UNSUPERVISED DEHAZING ON REAL-WORLD IMAGES USING GANS

A Thesis

by

ANJALI CHADHA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Zhangyang (Atlas) Wang
Committee Members,	Theodora Chaspari
	Xiaoning Qian
Head of Department,	Dilma Da Silva

May 2019

Major Subject: Computer Science

ABSTRACT

Dehazing is an important pre-processing step in almost all computer vision systems deployed in outdoor settings. Existing dehaze methods are either based on heuristic image priors, or on models trained with hazy-clear image pairs of the same scene. In practice, however, obtaining paired images isn't feasible, so researchers often add synthetic haze on clean images to create paired data sets. This might result in a domain shift when models trained on synthetic images are used for real-world outdoor settings. In this work, we propose UD-GAN (**U**nPaired **D**ehaze GAN), a novel generative adversarial network based dehazing model, which can generate clean images using only unpaired data. UD-GAN can not only be trained using a large repository of real-world clear and hazy images but it can also learn the characteristics of true haze better than other models trained on synthetic data. Moreover, our method is model-agnostic and would perform well even when the assumptions made by the physical model don't hold true. UD-GAN uses an attention-based generator and we explore two types of attention maps which can be used along with this generator. Finally, we compare the performance of our approach using full-reference metrics, no-reference metrics, and the accuracy in object detection. The qualitative and quantitative results generated by UD-GAN are on-par with the current state-of-the-art dehazing methods.

DEDICATION

To Papa

ACKNOWLEDGMENTS

The timely completion of this work wouldn't have been possible if it hadn't been for my advisor Prof. Zhangyang Wang. His invaluable guidance helped me veer through many obstacles during the course of my research. I would like to extend my sincere thanks to Prof. Theodora Chaspari and Prof. Xiaoning Qian for serving as my committee members and providing their constant support. I also appreciate the support of the CSE graduate office staff, especially, Karrie Bourquin, for their patience and help in meeting multiple deadlines for my thesis completion.

Special thanks to my roommates and friends: Isha, Vrushali, Hemangi, Tushar, Mayank, Christopher, and Puneet. Whenever I indulged in self-doubt or fear, they were always there for my rescue and provided me with a safe and loving ecosystem.

Finally, I would always be deeply indebted to my family for their incessant love and support. Sitting thousands of miles away, across the oceans, they have always been my ardent supporters instilling optimism in me during some of the most trying moments.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Prof. Zhangyang Wang and Prof. Theodora Chaspari of the Department of Computer Science and Engineering and Prof. Xiaoning Qian of the Department of Electrical and Computer Engineering. All work for the thesis was completed independently by the student.

Portions of this research were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing. There are no outside funding contributions to acknowledge related to the research.

NOMENCLATURE

UD-GAN	Unpaired Dehazing Generative Adversarial Networks
CNN	Convolutional Neural Networks
DCP	Dark Channel Prior
GAN	Generative Adversarial Networks
DNN	Deep Neural Networks
MSCNN	Multi-Scale Convolutional Networks
PSNR	Peak Signal-to-noise Ration
SSIM	Structural Similarity Index
IQA	Image Quality Assessment
SSEQ	SpatialSpectral Entropy-based Quality
BLIINDS	BLind Image Integrity Notator using DCT Statistics
mAP	Mean Average Precision

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Haze is an atmospheric phenomenon caused by the presence of dust, smoke, and other dry particulates [3]. When a reflected light from an object surface hit these particles, their course of propagation changes. This results in images with poor contrast and faint colors. The deterioration in image quality is directly proportional to the distance of the scene from the camera since very less amount of light reflected from the far scenery will reach the camera sensor.

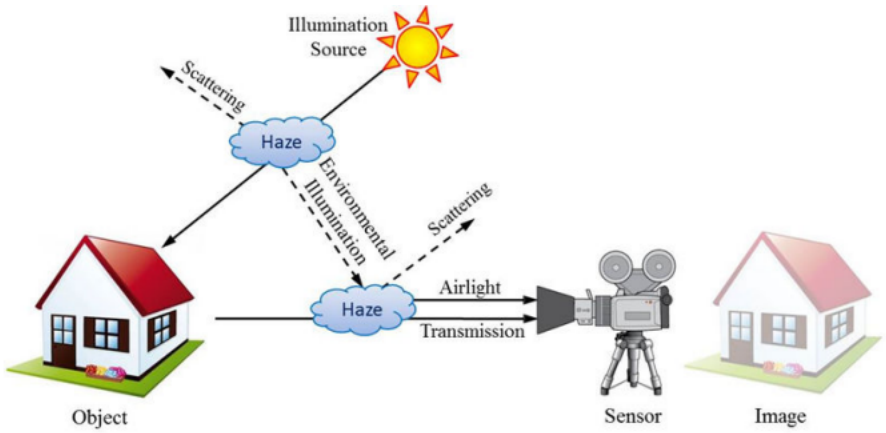

Figure 1.1: Image formation under hazy weather conditions and atmospheric scattering model (Reprinted from [1])

Outdoor computer vision applications such as autonomous vehicles [4], object tracking and recognition, and traffic surveillance cameras [5] perform best when the input images are clear. Weather conditions such as haze and fog work against the success of these applications by deteriorating the visibility of scenery. Likewise, hobbyists or professional photographers, specifically those interested in long-distance photography, also suffer from haze effects because what is seen with the naked eye is not what is captured by the camera sensor. Due to these concerns, image dehazing is a widely sought after problem and has been studied by researchers for over a decade.

