Deep Learning and its Applications
Project Presentation on Dehazing Images
Group-04

Nikhil T R, Kaustubh Verma, Aj R Laddha

June 10, 2019

Introduction - Haze

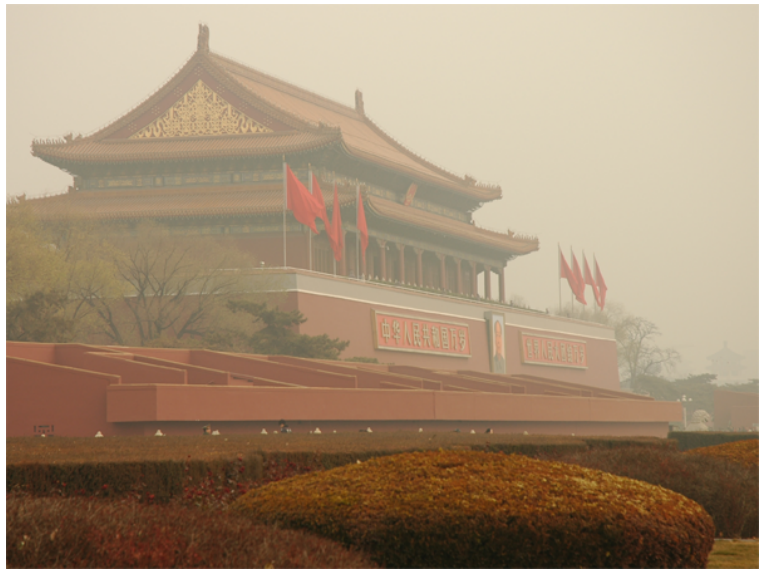


Figure: Hazy Image

Haze is traditionally an atmospheric phenomenon in which dust, smoke, and other dry particulates obscure the clarity of the sky.
[1] [2]

[1] https://en.wikipedia.org/wiki/Haze
[2] http://kaiminghe.com/cvpr09/

Removing Haze

Many editors like Photoshop come with an inbuilt dehaze filter.

Figure: Before and After

3

[3] https://blogs.adobe.com/jkost/2015/06/adobe-announces-camera-raw-9-1-for-photoshop-cc-and-lightroom-cc.html

Problem Statement

To remove haze from single image, while making it appear as natural as possible.

Motivation

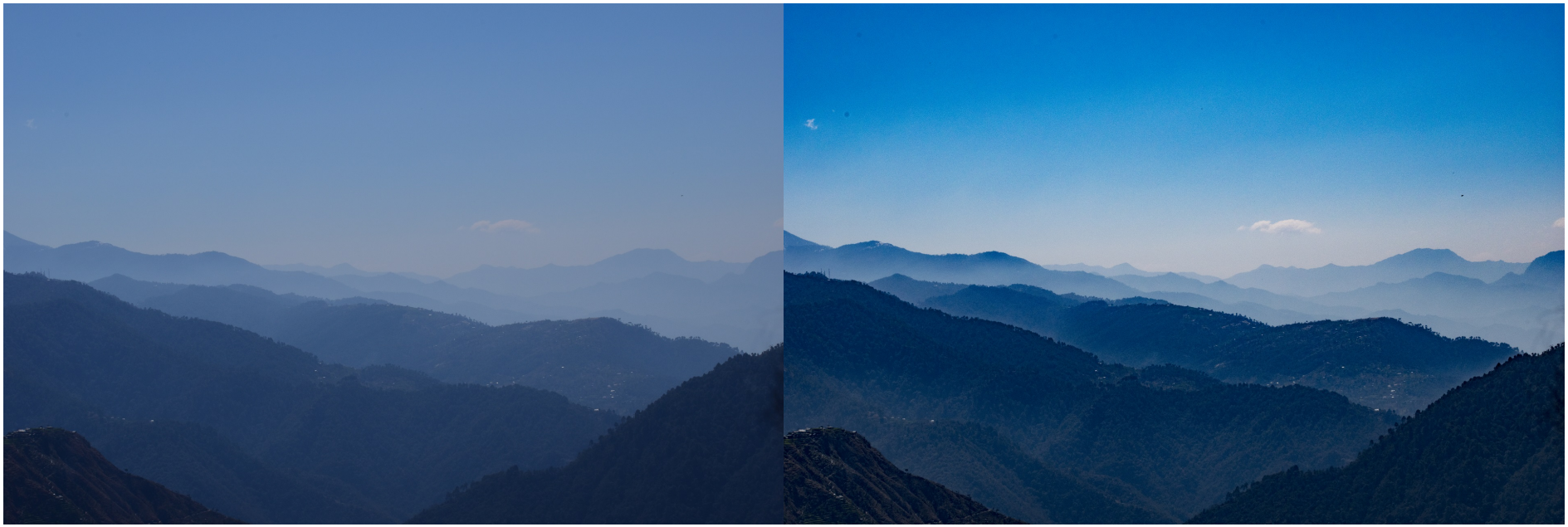
Inbuilt dehaze filter is not enough. For example, consider the images below:

(a) Raw (b) Dehaze using Photoshop

Picture Credits: Nikhil T R

This is the best quality image that was possible using the inbuilt filter. We can see clearly that second picture still has a lot of haze.

