# Unsupervised Single Image Hyperspectral Reconstruction

Divya Kothandaraman

Indian Institute of Technology Madras, India

**Abstract.** Reconstruction of the hyperspectral image from a compressively sensed image or an RGB image is a challenging task. The existing supervised regression methods have been shown to perform well in solving such ill-posed inverse problems as compared to traditional iterative methods, but their performance is data dependent and the reconstructed results exhibit inconsistent quality for different datasets. To overcome the drawbacks of supervised learning for hyperspectral reconstruction, we propose an unsupervised deep learning based pipeline that can reconstruct the hyperspectral image from a RGB image or compressively sensed (CASSI) data by solving an inverse optimization problem. We exploit the fact that the hyperspectral to RGB or CASSI image formation model is differentiable and can be used in a gradient-based learning algorithm. We utilize the imaging model to directly reconstruct the hyperspectral image while testing by optimizing in the neural network weight space. Our method is based on deep image prior and since it does not require a training step, it is robust to the problem of domain-shift. We show that even with a primitive network (UNet), our method competes well with much complex architectures that other supervised methods use.

**Keywords:** Hyperspectral Reconstruction · Single Image · Unsupervised

## 1 Introduction

Hyperspectral imaging systems acquire over-sampled electromagnetic signals from a narrow range of spectrum and hence provide highly precise information about the scene. They are widely used in many areas such as remote sensing, military, industry, etc and have many vision based applications such as image classification, segmentation, tracking, face recognition, document analysis, analysis of paintings, food inspection etc.

Hyperspectral(HS) imaging systems are either based on spectral scanning systems or on snapshot imaging. Scanning based systems like push broom and whisk broom scanners [4],[1] capture data using different band pass filters but are limited in spectral resolution. Snapshot HS imagers like camera-array systems tend to be very bulky and thus have limited applications. A compact snapshot HS imaging system is the Coded Aperture Snapshot Spectral Imaging(CASSI) [5],[10] which is based on compressive imaging. Such systems are expensive due to various optical elements required and hence are difficult to handle in practice. Thus, the cost of acquiring hyperspectral data is very high. However, reconstructing hyperspectral data from a RGB image is an affordable and convenient option.

The task of reconstructing hyperspectral images from RGB is highly ill posed. There is a significant information difference between a HS and the corresponding RGB image and this makes hyperspectral recovery an extremely ill-posed problem. Researchers have so far explored many data driven techniques to reconstruct the hyperspectral data-cube from a single RGB image. Supervised deep learning approaches require a huge amount of data to accurately model the problem. Additionally, they may not generalize well on images collected from diverse sources due to domain gap issues. Furthermore, hyperspectral ground truth is difficult to obtain. These issues call for an unsupervised domain independent approach for hyperspectral reconstruction.

In this paper, motivated by Deep Image Prior [9], we propose an unsupervised method that can be used to directly reconstruct the HS data from a single RGB image or from CASSI measurements by utilizing the hyperspectral to RGB/CASSI imaging model and optimizing in the neural network parameter space. We use an autoencoder to reconstruct the hyperspectral data-cube, followed by the application of a differentiable mapping to downsample the hyperspectral image back to the RGB / CASSI image. We then optimize the autoencoder's parameter space by minimizing a compound loss and this provides an optimal reconstructed HS image. Along with the mean squared error, our compound loss consists of an additional loss term for consistency between consecutive channels.

In case a hyperspectral dataset is available, we propose to first train our autoencoder network on the dataset. We then initialize the network with these weights for our single image reconstruction task. We observe that the initialization not only aids the training of images belonging to the dataset on which the supervised training was done, but also images from other datasets. We additionally show that using a primitive network such as UNet in our proposed method performs comparable with other carefully tailored and more complex architectures when used in supervised methods. We perform experiments on a few RGB images from different datasets and report comparisons with the current state-of-the-art methods. Though CASSI measurements are hard to obtain (the camera is not readily available), we simulate CASSI measurements and obtain the corresponding hyperspectral reconstructed data using our algorithm. The results prove that our algorithm not only works well on RGB images but also on CASSI measurements. In a nutshell, the contributions of this paper are as follows:

- We propose a training-free deep learning based approach based on Deep Image Prior [9] for reconstructing hyperspectral data from RGB or compressed measurements (CASSI). Apart from measurement loss, we also use channel-wise consistency loss for regularization.

