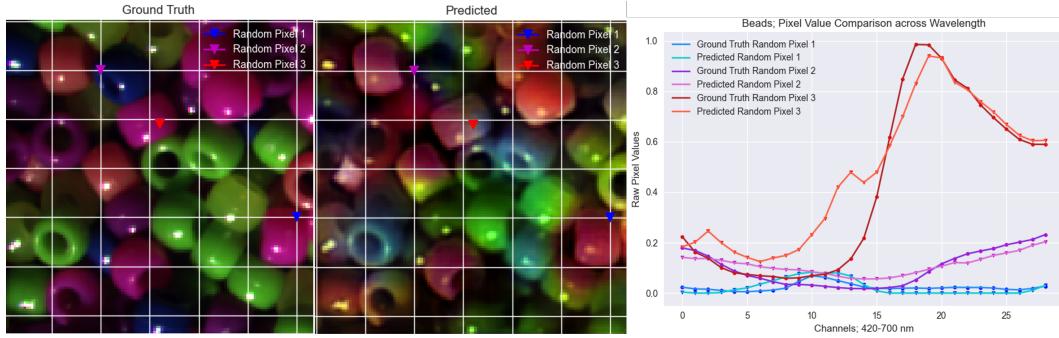


Figure 6: Pixel comparison of Pseudo-RGB images for HyperUnet_{LR}.

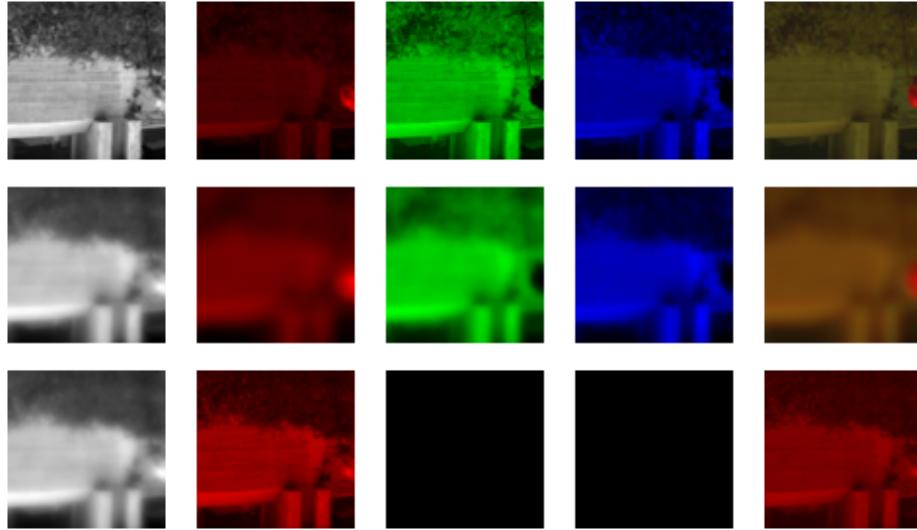


Figure 7: Comparison of Pseudo-RGB images between ground truth (top row), hyperspectral blurred images (middle row), and the HyperUnet_{SSIM} outputs.