Challenges



(c) Raw (d) Photoshop Dehaze

Picture Credits: Nikhil T R

Challenges we faced

1. Change in Colour of objects.
2. Shift of Saturation.
3. Recognising Transmission map.
4. Understanding Edges.
5. Maintaining Resolution.
6. Artifacts.

Datasets

NYU V2 Depth - haze to be synthesized
https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html
The dataset was made using injecting haze based on the depth information already available. In the end this gave us an image, with its hazy counterpart.

Methodologies Explored

- DehazeNet (2016)
  https://arxiv.org/pdf/1601.07661.pdf
- AOD-Net (2017)
  http://openaccess.thecvf.com/content_ICCV_2017/papers/Li_AOD-Net_All-In-One_Dehazing_ICCV_2017_paper.pdf
- Auto Encoder for learning the depth.
- Pix2Pix (2017) https://arxiv.org/pdf/1611.07004.pdf.
- Final Model - Single Image Haze Removal using a Generative Adversarial Network
  https://arxiv.org/ftp/arxiv/papers/1810/1810.09479.pdf

DehazeNet - Introduction

- Assumes an atmospheric scattering based model for haze. This is taken from another paper.

$$I(x) = J(x)t(x) + \alpha(1 - t(x))$$

- Emphasis on calculating the medium transmission map $t(x)$ of a given hazy image $I(x)$. The original image $J(x)$ can then be recovered using it.
- Use of a new activation function - BReLu (Bilateral Rectified Linear Unit)

Model's Task: Hazy Image $\rightarrow$ Transmission Map.
α is estimated from $I(x)$ itself.

Architecture

Layers:

1. Feature Extraction using Convolution and Maxout.
2. Multi Scale Mapping using Convolution.
3. Local Extremum using MaxPool.
4. Non Linear Regression using Convolution and BRelu

Is “end to end trainable”.

AOD(All-in-One Dehazing)-Net - Introduction

- Based on a re-formulated atmospheric scattering model.
- Generates the clean image through a light-weight CNN,rather than estimating the transmission matrix and the atmospheric light separately as previous model.
- Easy model allows AOD-Net into other deep models, e.g., Faster R-CNN, for improving high-level tasks on hazy images.
- The key to achieve haze removal is to recover a transmission matrix.However, the estimation is not always accurate, and some common pre-processing such as guild filtering or softmatting further distort the hazy image generation process

AOD-Net **Transformation Formulae**

- I(x) =J(x)t(x) +A(1-t(x))
  where I(x) is the hazy image, t(x) is the transmission matrix, A is a scalar parameter in the model which is related to ambient light.
- $t(x) = e^{-\beta * d(x)}$
  The transmission matrix depends on the depth d(x).
- Objective: minimize: J (x)=K(x)I(x) - K(x)+b

$$J(x) = \frac{1}{t(x)} I(x) - A\frac{1}{t(x)} + A$$

$$J(x) = K(x)I(x) - K(x) + b$$

where K(x) is defined as,

$$K(x) = \frac{\frac{1}{t(x)}(I(x) - A) + (A - b)}{I(x) - 1}$$

Model Details

Model Architecture

Composed of two parts

1.K-estimation module that uses five convolutional layers to estimate K (x)

2.Clean image generation module.

RELU used,momentum and the decay parameter set to 0.9 and 0.0001.

Loss function - Simple Mean Square Error (MSE).