We can formally define the term "dehazing" as the process of recovering an image with im-

proved visibility and better colors as if it was captured on a clear day. When the haze removal model takes only a single scene image as input, the process is called "*single image dehazing*". Single image dehazing methods can be largely categorized as prior-based and data-driven methods. Prior-based methods [6, 7, 8] make sophisticated image statistics assumptions, or use strong priors to recover the actual image scenery. However, when the assumptions made by these methods are violated, these priors no longer provide satisfactory results. Data-driven methods [1, 9, 10], on the other hand, experiments with different variations of CNN to perform dehazing. While deep learning methods learn to recover a clear image by training on large data set of clear and hazy images, prior based methods use only an input image. Deep learning based methods perform better in most situations due to this very reason.

Figure 1.2: An example of hazy image (left) and it's clear version (right) generated by UD-GAN

Most of the DNN-based models use a paired data set where every image scene is available in two variants – clear image and hazy image. Obtaining such kind of paired data sets is impractical for most real-world cases and often researchers add a layer of fake haze on clear images to work around this issue. Dehazing models which are trained in a supervised manner have a few downsides.

- First, in most real-world scenarios, it is impractical to retrieve the additional meta-data such as depth, 3-D models that helps in authentic paired images.
- Second, synthetically created hazy images don't represent the actual distribution of true hazy images and therefore models trained using this data may not perform well in the real-world settings.

One solution to overcome these drawbacks is – unsupervised learning. In unsupervised learning, the model will be fed with unpaired clear and images, thus eliminating the paired data constraint. CycleDehaze [11] and, DeepDCP [12] (inspired by Dark Channel Prior [8]) are recently proposed dehazing methods based on unsupervised learning approach. The problem of recovering a clear image from a hazy one bears a strong similarity with another category of computer vision tasks –image-to-image translation. Multiple unsupervised ways [13, 14] of image translation have been proposed recently. Inspired from these methods, our model uses generative adversarial network (GANs) to learn the low-level mapping between hazy and clear images in an unsupervised way

In this thesis work, we propose **UD-GAN** (**U**nPaired **D**ehaze GAN), a dehazing approach which can be trained using real-world images in a fully unpaired fashion. Figure 1.2 presents UD-GAN's results on a sample hazy image. Unlike other unsupervised methods which use a cycle-consistency loss to regularize the training, UD-GAN uses an attention-based generator network along with a combination of hybrid loss functions. Adding attention to the generator helps the model to focus on some specific areas in an image while creating a clear image. We further explore two types of attention maps and compare the model's performance using each attention map. This comparison underscores the importance of the attention map in our model. Most dehazing methods are based on the physical scattering model, however, our approach is independent of the scattering model. This gives our model an advantage over other model-dependant methods for situations where the relationship between the original scene and the hazy scene is fairly complex and cannot be captured by the scattering model. Finally, we evaluate our results using full-reference metrics, no-reference metrics, and its accuracy in object detection done under hazy conditions.

To the best of our knowledge, our method is the first one to use unpaired learning to recover hazy images by injecting visual attention in the generator. The remainder of this work is organized as follows. In Section 2, we first discuss an important physics-based model which explains the mathematical relationship between hazy and clear image. Following this, we review the past literature on dehazing ranging from prior-base classical methods to deep learning methods using both supervised and unsupervised learning. In Section 3, we discuss the detailed architecture of UD-GAN, our final loss function, and the two types of attention maps. In Section 4, we provide both visual and quantitative results of our approach and compare it with other state-of-the-art methods. In the last section, we conclude our results and also discuss future work to be done in this area.

2. BACKGROUND AND RELATED WORK

2.1 Atmospheric Scattering Model

The atmospheric scattering model [15, 16, 17], also referred as Koschmieder model is a widely used model to describe the generation of a hazy image. Figure 1.1 and 2.1 explains this phenomenon in action. The model can be formally described as:

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{2.1}$$

where $I(x)$ is observed hazy image, $J(x)$ is the real image to be recovered, $t(x)$ is the medium transmission matrix, A denotes the global atmospheric light, and x represents the pixels in the input hazy image I. The medium transmission map $t(x)$ is a distance dependent factor and is mathematically defined as:

$$t(x) = e^{-\beta d(x)} \tag{2.2}$$

where β is the atmosphere scattering coefficient and $d(x)$ is the distance between the camera and the scene point. Atmospheric light A is the light coming from an object at an infinite distance i.e diffusion of light by the haze and it is usually represented as a constant vector with 3 components in RGB space $A = (A_r, A_g, A_b)$

Based on the above model (2.1), hazy image $I(x)$ can be described as a linear combination of two factors:

i. ***Direct attenuation*** $(J(x)t(x))$ represents the amount of light scattered or decayed before it reached camera lens.

ii. ***Airlight*** $(A(x)(1 - t(x)))$ is a function of scene depth and global atmospheric light A. It represents the change in scene brightness due to environmental light scattering.