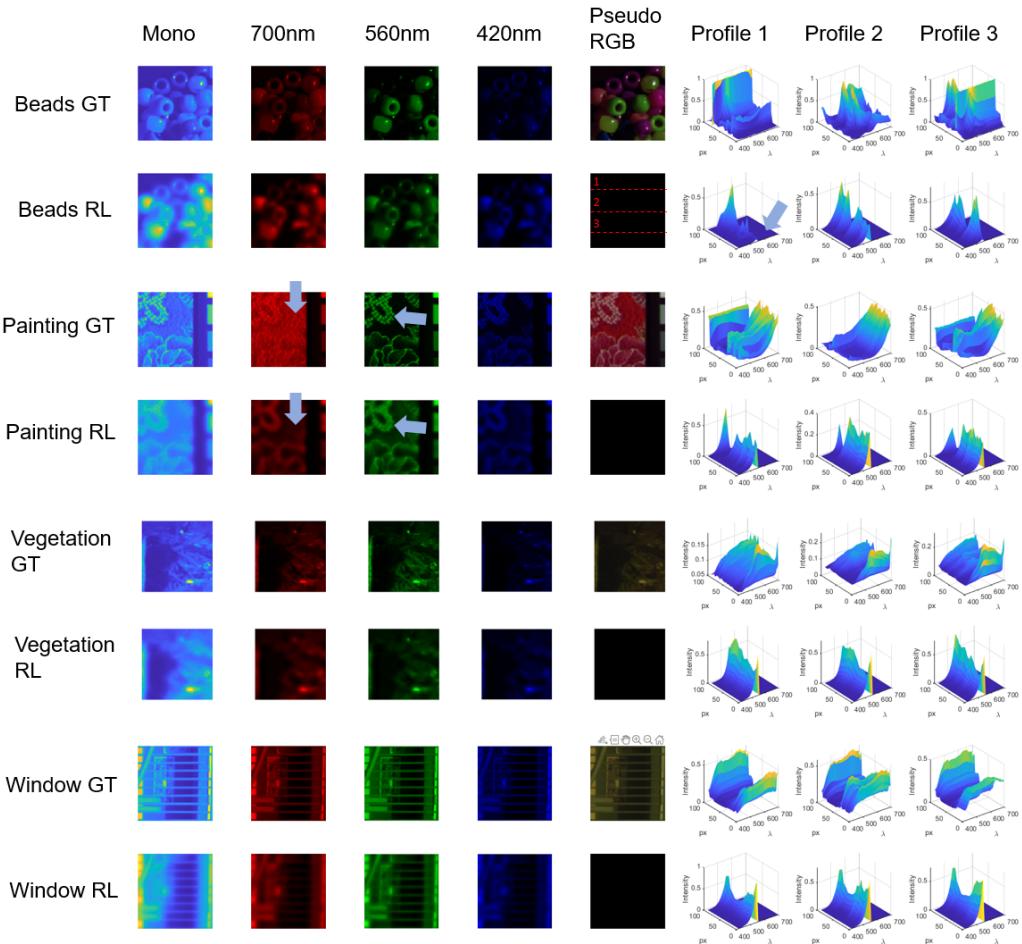


Figure 8: Pixel comparison between Ground Truth and Predicted test images for 3D RL.

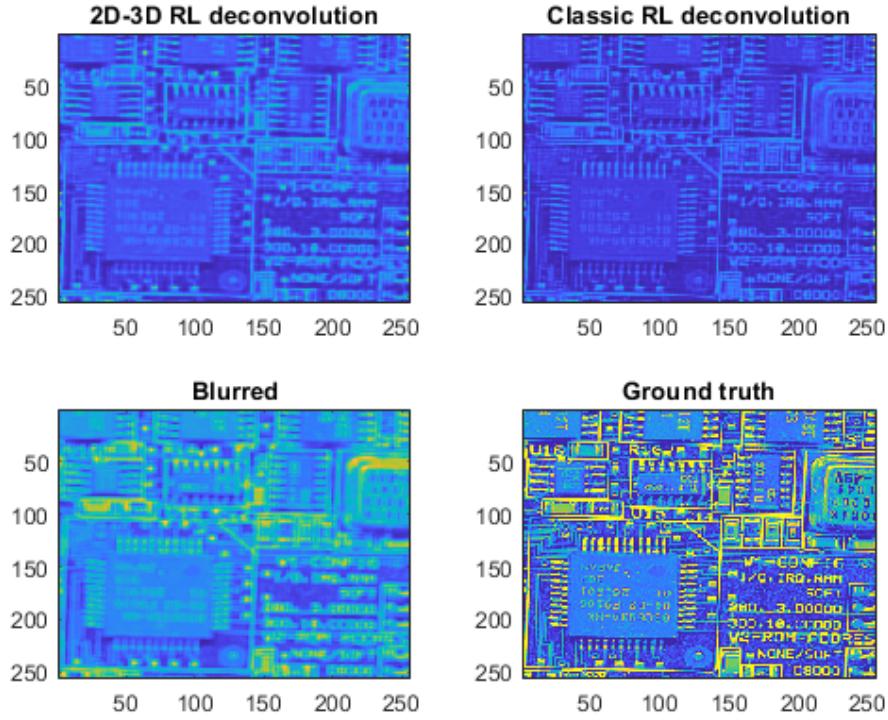


Figure 9: Comparison of 2D-3D and Classic RL deconvolution results. The images from top to bottom for each respective side is a comparison between ground truth and predicted test images.

