Spectral Reconstruction from RGB Images

A BTP Report

by

Names

Praveen Pallagani

Vishruth Thippabathuni

Harsha Vardhan Sajja

Dinesh Kumar Koppala

Roll Nos.:

S20180010127
S20180010175
S20180010151
S20180010085

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRICITY

DATE: 16-12-2021

2nd Semester Report

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRICITY

CANDIDATE’S DECLARATION

I hereby certify that the work which is being presented in the BTP entitled “**Spectral Reconstruction from RGB Images**” in the partial fulfillment of the requirements for the award of the degree of B. Tech and submitted in the Indian Institute of Information Technology Sri City, is an authentic record of my own work carried out during the time period from January 2021 to May 2021 under the supervision of Prof. Shiv Ram Dubey, Indian Institute of Information Technology Sri City, India.

The matter presented in this report has not been submitted by me for the award of any other degree of this or any other institute.

Praveen Pallagni
16-12-21

Signature of the student with date

(Praveen Pallagani)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of BTP Supervisor with date

(Prof. P Viswanath)

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRICITY

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the BTP entitled "**Spectral Reconstruction from RGB Images**" in the partial fulfillment of the requirements for the award of the degree of B. Tech and submitted in the Indian Institute of Information Technology Sri City, is an authentic record of my own work carried out during the time period from January 2021 to May 2021 under the supervision of Prof. Shiv Ram Dubey, Indian Institute of Information Technology Sri City, India.

The matter presented in this report has not been submitted by me for the award of any other degree of this or any other institute.

Thippabathuni Vishruth
16-12-2021

Signature of the student with date

(Vishruth Thippabathuni)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of BTP Supervisor with date

(Prof. P Viswanath)

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRICITY

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the BTP entitled "**Spectral Reconstruction from RGB Images**" in the partial fulfillment of the requirements for the award of the degree of B. Tech and submitted in the Indian Institute of Information Technology Sri City, is an authentic record of my own work carried out during the time period from January 2021 to May 2021 under the supervision of Prof. Shiv Ram Dubey, Indian Institute of Information Technology Sri City, India.

The matter presented in this report has not been submitted by me for the award of any other degree of this or any other institute.

S. Harsha Vardhan
16-12-2021

Signature of the student with date

(Harsha Vardhan Sajja)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of BTP Supervisor with date

(Prof. P Viswanath)

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRICITY

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the BTP entitled "**Spectral Reconstruction from RGB Images**" in the partial fulfillment of the requirements for the award of the degree of B. Tech and submitted in the Indian Institute of Information Technology Sri City, is an authentic record of my own work carried out during the time period from January 2021 to May 2021 under the supervision of Prof. Shiv Ram Dubey, Indian Institute of Information Technology Sri City, India.

The matter presented in this report has not been submitted by me for the award of any other degree of this or any other institute.

Koppala Dinesh Kumar
16-12-2021

Signature of the student with date

(Dinesh Kumar Koppala)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of BTP Supervisor with date

(Prof. P Viswanath)

ABSTRACT

Idea of the project is to reconstruct the spectral images from the RGB images. RGB images are taken in the near wavelength ranges of 380nm - 750nm. In this project we try to reconstruct 31 spectral images at distinct wavelengths between 400nm to 700nm like 410nm, 420nm etc.

Capturing spectral images is hard, both money and time consuming, everybody can't take spectral images, so we are looking at a cost efficient and easy way of generating spectral images. We used the GAN based model for generation of spectral images from the RGB images.

We looked at different approaches using GAN's for the efficient way to reconstruct the spectral images back from the RGB images. We are going discuss about those models and datasets used in this paper.

Contents

1. Introduction

RGB images are taken from the 380 nm - 750 nm wavelength range of visible spectrum of light.
Here we want to reconstruct 31 spectral images in a range of 380 nm - 750 nm wavelength.

Bands we are focusing on are:
[400 410 420 430 440 450 460 470 480 490 500 510 520 530 540 550 560 570 580 590 600 610 620 630 640 650 660 670 680 690 700]

Normal RGB images have only 3 channels (red, green, blue channels). Spectral images we are trying to construct will have 31 channels each belonging to one band mentioned above. Spectral images are very useful in domains like image analytics, medical examinations, space research projects etc.

They capture more than just the physical features of the object. Fruits freshness detection, oil content in water, seeds viability for generating are some of the examples of applications of spectral images.

2. Literature Survey

GANs are a clever way of training a generative model by framing the problem as a supervised learning problem with two sub-models: the generator model that we train to generate new examples, and the discriminator model that tries to classify examples as either real (from the domain) or fake (generated). The two models are trained together in a zero-sum game, adversarial, until the discriminator model is fooled about half the time, meaning the generator model is generating plausible examples.

GANs are an exciting and rapidly changing field, delivering on the promise of generative models in their ability to generate realistic examples across a range of problem domains, most notably in image-to-image translation tasks such as translating photos of summer to winter or day to night, and in generating photorealistic photos of objects, scenes, and people that even humans cannot tell are fake.