Dehazing methods based on this physical model estimate parameters A and $t(x)$ to recover clear scene $J(x)$. This is an under-constrained problem because there are three unknowns $A, t(x), J(x)$

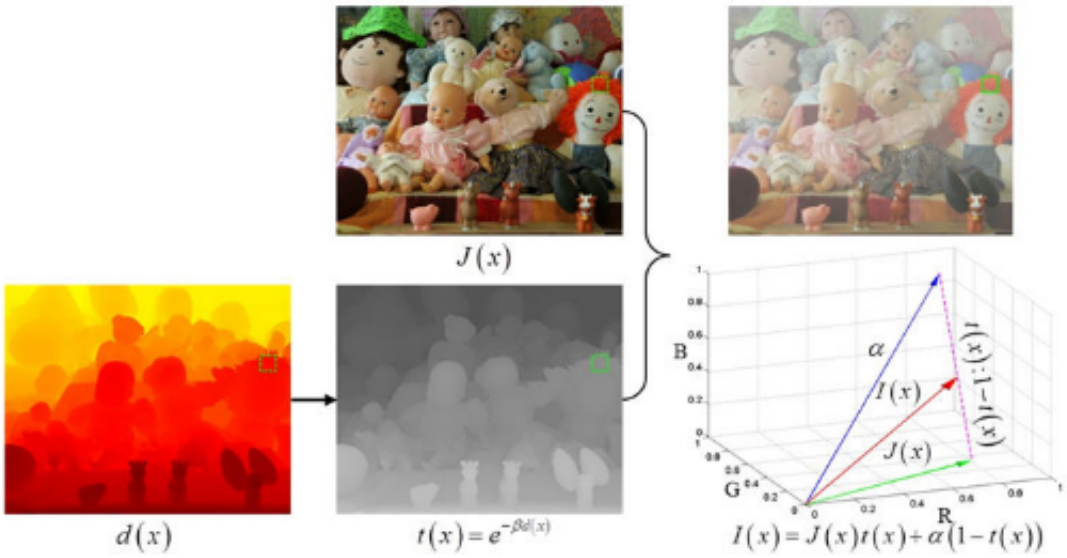

Figure 2.1: Different components in Atmospheric Scattering Model (Reprinted from [1])

and only a single input image $I(x)$.

2.2 Overview of Dehazing Approaches

Earlier approaches tried using several image enhancement techniques - contrast-based [18], histogram-based [19]. These methods didn't perform well because they didn't factor in the varying haze density in the image. We can widely classify existing image dehazing methods on the basis of their inputs:

i. *Multiple Images*: Different images are obtained under various weather conditions to perform dehazing [16, 17, 20]

ii. *Polarizing Filters*: To avoid the inconvenience of capturing images under various weather conditions, various filters were used to simulate different weathers [21]

iii. *Single Image + Additional Metadata* such as depth or 3-D model of the scene [22]

iv. *Single Image* dehazing methods uses only the hazy image as input.

Except for single-image based approach (iv), the other three methods listed above are not practical under real-life settings as we generally have a single hazy image with no additional data. In this

work, our main focus is on single image dehazing methods

2.3 Single Image Dehazing

Most state-of-the-art single image dehazing techniques are based on the atmospheric scattering model and approximate the critical parameters A and $t(x)$ using:

i. Classical Prior-Based methods

ii. Data-Driven methods (based on CNNs or GANs)

2.3.1 Prior-Based Methods

Classical prior methods depend on hand-crafted features using various properties of an image such as color, texture, and contrast to remove the haze from an input image. They use natural image priors and depth statistics. Fattal et al. proposed Independent Component Analysis (ICA) based minimal input approach for dehazing a color image [7]. This technique, however, was time-consuming and didn't perform well on dense-haze scenes. One of the most popular prior-based methods, Dark Channel Prior (DCP) [8], approximates the transmission matrix very reliably. It is based on an observation that at least one color channel has some pixels with very low intensities in most of the haze-free patches. However, DCP performance falls when the objects in an image are similar to the atmospheric light. Zhu et al. proposed a color attenuation prior and created a linear model for estimating the scene depth of the hazy image [23]. Meng et al. developed a contextual regularization dehazing method and explored inherent boundary constraints to restore the clear images [24]. Li et al. applied dehazing on video sequences by jointly estimating scene depth and recovering the clear latent image [25]. Berman et al. put forward a non-local prior based approach which is based on the assumption that each color cluster in the clear image becomes a haze-line in RGB space [6].

2.3.2 Data-Driven Methods

Based on deep learning, most data-driven methods either estimate the transmission map or directly recovers a clear image. Having a large knowledge bank in the form of training images

help these data-driven methods learn haze features better than prior based methods.

2.3.2.1 Supervised Learning

DehazeNet [1] proposed a supervised CNN-based method to learn the transmission matrix $t(x)$. Multi-scale CNN (MSCNN) [9] is another effective dehazing model which first predicts coarse scale transmission matrix and later refines it locally. Both DehazeNet and MSCNN estimates the transmission map using a trained model; there is an extra step in the workflow to recover the clear image. AOD-Net [10] removed this extra step by introducing an end-to-end design which directly generates a clear image as output without any separate or intermediate parameter estimation step. Another end-to-end dehazing model is based on densely connected pyramid architecture [26]. Gated Fusion Networks (GFN) [27] uses a fusion of white balancing, gamma correction and contrast enhancing techniques to achieve dehazing.

All the methods discussed in the last paragraph are *supervised*; all use dataset of paired (clear-hazy) images for training. Paired (clear-hazy) images are created either using depth meta-data information or by adding artificial haze by varying levels of A and β. Models trained on synthetic-haze images may not generalize well to real-world hazy images.