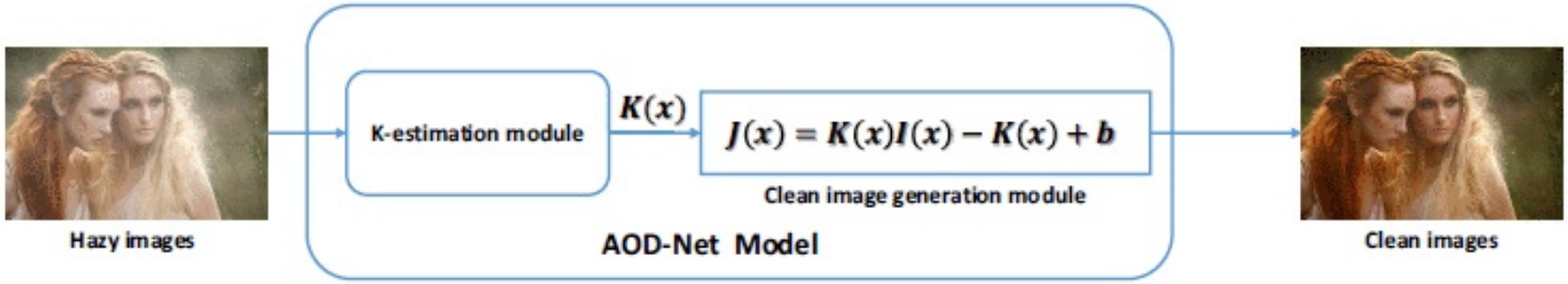
Model's Pros

- Improved PSNR(Peak signal-to-noise ratio),SSIM(Structure similarity index),quality of image
- Very fast processing speed.
- Can be embeded with other deep models.

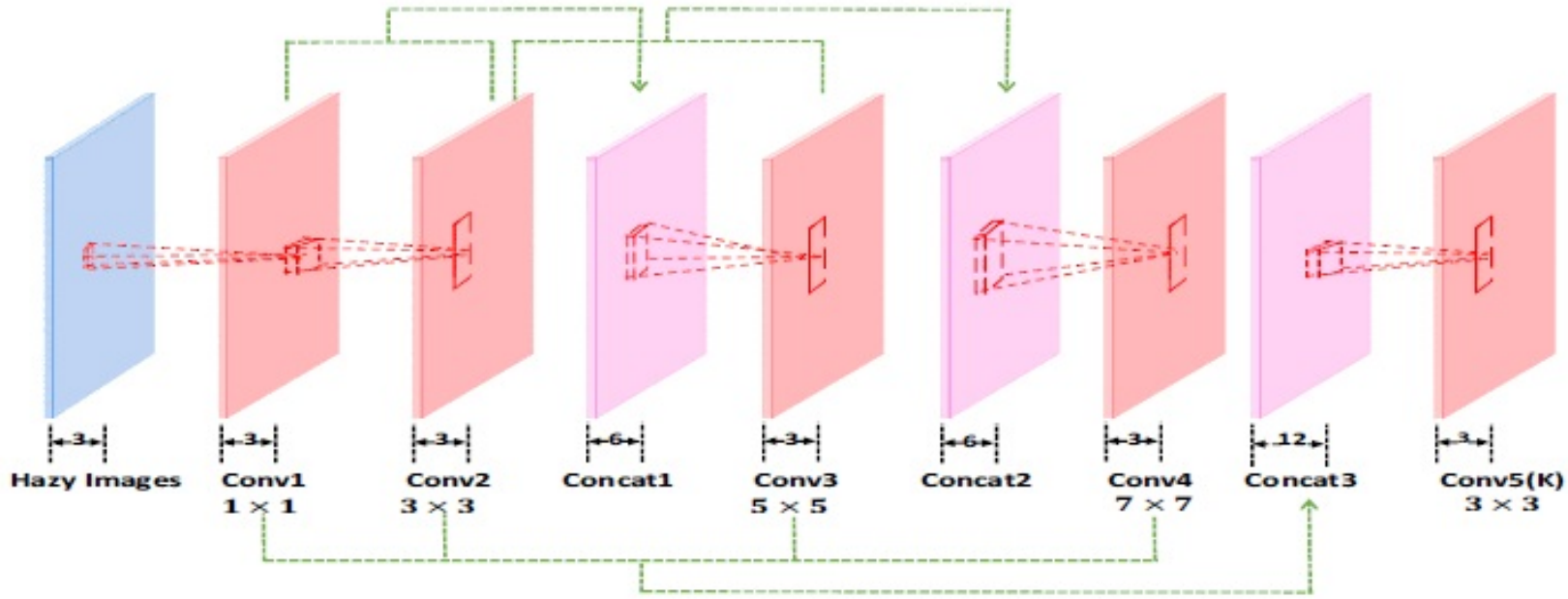
Dataset

Created synthesized hazy images by using the ground-truth images from depth Indoor NYU2 Depth Database.