Conditional generative adversarial network, or cGAN for short, is a type of GAN that involves the conditional generation of images by a generator model.

In cGANs, a conditional setting is applied, meaning that both the generator and discriminator are conditioned on some sort of auxiliary information (such as class labels or data) from other modalities. As a result, the ideal model can learn multi-modal mapping from inputs to outputs by being fed with different contextual information.

In a uni-modal experiment, Mirza and Osindero trained a cGAN on 784 dimensional MNIST images conditioned on their class labels. This generated results that were comparable with some other methods but were outperformed by non-conditional GANs.

Another experiment demonstrated automated image tagging using cGANs to generate (possibly multi-modal) distributions of tag-vectors conditional on image features. This showed promise and attracted further exploration of possible uses for cGANs.

3. Datasets Details

NTIRE 2020 SPECTRAL RECONSTRUCTION DATASET

- It contains 2 different tracks track1 and track2. Clean track and Real-World Track for RGB images. Corresponding Spectral Images are there for each RGB images.
- It consists of 31 spectral channels starting from 400nm to 700nm. With step size of 10nm between each channel like 400nm, 410nm, etc.
- It contains 450 training examples, 10 validation examples, 10 test examples.
- The Specim IQ camera provides RAW 512× 512px images with 204 spectral bands in the 400-1000nm range. For this challenge, manufacturer-supplied radiometric calibration has been applied to the RAW images, and the images have been resampled to 31 spectral bands in the visual range (400-700nm).
- Both RAW and radiometrically calibrated images have been made available to researchers. The radiometric calibration corrects for measurement biases introduced by the camera systems CMOS sensor, con- verting the recorded RAW per channel intensity data to accurate spectral measurements. “Lines” (image columns) with excessive interference are also removed by this process, resulting in a 482 × 512px image, resampled to 31 bands from 400nm to 700nm with a 10nm step.

CAVE DATASET

- The database consists of 32 scenes, divided into 5 sections. Spectral resolution reflectance data from 400nm to 700nm at 10nm steps (31 bands total). Each band is stored as a 16-bit grayscale PNG image.
- Each scene also contains a single representative color image, displayed using sRGB values rendered under a neutral daylight illuminant (D65).
- In total there are 32 different examples for all 31 channels.

NUS DATASET

- It contains spectral Images with 31 spectral channels ranging from 400nm to 700nm like other 2 datasets mentioned.
- Total 66 spectral Images examples are given.

4. Methodology

DEEP CONV NET MODEL

We used the Pix2Pix GAN model for the generation of these spectral images. This model takes input the RGB image and outputs 3 images in different spectral channels.

This Pix2Pix GAN is an adversarial network where internally we have 2 models Generator and Discriminator.

For the Generator we used the U-Net architecture, it involves down sampling of 3x256x256 images to bottleneck layer (512x8x8) from there again up sampling will be done to retain 3x256x256.

All the down sampling layers are as follows: Convolution, Batch Normalization and LeakyReLU with alpha (0.2) except the initial layer, it doesn't have Batch Normalization.

Bottleneck layers and up sampling first layers have dropout with a ratio of (0.5). All the up-sampling layers are as follows: Convolution, Batch Normalization and ReLU activation except the bottleneck layer it won't contain Batch Normalization.

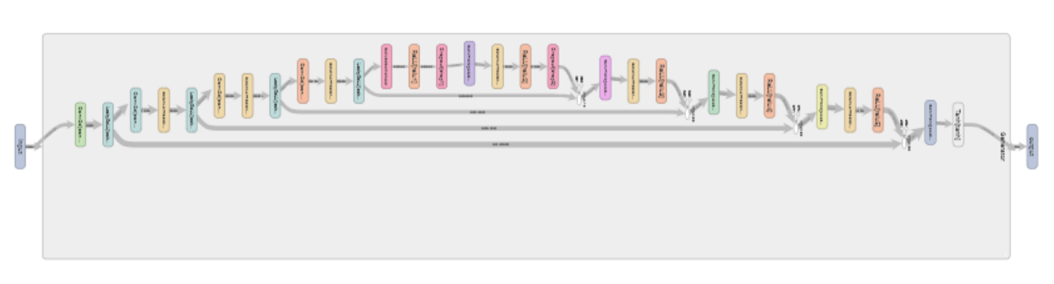


Figure 1.1 UNet Generator Network Figure.

Resolution Information of image after each layer of the network

256x256x3 --D-→ 128x128x64 --D-→ 64x64x128 --D-→ 32x32x256.

32x32x256 --D-→ 16x16x512 --D-→ 8x8x512 --U-→ 16x16x512 --U-→ 32x32x256.

32x32x256 --U-→ 64x64x128 --U-→ 128x128x64 --U-→ 256x256x3.

Skip connections: `[up1, d4], [up2, d3], [up3, d2], [up4, d1].`

HYPER UNET

This model is also very much like DEEP CONV NET, but the penalty constant lambda is hyper optimized using Optune library. Hyper parameter was selected from the range of 50 to 500 under uniform conditions by conducting 10 study cases.