2.3.2.2 Adversarial Learning

Generative Adversarial Networks (GANs) [28] have been successful in image generation [29, 30], image manipulation tasks, such as style transfer, image inpainting. General GAN architectures consist of a generator and a discriminator who are adversarially trained at the same time. The discriminator's task is to correctly distinguish real samples and the output of the generator; while the generator's task is to output fake images which can fool the discriminator. The key to the success of GANs is the concept of an adversarial loss that forces the images generated by the generator to be indistinguishable from real photos.

i. ***Supervised GAN-based methods***:

PO-GAN (Perceptually Optimised GAN) combined adversarial loss with perceptual loss and guided filtering to directly generate a haze-free image as model output [31]. Zhang et

al. proposed a model which relaxed the assumption of constant global atmospheric light A in physical scattering model [32]. Besides dehazing, there are other supervised GAN-based models to enhance/restore images such as - deblurring [33], super-resolution [34], deraining [35].

ii. ***Unsupervised GAN-based methods***:

Recently multiple unsupervised GANs have been developed to learn the inter-domain mapping without using paired input samples, such as cycleGAN [13], disco-GAN [36], and dualGAN [37]. CycleGAN [13] uses a cycle-consistency loss to build a mapping between two different domains. CycleDehaze [11] combines the principle of cycleGAN with a perceptual loss to build a dehazing model. Golts et al. introduced a network using the DCP (Dark Channel Prior) energy function as a loss to recover the haze-free image [12].

3. PROPOSED METHODOLOGY

In this section, we discuss the architecture of UD-GAN as shown in Figure 3.1. We will explain the attention-based generator network and the two types of attention layers we experimented with in this work. Besides the generator and discriminator network architectures, choosing a good loss function is essential as well, especially for training a reconstruction network based on CNNs [14, 38]. Therefore, we trained UD-GAN using a hybrid loss function combining relativistic-GAN based loss and perceptual loss. This section will further elaborate attention layers and loss functions.

3.1 Attentive-Generator Network

The purpose of a generator is to create clear images from hazy ones. We use U-Net [39] as the backbone of our generator architecture. U-Net is essentially an encoder-decoder network with additional skip connections to share the low-level information at different depth layers between the input and output. This also enables the generator to synthesize images of higher quality.

Usually, paired learning methods compare the ground-truth clear image and output clear image using traditional L_2 or L_1 errors for regularizing the training process. However, in the case of UD-GAN, we don't have paired images. Therefore, we need to find another way to regularize the generator network. Inspired from [35, 40], we introduce attention layers to this U-Net based generator architecture. These attention layers help reinforcing the integral details of the input hazy images in the generated output. Attention layers are resized to the size of various feature maps and added to the U-Net generator network at different depths as shown in the architecture in figure 3.1. To generate an attention map, we use two techniques explained in the following section. Choosing an appropriate attention map helps immensely in reducing both training time and generating better output images.