Model Details



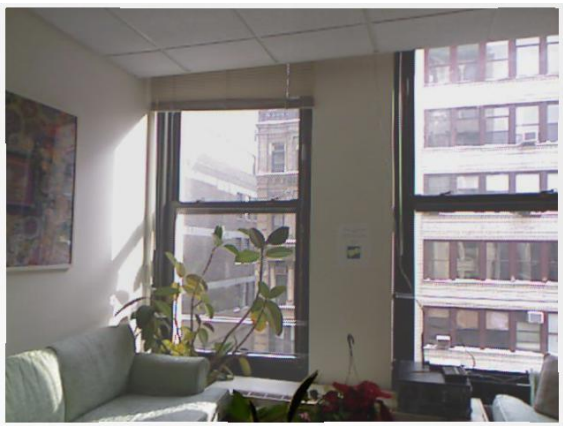
(a) The diagram of AOD-Net



(b) K-estimation module of AOD-Net

Figure: Model Architecture

Our Results



(a) Hazy Image

(b) Ground Truth

(c) Dehazed Image

(d) Hazy Image

(e) Ground Truth

(f) Dehazed Image

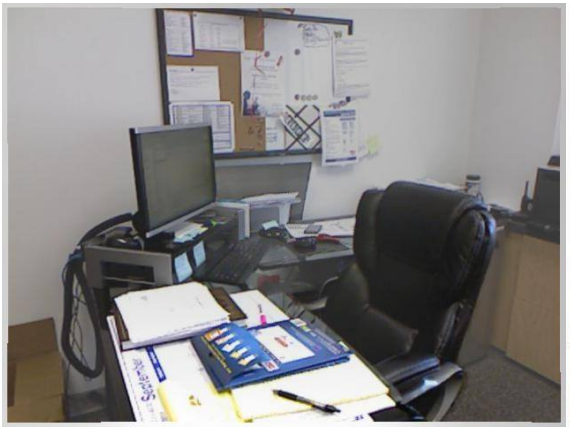
Observations

1. The performance is not satisfactory.
2. The temperature has been altered in the generated image.
3. Some colour translation has taken place. True colour is not retained in the generated image.

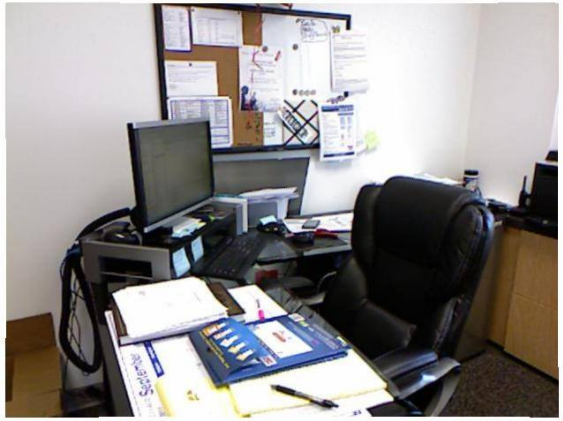
Auto Encoder

- This architecture was provided by Dr. Aditya Nigam for testing out for our purpose.
- We trained this auto encoder to learn the mapping from Hazy to Dehazed images.
- **Loss:** MSE