In each study case, so far best performed value will be stored and returned to the model. This Hyper parameter optimization requires build of a objective function to minimize.

Here our objective function is to minimize the RMSE value of the overall model. So, each time parameter will be selected in such a way to decrease the RMSE value of the model.

RESNET

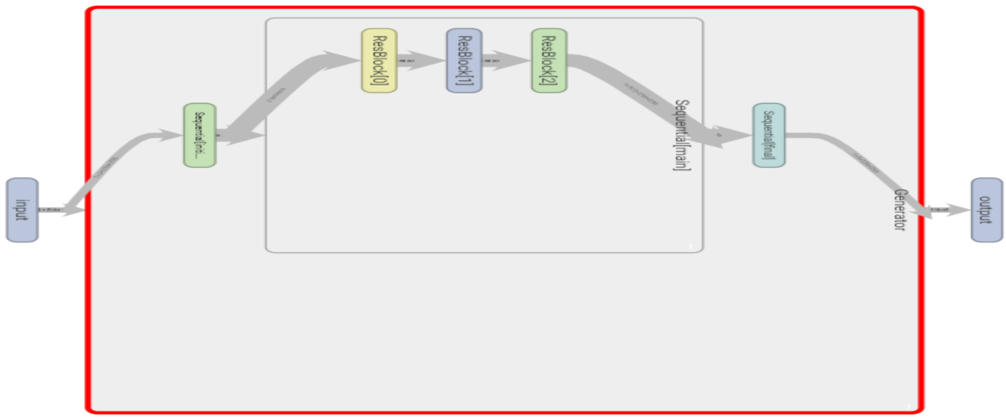


Figure 1.2 Resnet Generator Network Figure

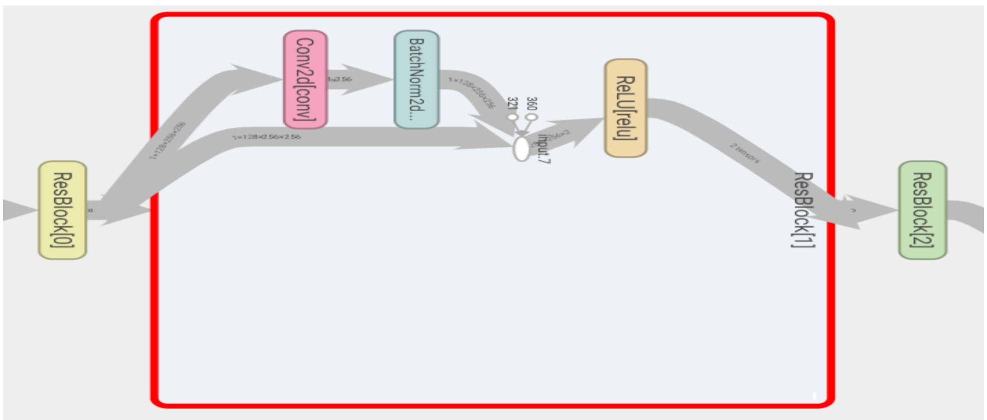


Figure 1.3 Resnet block internal design.

This is a Pix2Pix GAN model in which Resnet is used as Generator. It consists of Initial layer, 3 residual blocks and final layer.

Each Resnet block contains 1 conv layer, batch normalization and then Relu activation function. Input will be concatenated channel wise after doing the convolution and batch normalization.

Resolution Information of image after each layer of the network

256x256x3 → 256x256x64 → 256x256x128 → 256x256x256 → 256x256x512 → 256x256x3

Keeping the resolution same we changed the depth of the image in each layer.

DISCRIMINATOR

For the Discriminator network we used the Patch GAN.

The input for the network is channel wise concatenation of input and output of the generator.

Work of the Discriminator is to classify whether the generated output of the Generator is fake or real.

From the given input after applying some down sampling, it will output a 30x30x1 patch with probability of the patch being real.

For the real image our target is to get a patch of 1's and for fake images our target is to get a patch of 0's.

Each cell in the patch corresponds to a 70x70 patch in input.

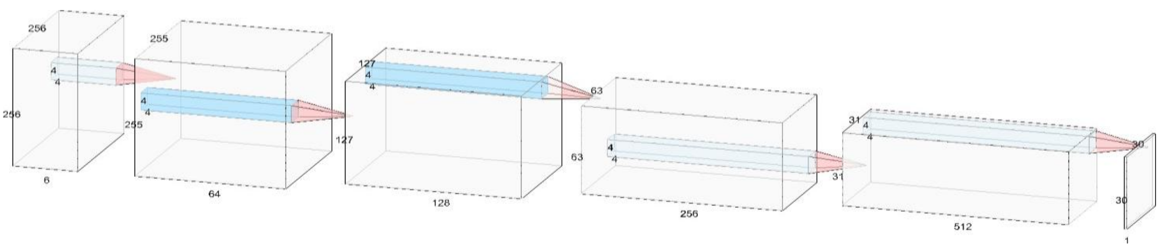


Figure 1.4 Discriminator Network Figure.