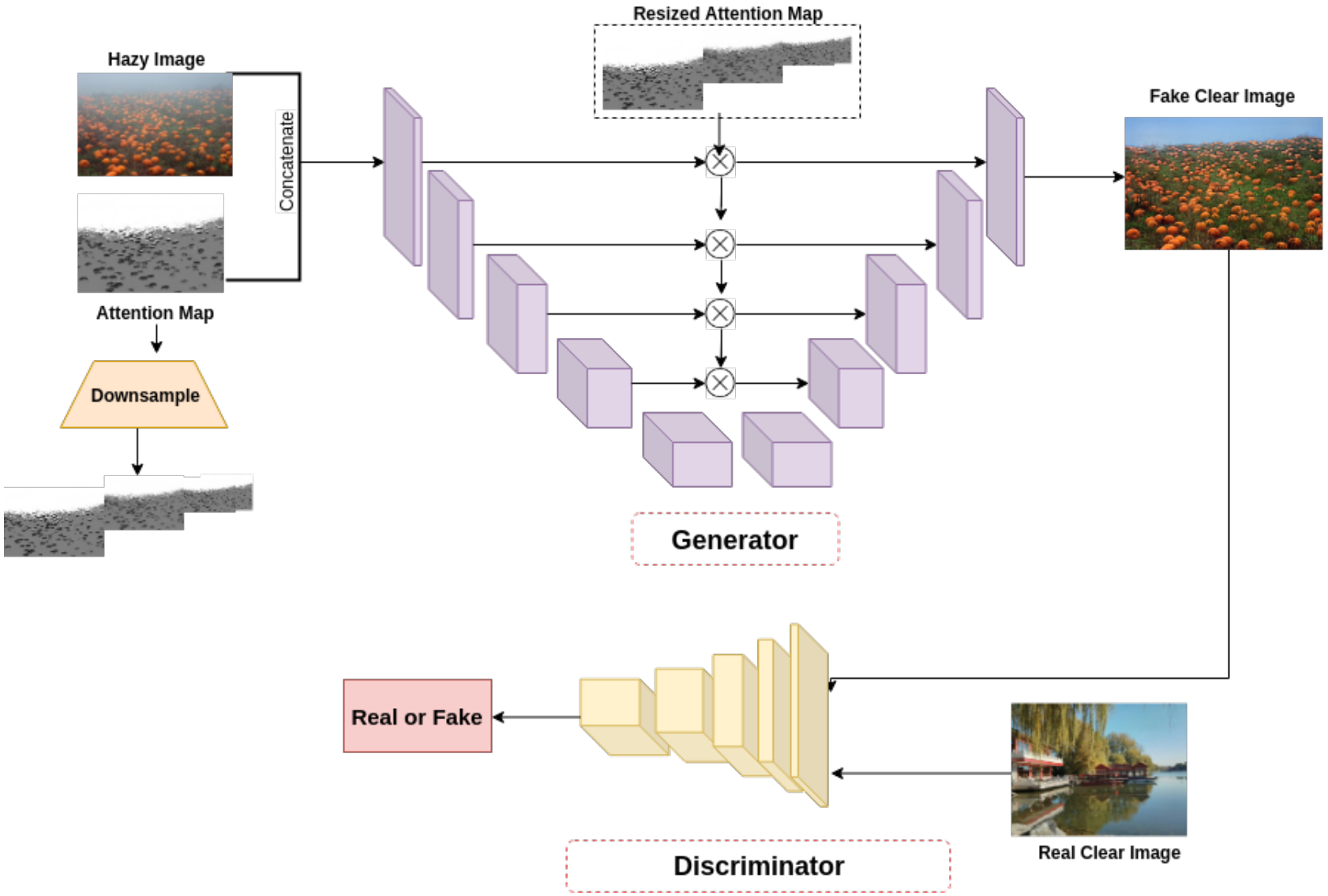

Figure 3.1: This figure represents an overview of UD-GAN architecture explaining the flow through a sample hazy image. In the pre-processing step, attention-maps are created for a sample hazy image, followed by passing both the original image and the resized attention maps to the generator. The output generated by the generator is passed to the discriminator for real/fake test. Note: this image doesn't show the calculation of perceptual loss which is also a part of the loss function.

3.1.1 Hue-Disparity based Attention

Hue-Disparity between an input image $I(x)$and it's semi-inverse image $I_{si}(x)$ is an indicator of haze level at a given location in an image [41] semi-inverse image is defined using below equation where $c \in (r, g, b)$

$$I_{si}(x) = max[I^c(x), 1 - I^c(x)]$$

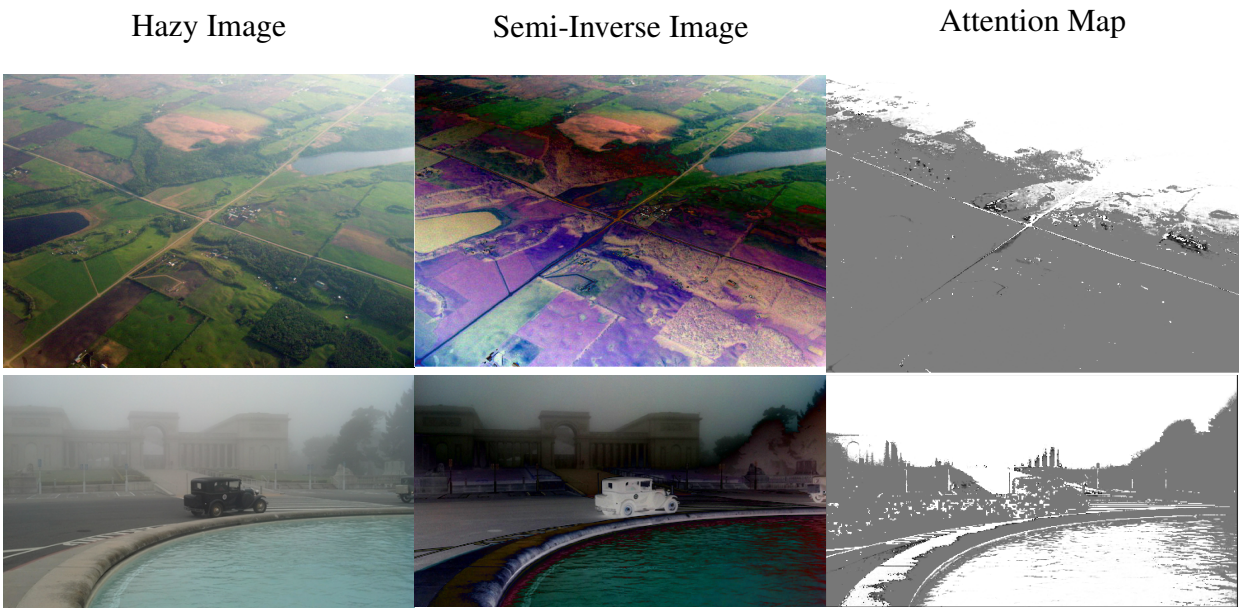

Figure 3.2: Hue Disparity Based Attention Map. The first column represents the input hazy image, the middle one is the semi-inverse image where the pixels with the blue/purple color represents the haze-free pixels. The final column represents the attention map used by the generator for the corresponding input hazy image, where brighter (whitish) pixels represents the hazy image area where we want our generator to specifically work on.

Hue-disparity feature is formally defined as below where superscript h represents the hue channel in HSV space:

$$H(x) = |I^h_{si}(x) - I^h(x)|$$

As described in Fig. 3.2, image patches with higher haze density have higher (brighter) values in the corresponding attention map.

- For haze-free regions, there will be at least one channel in the original image with small values, that value will be replaced by the semi-inverse operation, leading to a large change in hue values. In semi-inverse images, these haze-free regions will have dark blue or purple color as shown in Fig. 3.2.