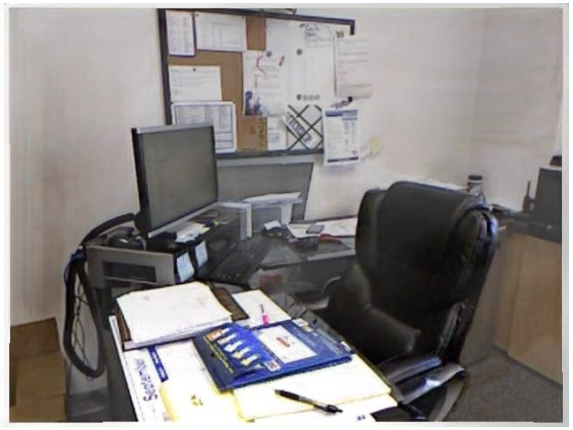
Results



(g) Hazy Image



(h) Ground Truth



(i) Dehazed Image

(j) Hazy Image

(k) Ground Truth

(l) Dehazed Image

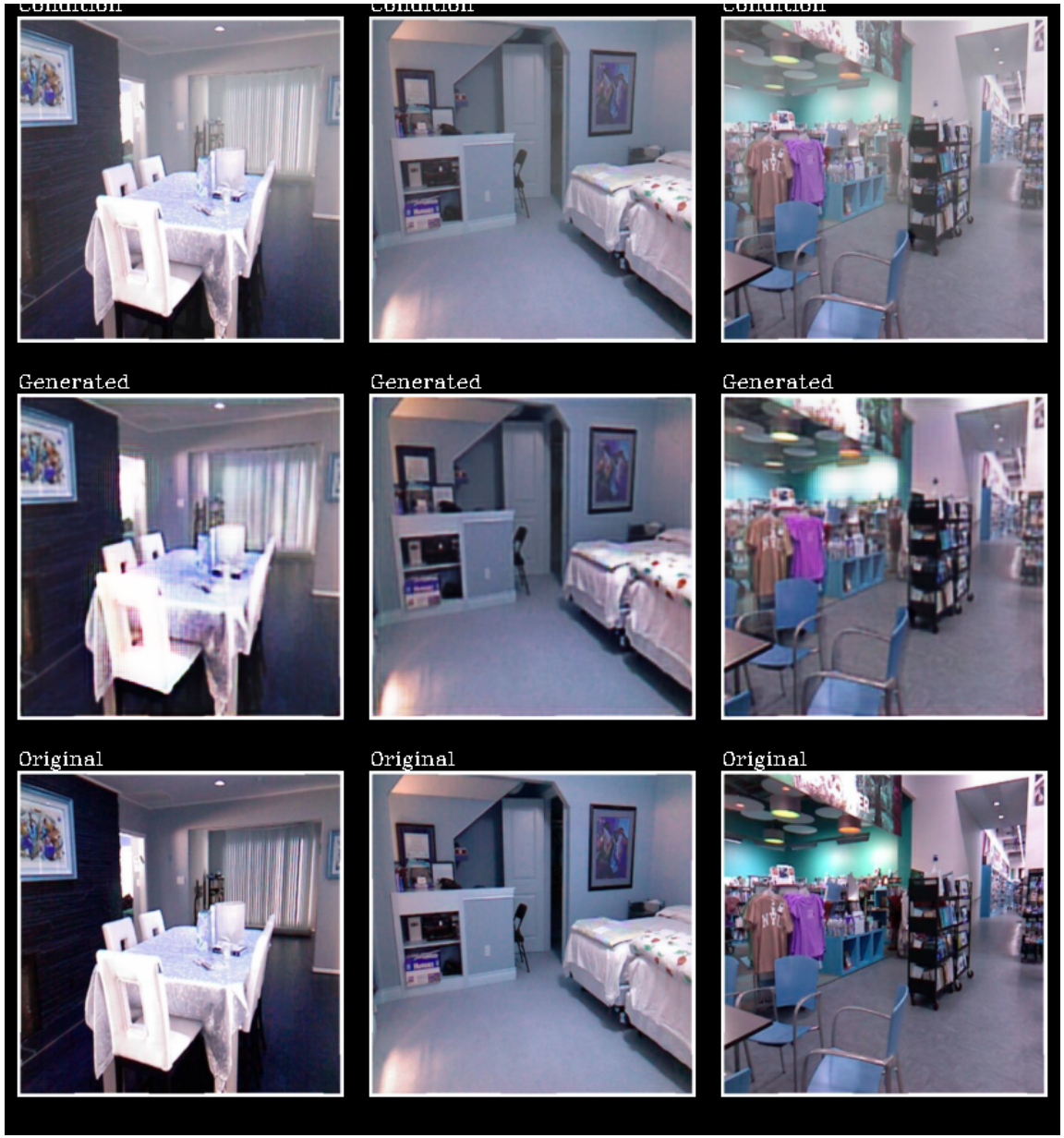
Observations

1. The performance is not satisfactory.
2. Artifacts were observed as in image (c).
3. A lot of haze is still left out as in image (f).

Pix2Pix

Pix2Pix uses a Conditional GAN technique for image to image translation. We thought it would be a good idea to try out a similar translation form Haze to Dehazed images.
Pix2Pix uses U-Net architecture for Generator and Patch-GAN for discriminator with Adversarial training.
PixPix has been observed good results on many image to image translation problems.

Results


Generated
Generated
Generated
Original
Original
Original

Results


Generated
Generated
Generated
Original
Original
Original

Observations

1. Pix2pix gives better results as compared to the previous AOD and AE models.
2. Artifacts were noticed after certain epochs in training. This could probably be because of the L2 loss function that was used.
3. Upon further scrutiny, we noticed that the model is not able to maintain sharpness.

Final Model

Model Details:

- The model replaces the conventional U-Net by the 54-Layer Tiramisu. The latter is extremely parameter efficient, and has state of the art performance in semantic segmentation.
- It also uses the Patch Discriminator to reduce artifacts.
- Smooth L1 Loss is added to the standard conditional GAN loss.

Variations:

1. **V1:** Only Smooth L1 Loss added to standard CGAN Loss.
2. **V2:** Smooth L1 and Perceptual Loss added to standard CGAN Loss.