Resolution of images after each layer in the network.

256x256x6 → 255x255x64 → 127x127x128 → 63x63x256 → 31x31x512 --D→ 30x30x1.

5. Experimental Setup

Loss functions and optimizers

For the generator we used the combination of conditional adversarial loss and L1 loss together, and for the discriminator we used the binary cross entropy loss to classify the patch real or fake.

Generator loss: $\square^{*} = \arg\min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G)$

Discriminator Loss: $(Y)(-\log(Y_{pred})) + (1-Y)(-\log(1-Y_{pred}))$

Adam optimizers with learning rate of 2e-4 and beta values of (0.5,0.999).

Hyper optimized value of lambda and loss values of Generator and Discriminator in some region of wavelengths.

1. In range of 400nm – 420nm, lambda 103.97, Gloss 4854.71, Dloss 92.06.
2. In range of 520nm – 540nm, lambda 431.63, Gloss 7619.67, Dloss 17.02.
3. In range of 610nm – 630nm, lambda 256.45, Gloss 6578.48, Dloss 71.51.
4. In range of 680nm – 70nm, lambda 487.33, Gloss 4321.22, Dloss 66.01.

Evaluation Metrics of the Models.

RMSE (Root mean square error).
It involves difference between the predicted pixel value to the ground truth pixel value. **$|P_{gt}|$** is total number of pixels in the image.

$$RMSE = \sqrt{\frac{\sum_{i,c}\left(P_{gt_{i_c}} - P_{rec_{i_c}}\right)^2}{|P_{gt}|}},$$

Figure 2.1 RMSE Formula

RRMSE (relative RMSE)
This indicator is calculated by dividing RMSE with average value of measured data**.** Relative RMSE by dividing the error by the ground truth pixels value, thus preventing a bias towards low errors in lower pixel values.

$$rRMSE = \frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{(I_E^{(i)} - I_G^{(i)})^2}/I_G^{(i)}\right)$$

Figure 2.2 rRMSE Formula

6. Results

Predicted Spectral Images with RGB image and corresponding 3 spectral channels.