- On the contrary, for hazy areas in the input image, all channels will have high value, then the semi-inverse operation will return the same values. Hence, there will be no visual observable difference for hazy images.

3.1.2 Illumination-based Attention

Besides the hue-disparity attention mechanism, we also tried using the illumination channel I of the input image for creating an attention map. Since human eyes don't perceive RGB colors uniformly, we use the following standard formula to get the brightness of the image

$$I = 0.299 * R + 0.587 * G + 0.114 * B \tag{3.1}$$

We use $(1 - I)$ as our attentional map. Figure 3.3 represents the hazy images and their corresponding illumination based attention maps. Although this attention-map isn't the direct representation of the spatially varying haze densities of the input image, we found in our experiments that network using such attention mechanism does generate good dehazed results.

3.2 Adversarial Loss

UD-GAN's goal is to learn the underlying mapping between hazy (source) domain and clear (target) domain. To achieve this, our network contains a discriminator D and a generator G. For calculating the adversarial loss, we use a relativistic discriminator [42], which enforces the property that training the generator should not only increase the probability that fake data is real but also decrease the probability that real data is real. We can do this by making the discriminator relativistic, i.e D depends on both real and fake data. The standard relativistic discriminator function can be described as below:

Figure 3.3: Illumination Based Attention Map. The first column represents the input hazy image, the second column uses Eq.3.1 to generate a gray image representing its brightness. The last column is inverse of the illumination channel and it will be used as an attention map input to the generator.

$$D_{Ra}(x_r, x_f) = \sigma(C(x_r) - \mathbb{E}_{x_f \sim P_{fake}}[C(x_f)]) \tag{3.2}$$

$$D_{Ra}(x_f, x_r) = \sigma(C(x_f) - \mathbb{E}_{x_r \sim P_{real}}[C(x_r)]) \tag{3.3}$$

We use x_r and x_f to represent the real hazy and fake hazy images respectively, and C represents the discriminator network. We apply the relativistic property to the least-squares GAN (LSGAN) [43] and use it as adversarial loss for our generator and discriminator network.

3.3 Perceptual Consistency Loss

The adversarial loss only penalizes the network when the generated output image doesn't match the characteristics of clear images. It doesn't ensure that contextual details of the input images are preserved in the generated output. Therefore, inspired by [44], we use Perceptual Loss, L_P, to force perceptual similarity between the image produced by the generator and the input hazy image.

Formally, it can be defined as:

$$L(I^H) = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} ||\phi_{i,j}(I^H) - \phi_{i,j}(G(I^H))||^2 \tag{3.4}$$

$$L(I^H) = L_P \tag{3.5}$$

where I^H represents the input hazy image and $G(I^L)$ denotes the image generated by the generator network. $\phi_{i,j}$ represents the feature maps extracted from a VGG-16 model pre-trained on Image Net [1]. i and j denotes the i-th max pooling layer and j-th convolutional layer respectively. We experimented by extracting the features from both low-level ($conv_{2,2}, conv_{3,3}$) and high-level ($conv_{5,1}$) layers of VGG-16 network.

3.4 Overall Loss Function of UD-GAN

The final UD-GAN loss is composed of a linear combination of the three aforementioned losses:

$$L_{total} = \lambda_1 L_{RaGan} + \lambda_2 L_p \tag{3.6}$$

where λ_i represents the contribution of a given loss function to the total loss.

[1] Using VGG-16 pre-trained weights provided by this Github repo

4. EXPERIMENTS AND RESULTS

This section describes the training and testing dataset, followed by our experimental setup for training the dehazing model. Finally, we compare the results of our proposed approach with the existing dehazing methods

4.1 Dataset

We use the *RESIDE* (REalistic Single Image DEhazing) dataset [2] for training and evaluating the performance of our network. It contains both indoor and outdoor training images.

4.1.1 Overview of RESIDE

This dataset is further divided into five subsets, each serving different training and evaluation purpose.

i. *ITS* (Indoor Training Set) with 13,900 *synthetic indoor* images from NYU2 [45] and Middlebury stereo datasets [46]

ii. *OTS* (Outdoor Training Set) contains 2,057 paired clear images and the corrresponding synthetic hazy images

iii. *HSTS* (Hybrid Subjective Testing Test) has mix of synthetic and real-world hazy images, each containing 10 images

iv. *SOTS* (Synthetic Objective Testing Sets) includes both indoor and outdoor sections called *SOTS-indoor* and *SOTS-outdoor*, each containing 500 images

v. *RTTS* (Real-World Task-Driven Testing Set) provides 4,332 real-world images obtained from the web. Each image in *RTTS* is annotated with object bounding boxes and 5 categories - person, bicycle, bus, car, or motorbike

vi. *RTTS - Unannotated* contains 4,807 unannotated real-world hazy images

All the synthetic hazy images are created by first collecting clean haze-free images along with their depth meta-data, followed by using various combinations of the A and β parameters in the physical model (2.1)

4.1.2 Unpaired Dataset

Dehazing task can be construed as an image-to-image translation task where the source domain is a hazy image and target domain is a clear image.

4.1.2.1 Training Data

To create the datasets for source and target domains, we further merge or split the categories in RESIDE dataset to suit our purpose.

i. *Hazy Domain* (Source) contains RTTS - Unannotated [vi] images

ii. *Clear Domain* (Target) combines clear images from ITS & OTS [i, ii]

4.1.2.2 Evaluation Dataset

We use *SOTS-outdoor*, *HSTS-synthetic* and *RTTS-annotated* datasets of hazy images to evaluate the performance of the model qualitatively and quantitatively.