Metrics:

1. **SSIM:** Structural Similarity Index
2. **PSNR:** Peak Signal to Noise Ratio

Model Architecture

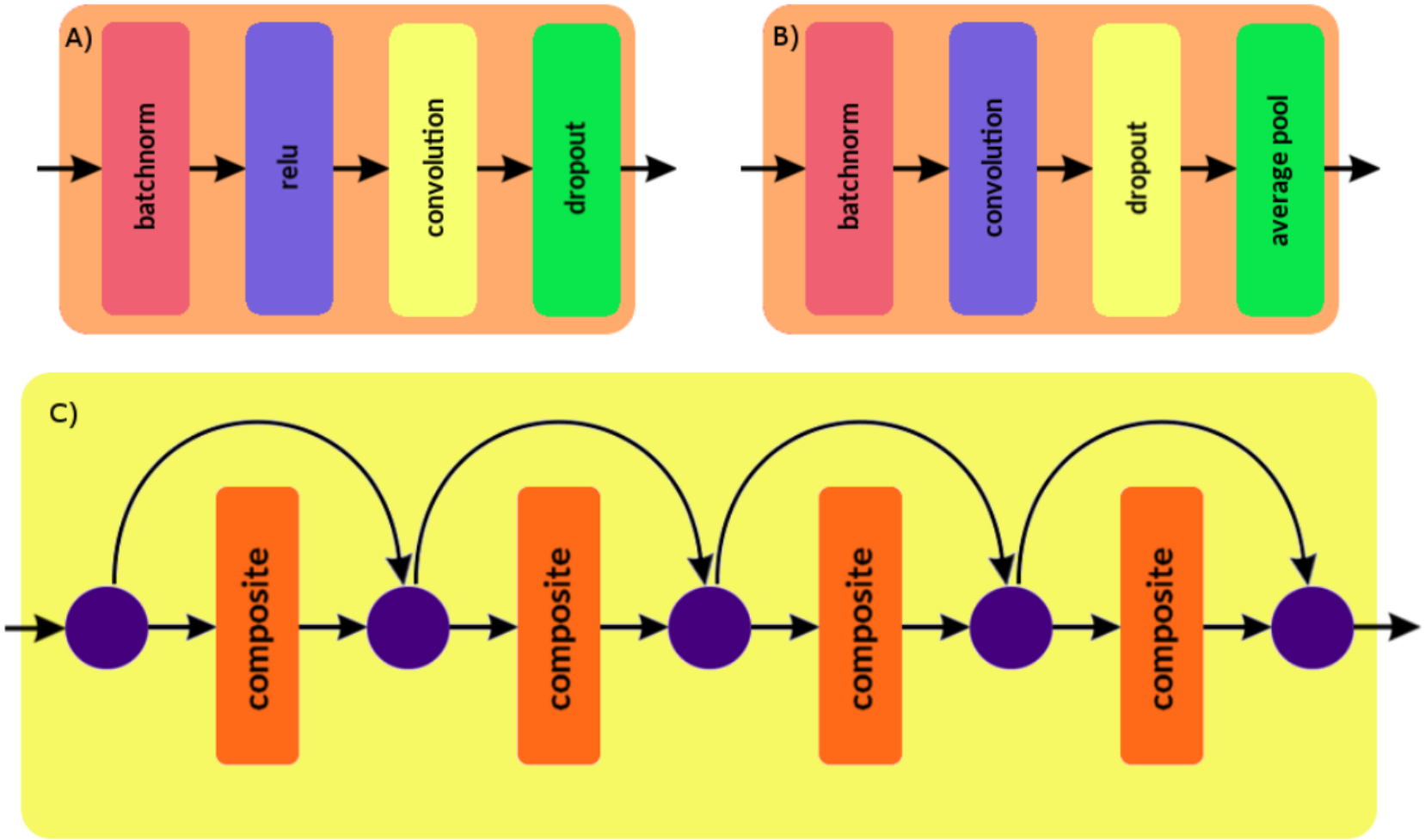


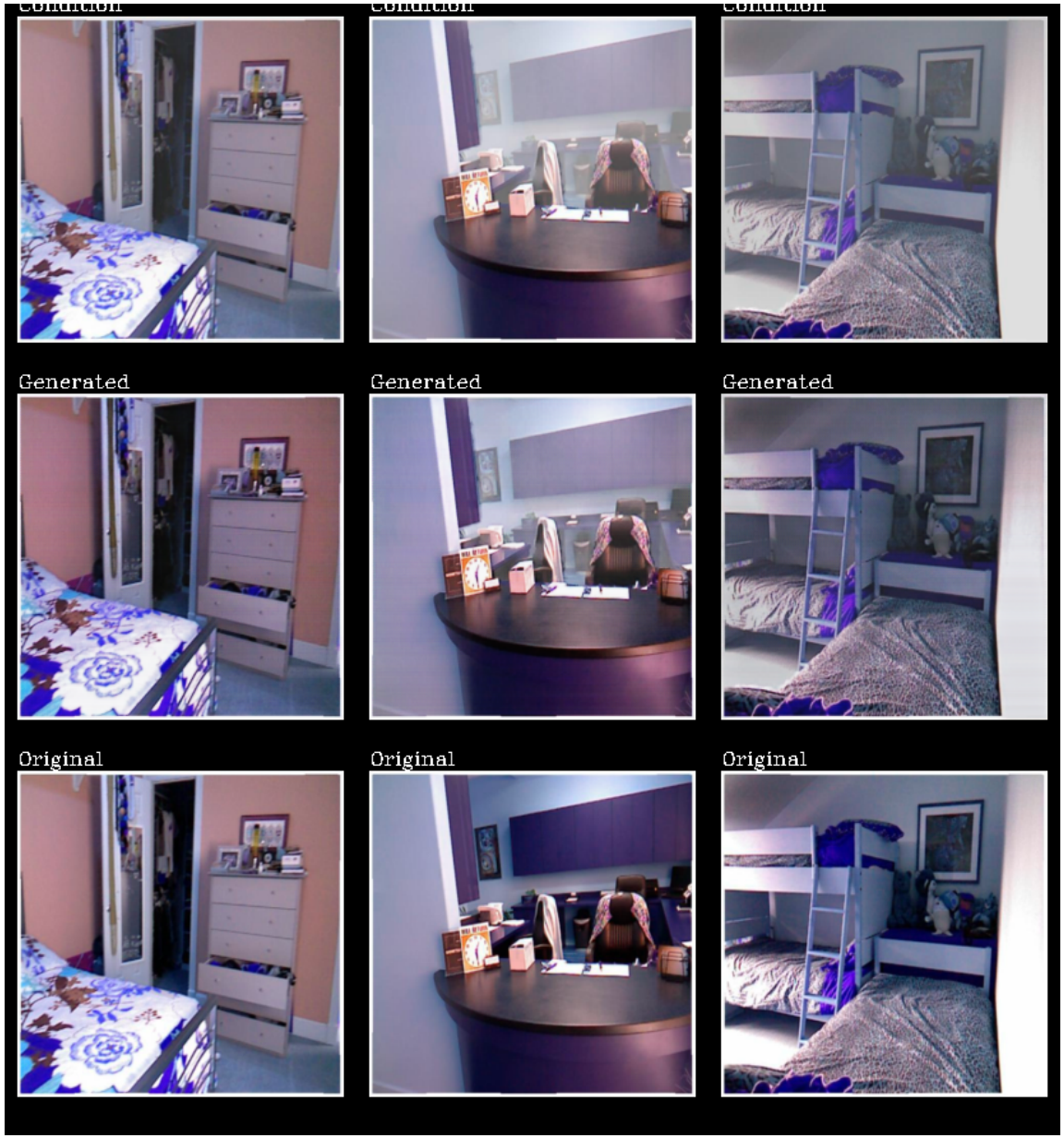
Fig. 3: Expanded view of the various components of the generator. **A: Transition Down** (TD) layer; **B: Composite Layer** (within a dense block); **C: Dense Block** with 4 composite layers. Curved lines ending on purple dots imply a concatenation operation. The Transition Up layer comprises only of a transpose convolution operation, and hence not shown here.

Model Architecture

Generator	
Operation	**Output Shape**
Input Conv	256,256,48
DB Encoder 1	256,256,96
TD 1	128,128,48
DB Encoder 2	128,128,96
TD 2	64,64,48
DB Encoder 3	64,64,96
TD 3	32,32,48
DB Encoder 4	32,32,96
TD 4	16,16,48
DB Encoder 5	16,16,96
TD 5	8,8,48
DB Bottleneck	8,8,228
TU 5	16,16,48
DB Decoder 5	16,16,192
TU 4	32,32,48
DB Decoder 4	32,32,192
TU 3	64,64,48
DB Decoder 3	64,64,192
TU 2	128,128,48
DB Decoder 2	128,128,192
TU 1	256,256,48
DB Decoder 1	256,256,192
Output Conv	256,256,3
Discriminator	
Operation	**Output Shape**
Layer 1	128,128,64
Layer 2	64,64,128
Layer 3	32,32,256
Layer 4	31,31,512
Layer5	30,30,1