- Our method does not need any training data. However, if such a dataset is available, then we can use it to pre-train our network. We show that initializing our network with the pre-trained weights leads to faster convergence.
- We show that even with a simple network, such as the UNet, we obtain results comparable to complex architectures, which require multi-GPU training.

## 2 Related work

Hyperspectral Image: RGB images contain three channels, red, blue and green at three distinct wavelengths which when combined seem realistic to the eye. Hyperspectral images [3] contain a lot more information that could be used for diverse applications. A hyperspectral camera divides the visible spectrum into thin slices, and thus hyperspectral images contain multiple 2D slices at different wavelengths. A hyperspectral image is generally represented as a cube containing two spatial dimensions(grayscale images at each wavelength representing reflectance data) and one spectral dimension (wavelength).

Deep Image Prior [9]: In many adversarial networks, it is observed that the structure of the generator network is capable of capturing low-level image statistics prior to any learning. Hence, a randomly-initialized neural network can be used as a handcrafted prior. For a network to be robust, the structure of the network needs to resonate with the structure of the data. Though this is contrary to the common belief that learning is required to build good image priors, this method has been shown to be useful in a number of image restoration problems like denoising and inpainting where the image prior is required to integrate information lost in the degradation processes. The striking aspect of this algorithm is that no aspect of the network is learned from the data and the weights are randomly initialized. Additionally, this is an unsupervised learning approach thus eliminating the need for ground truth. Training is not required. The network directly operates on the test image by optimizing in the neural network weight space. Typically, a small learning rate in the order of 0.001 is used.

Supervised Hyperspectral Reconstruction from RGB image: Supervised hyperspectral reconstruction is a problem that has been widely explored in the past. In HSCNN [11], the current state-of-the-art, the RGB image is rst upsampled in the spectral dimension through simple interpolation. The network then learns a mapping which is represented as a deep convolutional neural network (CNN). An improvement over this is HSCNN+ [8]. The plain convolution layers in HSCNN are replaced by residual blocks. The residual blocks in this model which is called HSCNN-R boosts the model's performance. Replacing the residual block by a dense block with a novel fusion scheme gives HSCNN-D. This makes the model much deeper thus allowing it to learn a much more complex function. However, increase in network depth increases the computational complexity. Additionally, these data driven methods are highly dependant on the metamerism of the inputs from a RGB camera. Also, one of the most important and commonly ignored aspect is the absence of any concrete justification for the optimality of RGB channels for reconstruction of HS image and not considering the possibility of any other channel as an optimal solution for this problem. Traditional methods to obtain hyperspectral reconstruction from RGB/CASSI images include TwIST [12].

CASSI: CASSI [5], [10] is a single-shot spectral imaging approach. It is based on the concept of compressive sensing. Light from the scene which we wish to capture is passed through a prism. The prism being a dispersive element splits light into various components on the basis of their wavelengths. Light is then passed though a binary coded aperture, which spatially blocks and admits light to pass through. Another prism is then placed beyond the coded aperture which unshears this filtered light. This produces a compressive hyperspectral image in which each pixel value is the integral over the wavelengths and spatial components of the data-cube. TwIST [12] is one traditional method to reconstruct hyperspectral images from CASSI data. Deepcassi [2] uses a spectral prior followed by ADMM optimization for the same task.

# 3 Unsupervised Single Image Hyperspectral Reconstruction

## 3.1 Data

We perform experiments on two hyperspectral datasets - BGU ICVL and CAVE. Hyperspectral data has information captured at wavelengths ranging from 400nm to 700 nm at 10 nm intervals. CASSI and RGB images are simulated from the provided hyperspectral data.