Figure 3.1 Result 430-450nm

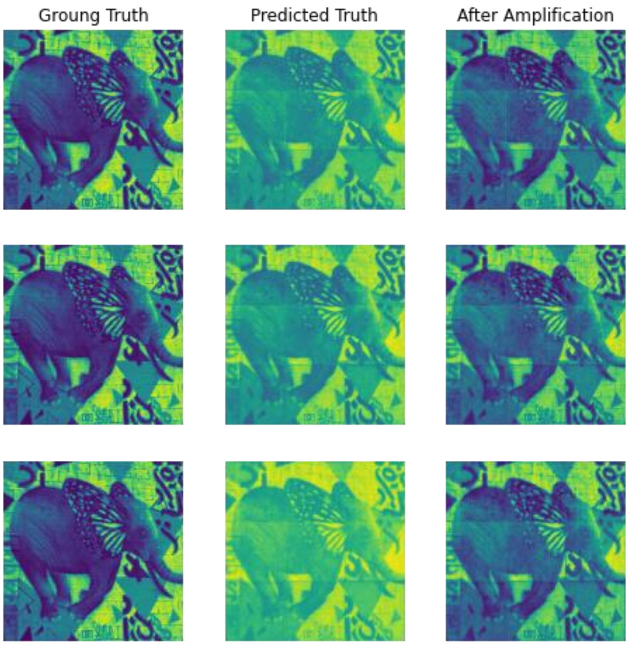


Figure 3.2 Result 520-540nm

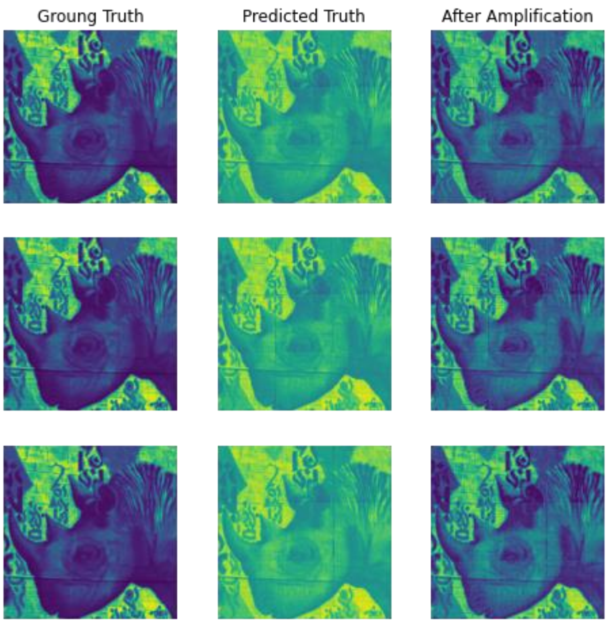


Figure 3.3 Result 610-630nm

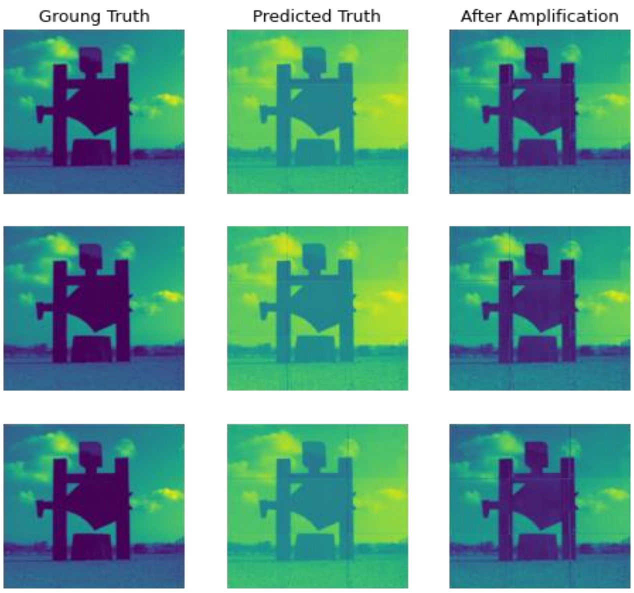
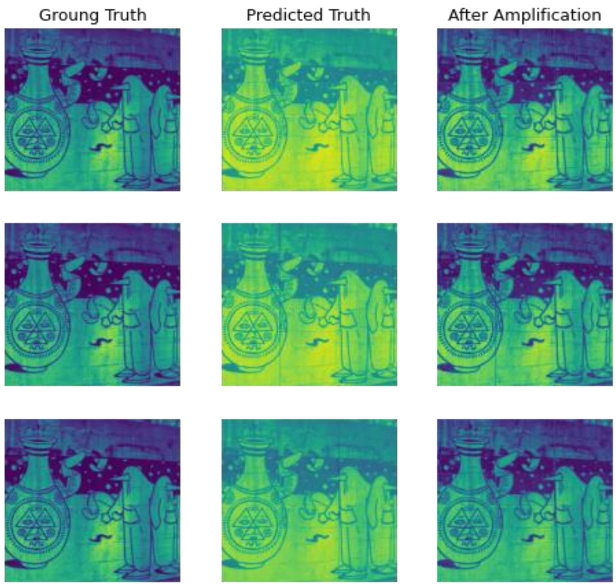


Figure 3.4 Result 680-700nm



Loss curves of model Generator and Discriminator

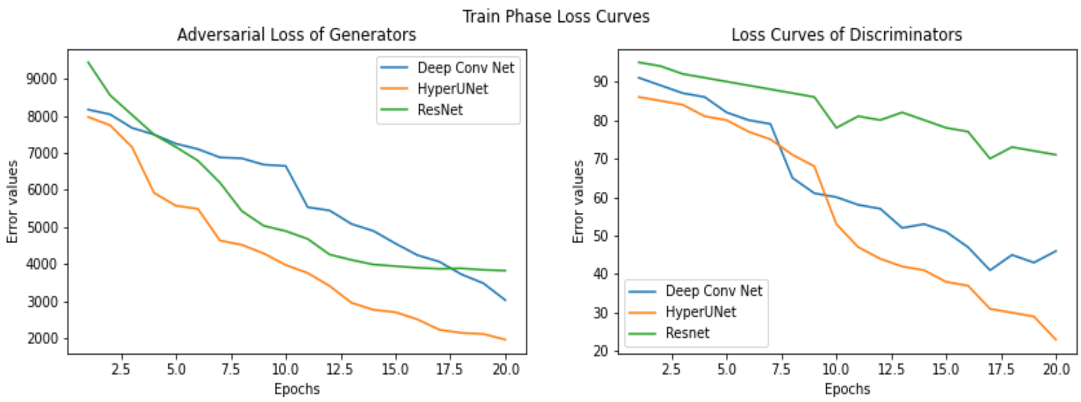


Figure 4.1 Loss curves of all Generators and Discriminators.

Channels wise loss curves of generator and discriminator.

Epochs on x-axis and score on y-axis.

Channels: 400nm, 410nm, 420nm.