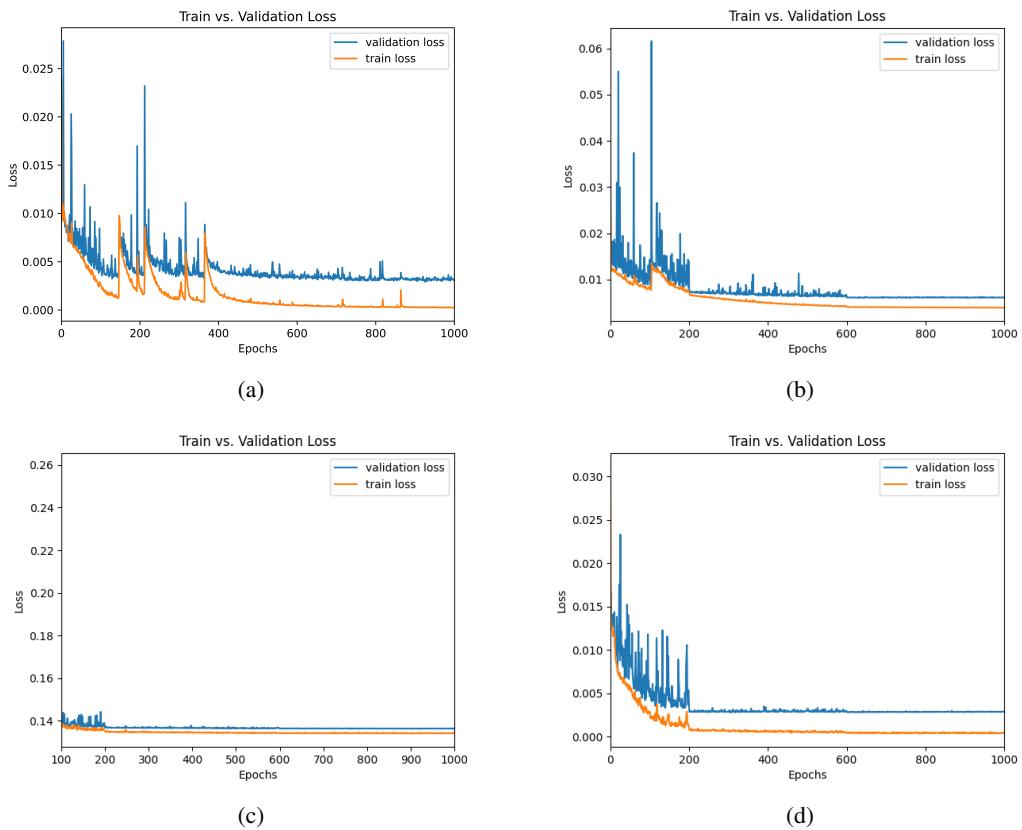


Figure 10: Loss curves for various experiments. **(a)** Loss curve for HyperUnet. **(b)** Loss curve of ReducedUnet Architecture. This was done as an exploration of the overfitting issue seen in previous models and also because of the limited time and computing resources available. **(c)** Loss curve for HyperUnet_{SSIM}. **(d)** Loss curve for HyperUnet_{LR}.