The input image is either a 3-channel RGB image / 1-channel downsampled CASSI image (grayscale). To downsample hyperspectral data to RGB images [3] , we use illumination data which represents the spectra of blue skylight with correlated colour temperature (CCT) 25000 K, daylight with CCT 6500 K, and evening sunlight with CCT 4000 K. To generate the CASSI image from its corresponding hyperspectral channels [6], a 2D binary(0s and 1s) random mask is initialized. The height of the random mask equals the height of the desired CASSI and hyperspectral image. The width equals the sum of the width of the CASSI image and the number of hyperspectral channels. A 3D mask is now generated. This 3D mask has the same spatial dimensions as the 2D mask. It has 31 channels(same as hyperspectral image). Channels of the 3D mask are shifted versions of the 2D mask. The 3D mask is then spatially cropped to the size of the hyperspectral image. Element wise multiplication of the 3D mask with the hyperspectral image followed by addition across channels gives us the corresponding 2D CASSI image.

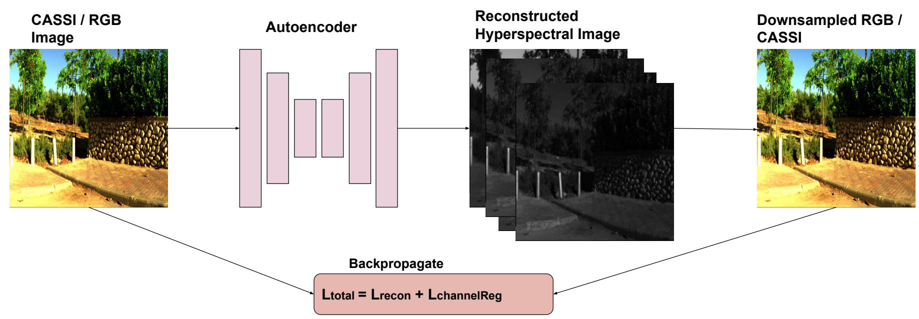


**Fig. 1.** Our Method: The RGB/CASSI image is passed through an autoencoder to obtain the reconstructed hyperspectral image. It is then downsampled by a suitable forward mapping followed by the application of losses.

### 3.2 Architecture and Training Details

The key advantage of our method is that it just needs just the test image. Typically, priors, training data, initialization, etc are not required though we initialize our network with pre-trained supervised weights for faster convergence. However, we notice that the final results don't vary much on when we initialize with pre-trained supervised weights as opposed to random initialzation. The key advantage is that training time is reduced to less than 1/10th of the time it takes to train when random initialization is done due to the decrease in the number of iterations that it takes to converge.

Our pipeline is depicted in Figure 1. The architecture of our autoencoder is a simple UNet. It has 3 input channels(3 for RGB, 1 for CASSI) and 31 output channels(hyperspectral). The UNet has 4 downsampling and 4 upsampling layers. We use a sigmoid layer at the end to constrain the output between 0 and 1. The UNet is initialized with the weights of the same UNet trained in a supervised manner on the BGU ICVL dataset. In each iteration, the original RGB image is passed through the UNet to generate a hyperspectral image. The output hyperspectral image is downsampled to its corresponding RGB / CASSI image by the use of a suitable forward function.The downsampled RGB image and the original RGB image are used to calculate the mean squared error $\mathrm{L}_{recon}$. We additionally add a channel consistency loss $\mathrm{L}_{channelReg}$ for continuity in the pixel values between consecutive channels. Training is done till convergence which takes about 1000 iterations. A very small decaying learning rate of the order of 0.005 is used. We observe that careful hyperparameter tuning is required to generate good results.

$$L_{total} = L_{recon} + L_{channelReg} \tag{1}$$

where $L_{recon}$ is the mean squared error between input RGB/ CASSI image and downsampled reconstructed hyperspectral image and $L_{channelReg} = \Sigma$(mean squared error between consecutive channels in the reconstructed hyperspectral image). It basically serves as the regularization for the difference between consecutive channels, we hence call it 'channel regularization loss'.

## 4 Results and Comparisons

For RGB to hyperspectral reconstruction, we perform comparison with the supervised hyperspectral reconstruction method, HSCNN [11]. For hyperspectral reconstruction from CASSI images, we perform comparisons with deepcassi [2]. We demonstrate visual results, spectral plots (in the supplementary section) and quantitative results on a few images from 2 different datasets (more results in supplementary section). Quantitative results are reported in terms of Signal to Noise Ratio (in decibels). Min-max normalization is performed on all data before analysis.