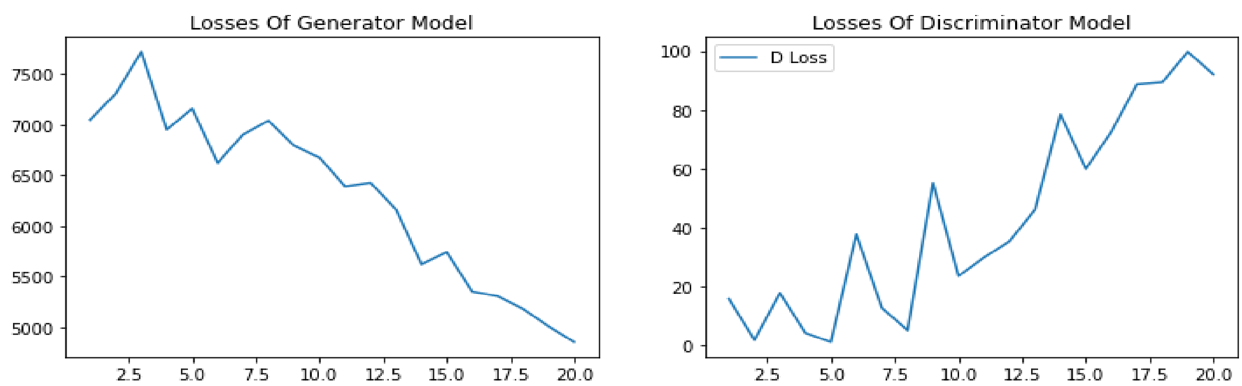


Figure 4.2 Loss curves for 400nm, 410nm, 420nm channels.

Channels: 430nm, 440nm, 450nm.

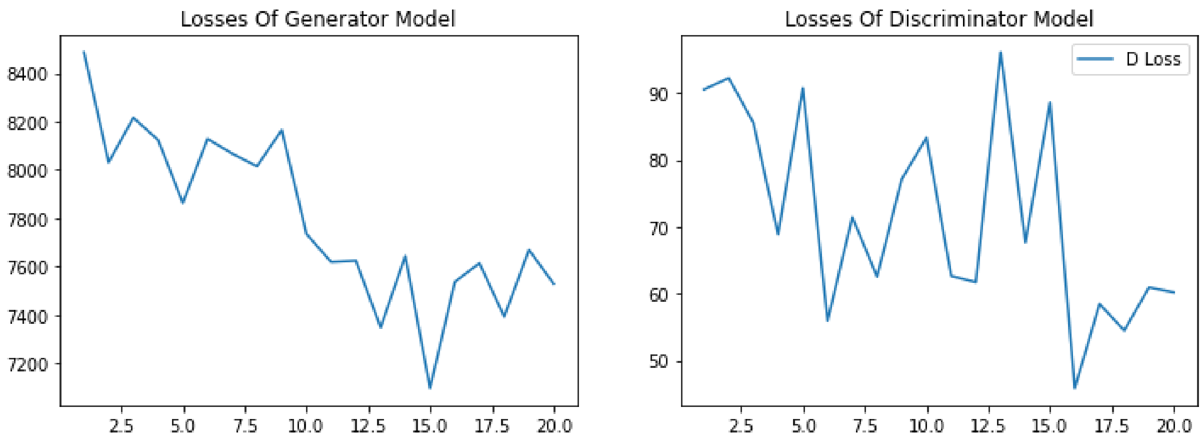


Figure 4.3 Loss curves for 430nm, 440nm, 450nm.

Channels: 460nm, 470nm, 480nm.

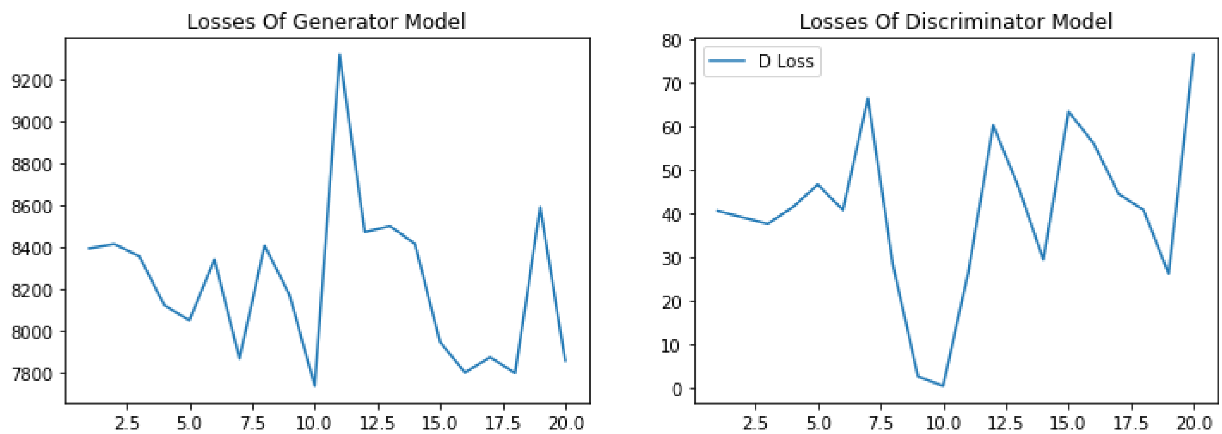


Figure 4.4 Loss curves for 460nm, 470nm, 480nm.

Channels: 490nm, 500nm, 510nm.