4.2 Training Details

We implemented this method in PyTorch and it uses the official CycleGAN [13] code [1] as the base and builds on top of it. In terms of hardware, all the training is done using two Tesla K80 Nvidia's GPUs. To enrich the training data set, we perform data augmentation on ITS, OTS, and RTTS - Unannotated datasets. Each image is randomly cropped to the size of 256 X 256 and the cropped image can be further flipped horizontally or vertically. During the training, we used Adam optimizer with learning rate 10^{-4} and batch size 30. The network weights are initialized using Gaussian initialization with zero mean and variance of 0.02.

[1] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

4.3 VGG-16 Features for L_p

As explained in the methodology section, we are using perceptual loss L_p to force the generated output to be similar to the input hazy image. The effect of L_p on the final dehazed output varies with the chosen VGG-16 layer for feature extraction. In this section, we perform experiments using three different VGG-16 conv layers – $conv_{2,2}, conv_{3,3}, conv_{5,1}$, and analyze if the choice of VGG-16 feature layer impacts the performance of the UD-GAN model in any way. Unless otherwise mentioned, all the results in this sub-section use Illumination-Based Attention.

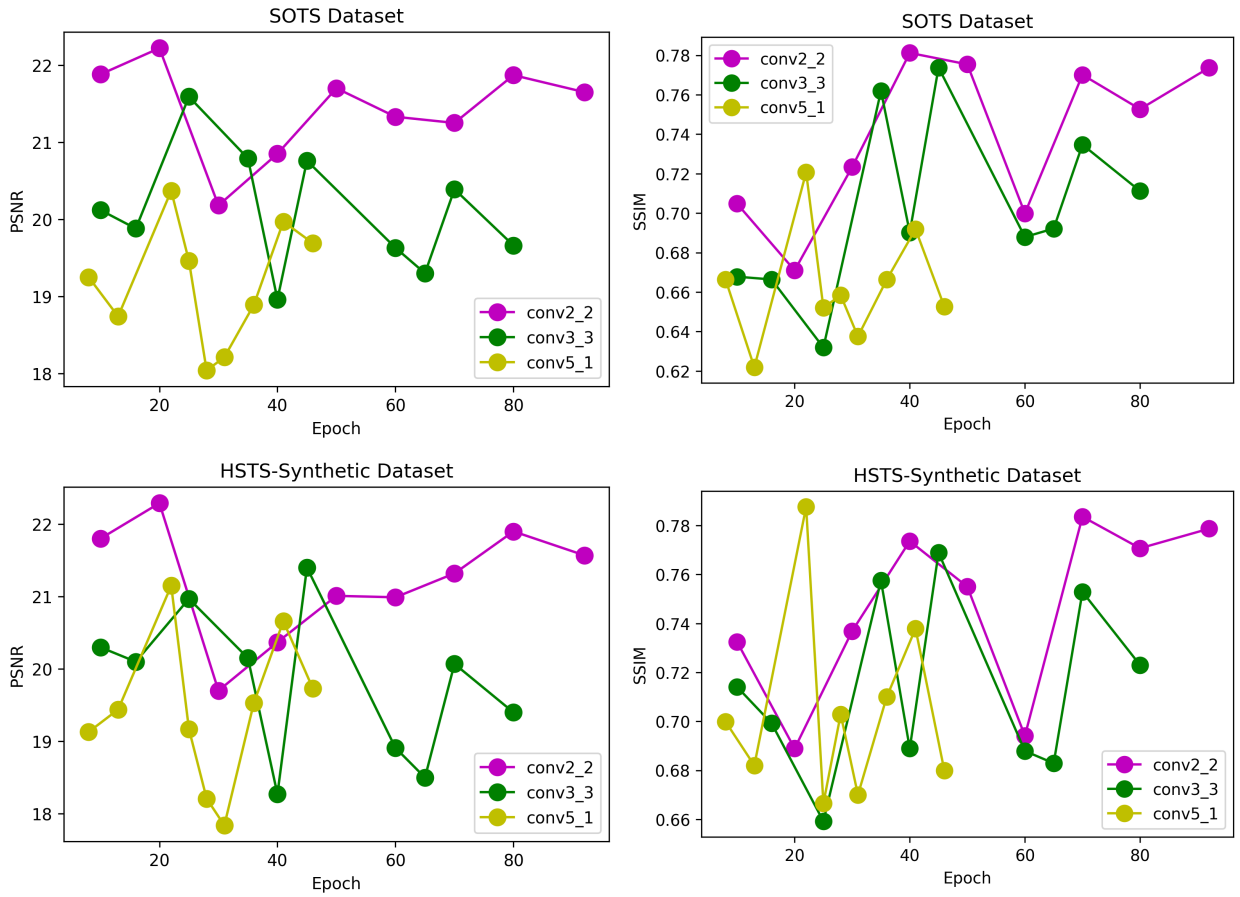

Figure 4.1: This figure contains four scatter plots presenting our analysis on the effect of the choice of VGG-16 feature layer for L_p on the full-reference metrics - PSNR and SSIM. The first row represents the PSNR and SSIM plots of dehazed images in the *SOTS* dataset. Similarly, the second row, represents the results evaluated on *HSTS-Synthetic* dataset

Fig. 4.1 summarises our findings in this regard. *conv5_1* (yellow color) performs the worst out of the three feature layers for both HSTS and SOTS dataset. In fact, we only trained the *conv5_1* for 44 epochs because the visual results (Fig. 4.2) were also poor in quality and we didn't see the benefit of training it for more epochs. This is why, in all the four plots, yellow line for *conv5_1* can be seen extending till 45^{th} epoch only. Between *conv2_2* (magenta)and *conv3_3* (green), *conv2_2* seems to perform better and consistently gives better PSNR and SSIM results throughout the epoch training.