### 4.1 RGB to Hyperspectral on BGU ICVL dataset

In this section, we demonstrate the results that our method achieves on the BGU ICVL dataset. The input RGB images are of size 1024 x 1024 x 3. The output images have a resolution of 1024 x 1024 x 31. We plot only 5 channels for space constraints. These channels correspond to the wavelengths 430nm, 490nm, 550nm, 610nm, 670nm. Our single image unsupervised method performs comparably well with the current state-of-the-art supervised method HSCNN++.

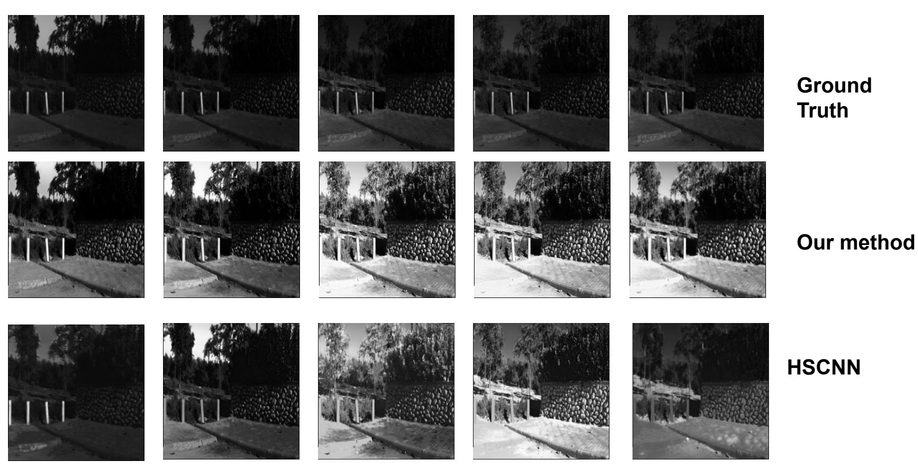


**Fig. 2.** RGB to hyperspectral, Image 1, BGU ICVL dataset, SNR: 26.77

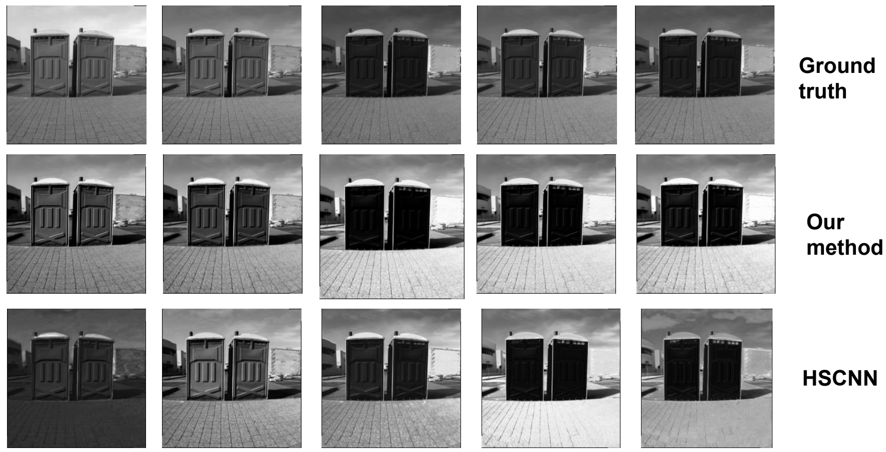


**Fig. 3.** RGB to hyperspectral, Image 2, BGU ICVL dataset, SNR: 35.22

### 4.2 RGB to Hyperspectral on CAVE dataset

To demonstrate the robustness of our algorithm across dataset, we show the performance of our pipeline on another dataset, the CAVE dataset.

### 4.3 CASSI to Hyperspectral on BGU ICVL Dataset

As cameras that capture CASSI data are not readily available, such data is difficult to capture. Using hyperspectral data, we simulate CASSI measurements. We then reconstruct hyperspectral images from these simulated CASSI measurements using our pipeline. We perform comparisons with deepcassi [2]. The results are displayed below.