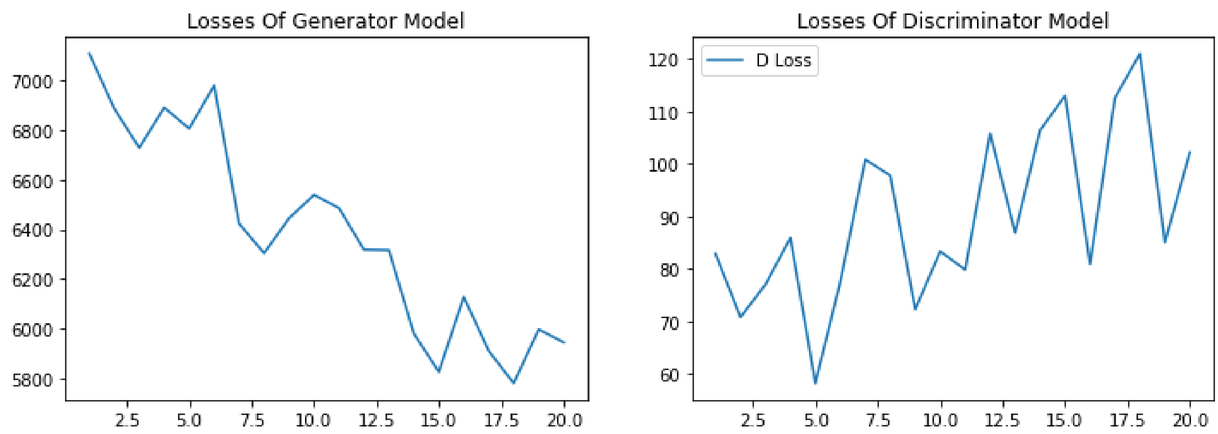


Figure 4.5 Loss curves for 490nm, 500nm, 510nm.

Channels: 520nm, 530nm, 540nm.

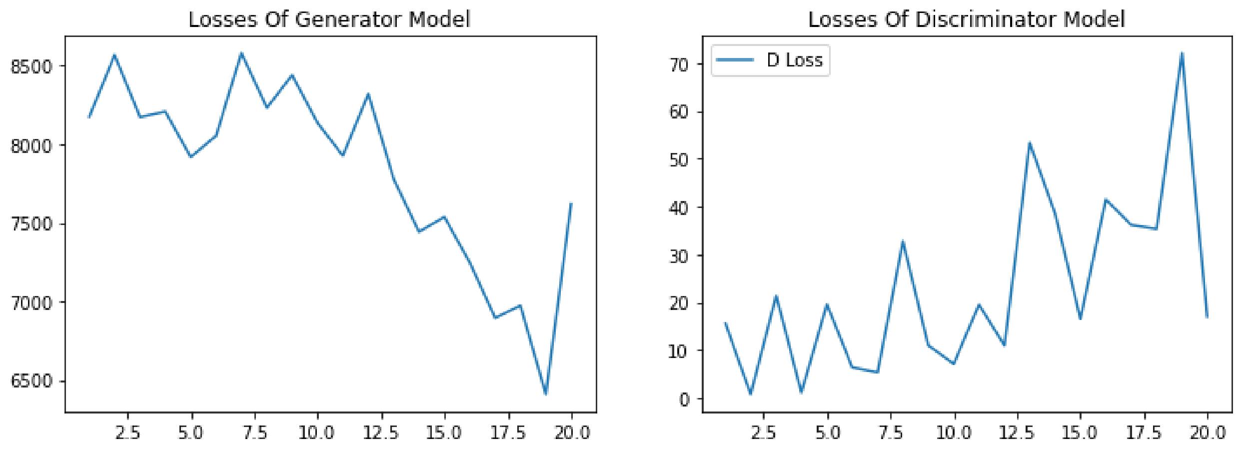


Figure 4.6 Loss curves for 520nm, 530nm, 540nm.

Channels: 550nm, 560nm, 570nm.

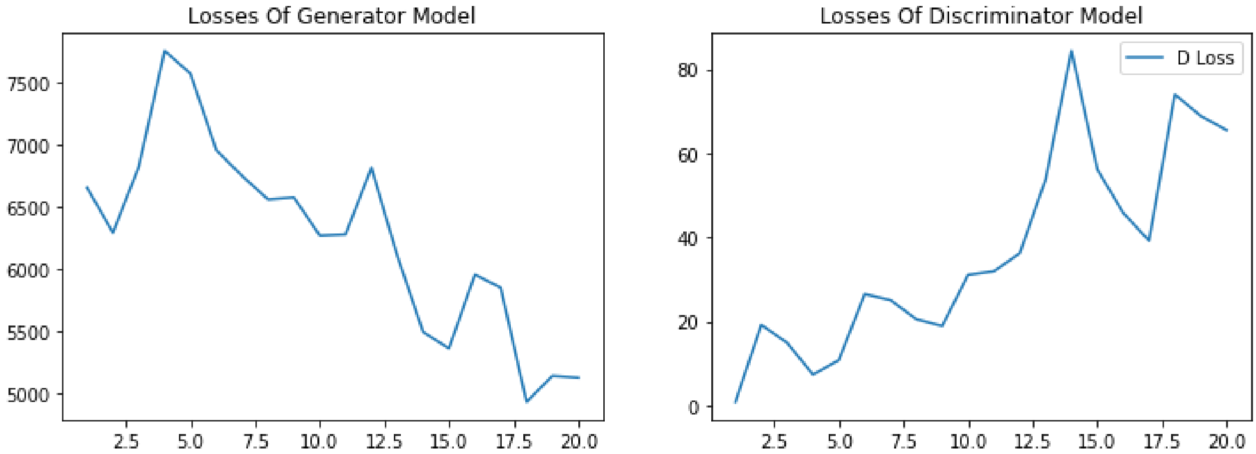


Figure 4.7 Loss curves for 550nm, 560nm, 570nm.