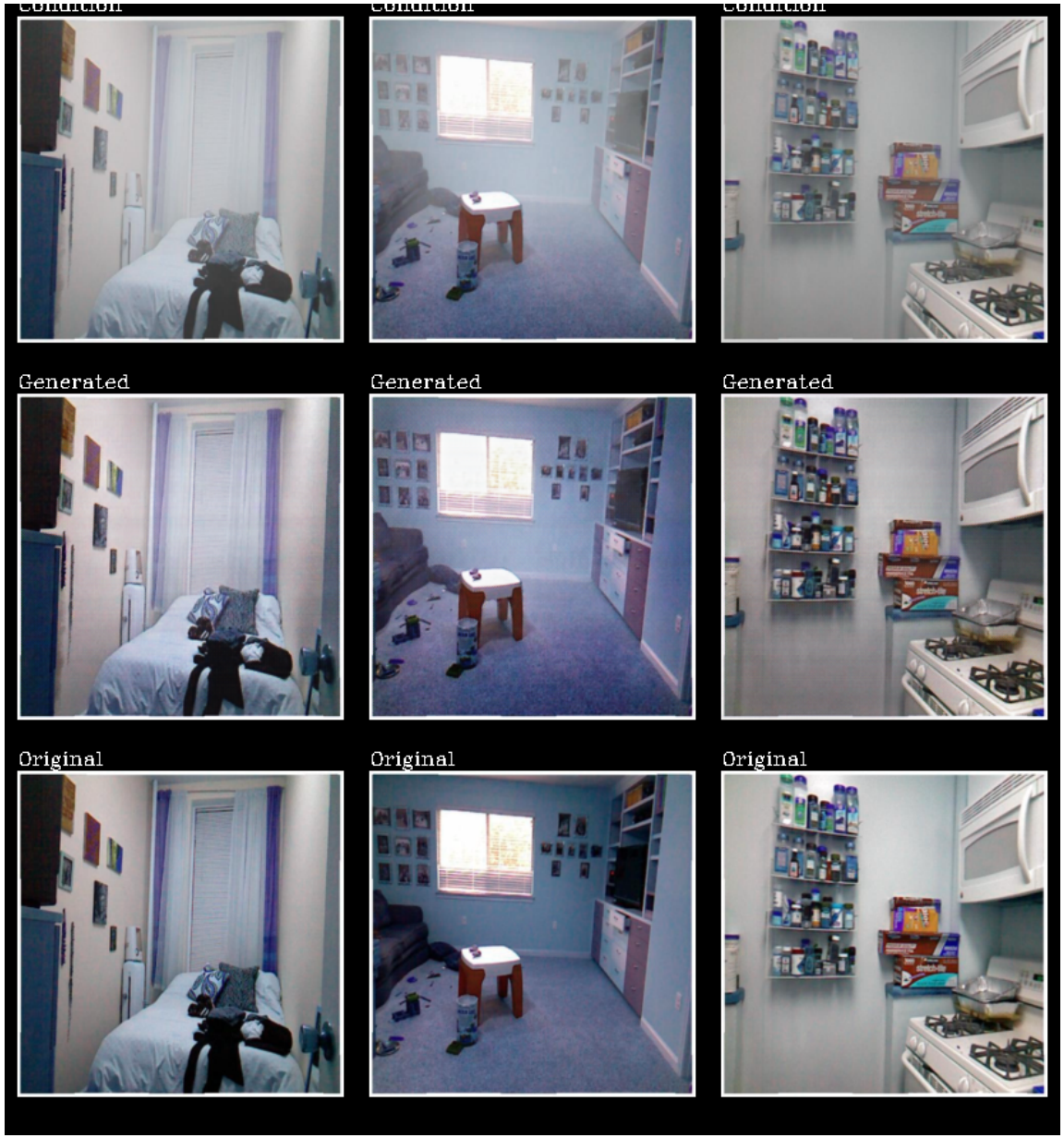
Table 1: Output shapes of the images when passed through each component of the generator and discriminator. Input shapes are 256,256,3 for generator and 256,256,6 for the discriminator.

Results - V1 - Only L1



Results - V1 - Only L1

Results - V2 - L1 Perceptual



Results - V2 - L1 & Perceptual

Observations

1. As can be seen from the above images of both V1 and V2 the results are much better than observed in previous model.
2. Also it can be observed that when using weighted loss of smooth L1, perceptual loss and pix2pix loss the details are much improved, as can be seen in V2 model.
3. The observed PSNR and SSIM values for this model was much higher with average SSIM on test data being 0.87 and average PSNR being 23.56