**Fig. 4.** Comparison between HSCNN++, ground truth (input RGB) and our method for RGB images downsampled from reconstructed hyperspectral images; Our method produces better results.

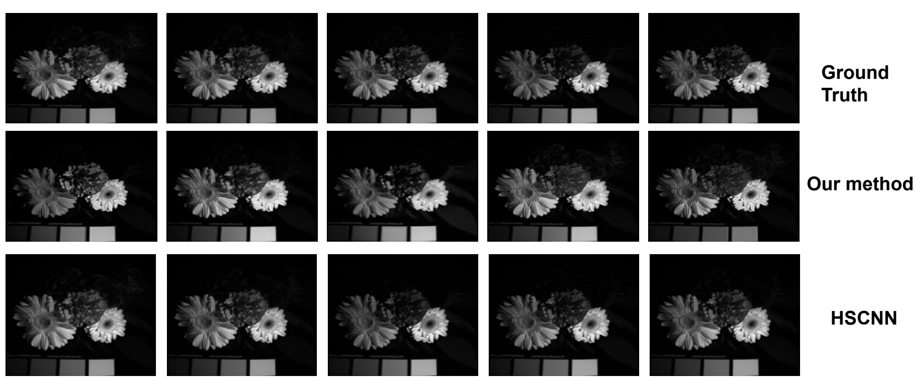


**Fig. 5.** RGB to hyperspectral, flowers, CAVE dataset, SNR:26.02

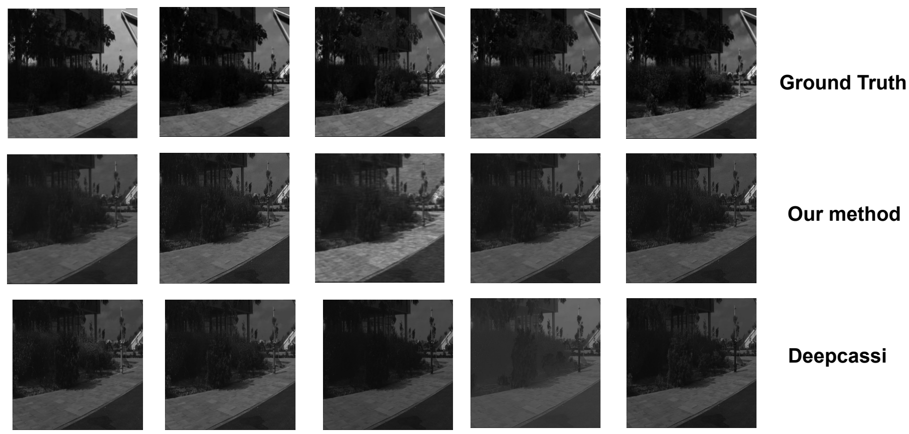


**Fig. 6.** CASSI to hyperspectral, Image 3, BGU ICVL dataset; Our method's results are less hazy in comparison to deepcassi's results. SNR:18.6

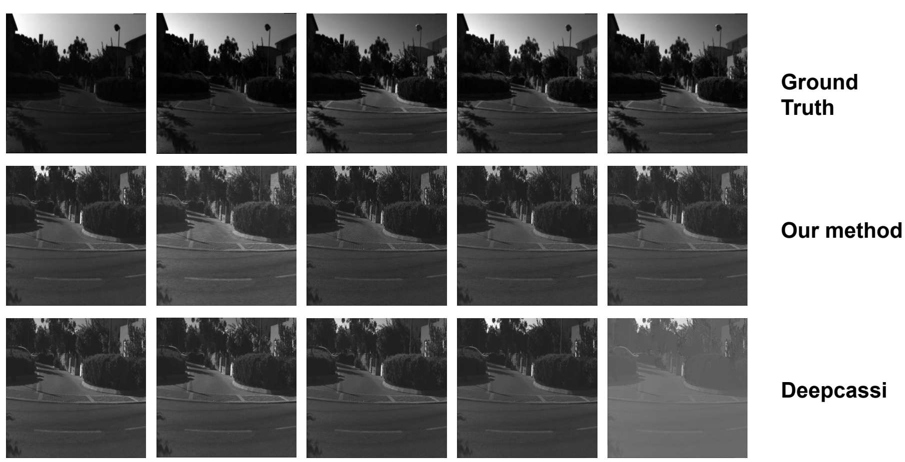


**Fig. 7.** CASSI to hyperspectral, Image 4, BGU ICVL dataset; Our method's results are less hazy in comparison to deepcassi's results.SNR:18.35