Likewise, charts are they’re till the wavelength of 700nm.

7. Evaluation

Models Comparison

Here we compared the 3 of our models DEEP CONV NET, HYPER-UNET and RESNET. Compared on the metrics of RMSE and rRMSE values.

		NTIRE 2020 dataset (Primary)		CAVE dataset		NUS dataset	
		train	valid	train	valid	train	valid
Deep ConvNet	RMSE	16.451	17.998	17.451	16.998	16.221	17.023
	rRMSE	0.767	0.643	0.737	0.713	0.761	0.839
Hyper UNet	RMSE	5.148	5.023	7.993	7.148	5.392	5.109
	rRMSE	0.473	0.464	0.703	0.695	0.656	0.612
Res Net	RMSE	10.352	11.821	14.894	15.821	14.237	15.452
	rRMSE	0.631	0.642	0.731	0.732	0.791	0.603

Table 1.1 Our models comparison.

Models Evaluation with existing models.

	NTIRE 2020				CAVE dataset				NUS dataset			
	Arad	A+	CNN	HyperNet	Arad	A+	CNN	HyperNet	Arad	A+	CNN	HyperNet
RMSE	–	–	–	5.085	5.61	2.74	2.55	7.0705	4.44	2.92	2.86	5.7505
rRMSE	–	–	–	0.418	0.499	0.426	0.469	0.769	0.190	0.142	0.152	0.684

Table 1.2 Comparison of best model with existing models.

8. Conclusion

The Pix2Pix GAN used was giving reasonable results in generating the images. We try to extend this idea of Pix2Pix GAN to generate the 31 spectral channels from 400 nm to 700 nm. We tried to generate 3 channels at a time to generate all the channels.

The U-Net architecture in the generator was big and it takes more running time during runtime. So, we are tried to replace the U-Net architecture with the RESNET model in our generator to make the model fast and it worked runtime got reduced.

We worked on relatively new dataset NTIRE 2020 which is an updated version of NTIRE 2018. We found a smaller number of previous works on this dataset, so we took this dataset as our primary dataset.

REFERENCES

1. Can, Yigit Baran, and Radu Timofte. "An efficient CNN for spectral reconstruction from RGB images." arXiv preprint arXiv:1804.04647 (2018).
2. Zhao, Yuzhi, et al. "Hierarchical regression network for spectral reconstruction from RGB images." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.
3. Koundinya, Sriharsha, et al. "2d-3d cnn based architectures for spectral reconstruction from rgb images." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018.
4. Stiebel, Tarek, et al. "Reconstructing spectral images from rgb-images using a convolutional neural network." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018.
5. Arad, Boaz, et al. "Ntire 2020 challenge on spectral reconstruction from an rgb image." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.
6. Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial networks." arXiv preprint arXiv:1406.2661 (2014).
7. Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-image translation with conditional adversarial networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134. 2017.
8. Liu, Xu, et al. "Multispectral image reconstruction from color images using enhanced variational autoencoder and generative adversarial network." IEEE Access 9 (2020): 1666-1679.
9. Aeschbacher, Jonas, Jiqing Wu, and Radu Timofte. "In defense of shallow learned spectral reconstruction from rgb images." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017.

List of Figures

List of Tables

List of Abbreviations and Symbols

- **GAN:** Generative Adversarial Network.
- **Pix2Pix GAN:** Pixel to Pixel Generative Adversarial Network
- **Resnet:** Residual Network.
- **cGAN:** Conditional generative adversarial network.
- **Conv:** Convolution layer.
- **Batch Norm:** Batch Normalization.
- **ReLU:** Rectified linear unit.
- **Gloss:** Generator loss.
- **Dloss:** Discriminator loss.
- **NM:** nano meter.
- **RMSE:** Root Mean Square Error.
- **rRMSE:** Relative Root Mean Square Error.
- **LeakyReLU:** Leaky rectified linear unit.

ACKNOWLEDGEMENTS

First, we want to thank our mentor Dr. Shiv Ram Dubey sir and Dr. P Viswanath sir for guiding us in this project and for the suggestion we got from both during the preparation for the evaluations.
Team members are also shown their active participation in every step of the project. In the beginning we started from scratch like learning the concepts of GAN’s etc.
With the help of this project, we got to know about new topics and team effort made us get some good results at the end.

For the technical support our mentors were there always to help us. They clarifies / replies to our doubts right away.
Workload split between members and work division; all went well. We are so happy to be part of this kind of project with our mentors.