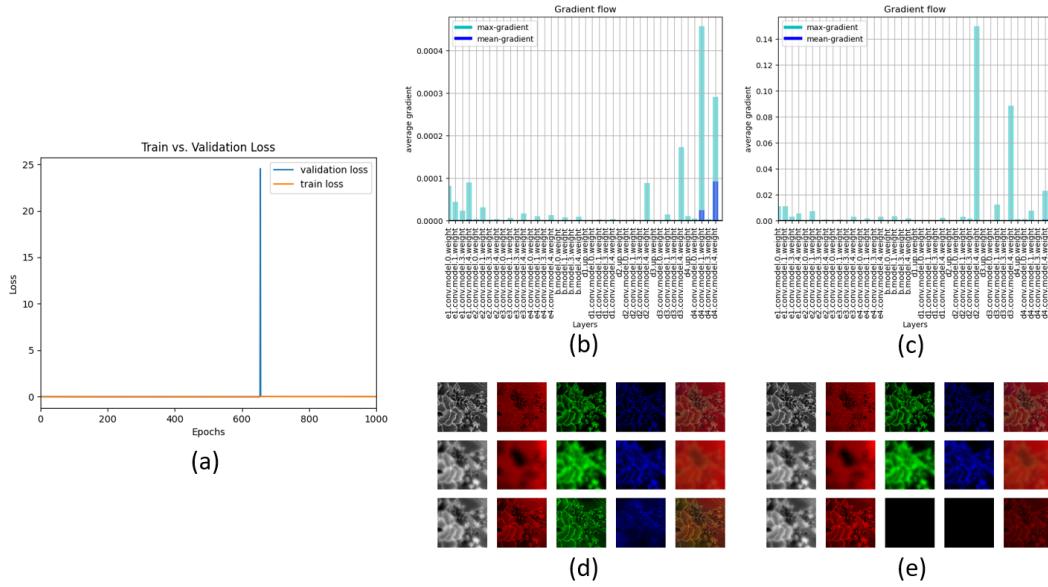


Figure 11: Example of training instability. (a) loss curve which presents an extreme spike at epoch 656. (b) Plot of max and average gradients for each layer at epoch 655. (c) Plot of max and average gradients for each layer at epoch 656, showing a 100-fold increase during a sudden gradient explosion. (d) Flowers test sample prior to gradient explosion. (e) Flowers test sample after gradient explosion, presenting loss of two color channels.

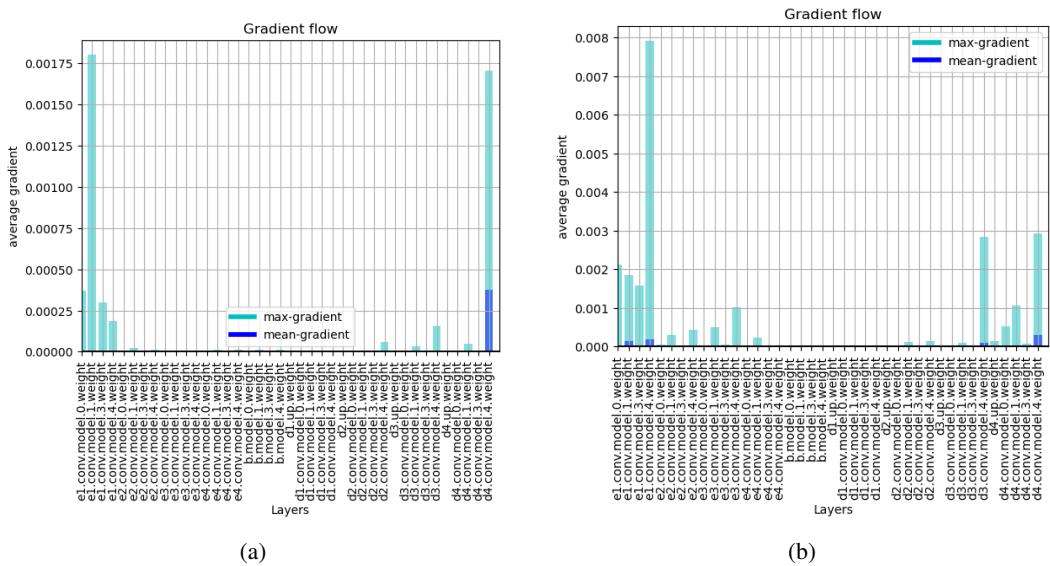


Figure 12: Gradient plot examples for gradient analysis. (a) Gradient size plots for HyperNet_{LR}. (b) Gradient size plots for HyperNet_{SSIM}.