## 5 Conclusion

In this paper, we propose a method for single image unsupervised hyperspectral reconstruction from RGB and CASSI images. Apart from mean squared loss which is traditionally used, we propose the use of an additional loss term, the channel consistency loss to preserve continuity between successive channels. We additionally initialize our network with weights trained in a supervised manner to speed up training. We demonstrate results comparable to the state-of-the-art using a simple network architecture, the UNet, as opposed to more complex and deep networks that many other supervised methods use.

## References


1. N. Brusco, S. Capeleto, M. Fedel, A. Paviotti, L. Poletto, G. M. Cortelazzo, and G. Tondello. A system for 3d modeling frescoed historical buildings with multispectral texture information. Machine Vision and Applications, 17(6):373 393, 2006.
2. 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 (TOG), 36(6):218, 2017.
3. D. H. Foster and K. Amano. Hyperspectral imaging in color vision research: tutorial. JOSA A, 36(4):606627, 2019.
4. N.Gat. Imaging spectroscopy using tunable filters: a review. In Wavelet Applications VII, volume 4056, pages 5064. International Society for Optics and Photonics, 2000.
5. M. Gehm, R. John, D. Brady, R. Willett, and T. Schulz. Single-shot compressive spectral imaging with a dual disperser architecture. Optics express,15(21):1401314027, 2007.
6. M. Nishimura. Compressive Hyperspectral Image Reconstruction with a Deep Generative Prior. http://stanford.edu/class/ee367/Winter2018/nishimuraee367win18report.pdf.
7. Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 939947, 2018.
8. Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 939947, 2018.
9. D.Ulyanov,A.Vedaldi,andV.Lempitsky. Deep image prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 94469454, 2018.
10. A. Wagadarikar, R. John, R. Willett, and D. Brady. Single disperser design for coded aperture snapshot spectral imaging. Applied optics, 47(10):B44B51, 2008.
11. Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, and F. Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In Proceedings of the IEEE International Conference on Computer Vision, pages 518525, 2017.
12. J. Bioucas-Dias and M. Figueiredo. A new twist: Two step iterative shrinkage/thresholding algorithms for image restoration. IEEE Trans. Image Process., 16(12):2992 3004, 2007.



**Acknowledgements** We would like to thank Prof. Kaushik Mitra for his guidance and support. We would also like to thank Honey Gupta for her inputs.

## 6 Supplementary Material

We demonstrate results and comparisons on a few more images here.

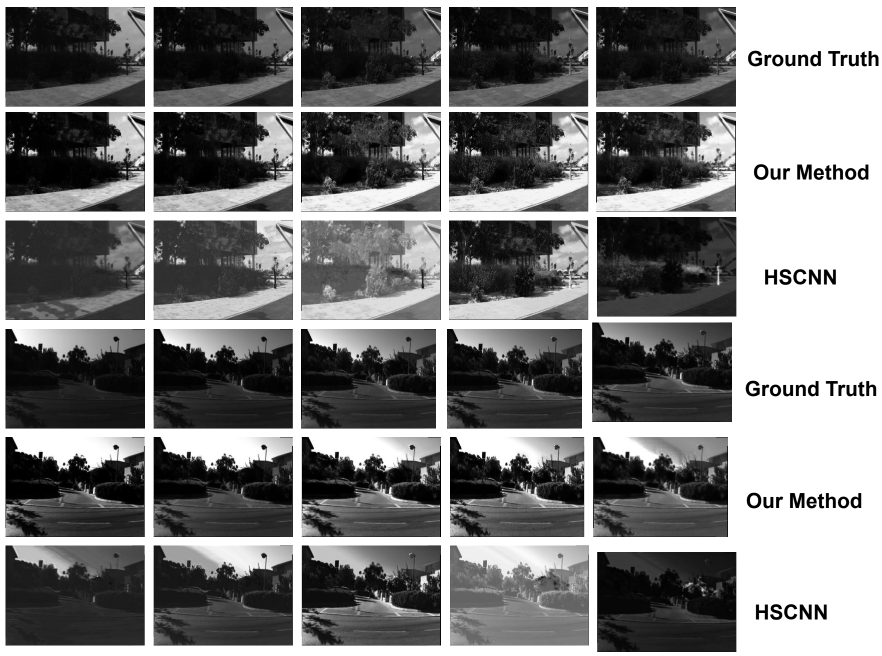


**Fig. 8.** RGB to hyperspectral, Image 3 and 4, BGU dataset,SNR:30.45 and 23.97 respectively

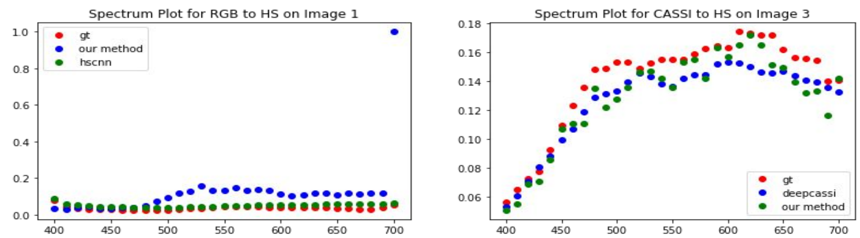


**Fig. 9.** Spectral Plots for images from BGU ICVL dataset

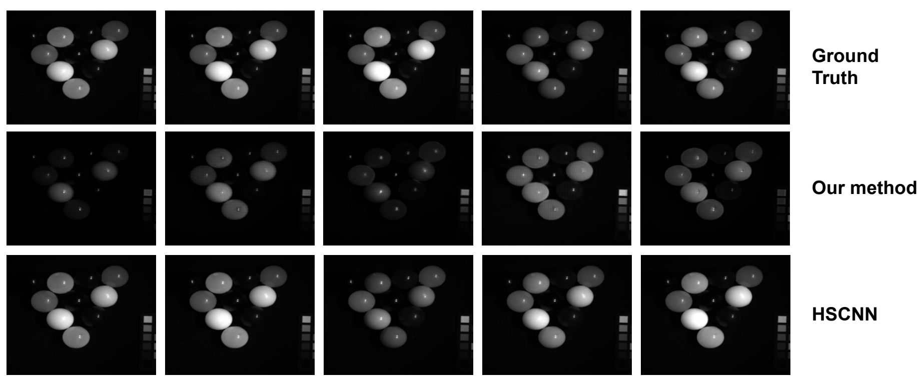


**Fig. 10.** RGB to hyperspectral, superballs, CAVE dataset,SNR:28.86

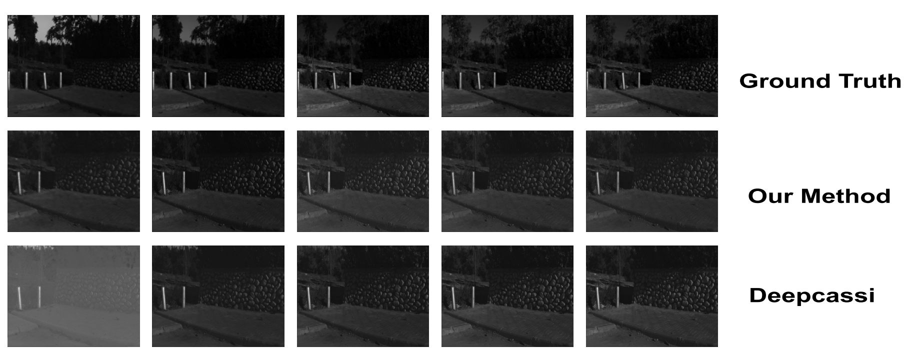


**Fig. 11.** CASSI to hyperspectral, Image 1, BGU dataset,SNR:16.4