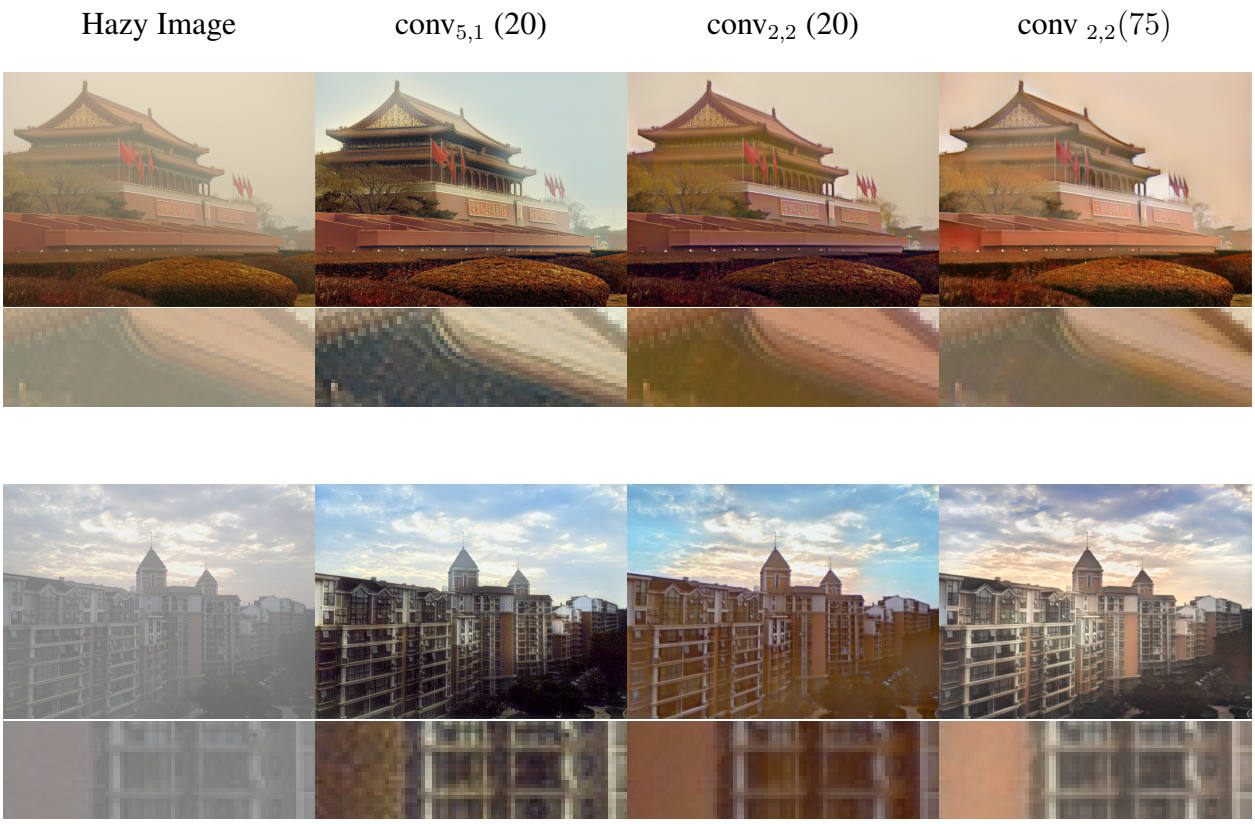

Figure 4.2: Visual comparison of dehazed images obtained from models trained using *conv5_1* and *conv2_2* for calculating feature loss. Row 2 and 4 zooms into the specific parts of the dehazed images to help in visualizing the semantic details of the images. The number in parenthesis, next to VGG-16 feature name represents the number of epochs trained to obtain the image.

It is well known that the PSNR and SSIM values aren't always representative of the quality of the restored images, therefore, we will visually inspect the images generated by using *conv5_1*

and *conv2_2* VGG-16 feature. In Fig. 4.2 we don't include the visual results for *conv3_3* since the visual results are very similar to *conv2_2*. We observe that dehazed image obtained by using *conv5_1* model has patchy artifacts whereas the dehazed results from *conv2_2* model trained for 20 and 75 epochs are free from such artifacts. Based on both the visual and quantitative results, in our final model, we chose *conv2_2* for calculating the perceptual loss, L_P.

4.4 Evaluating the Attention Layer

We evaluate the performance of UD-GAN using two different attention layers. For the sake of representation, we will refer our model using illumination-based attentive layer as *UD-GAN*$_I$, and model using hue-disparity based attentive layer as *UD-GAN*$_{HD}$ where subscript *I* refers to the **I**llumination and *HD* refers to **H**ue **D**isparity.

- On the one hand, *UD-GAN*$_I$ takes at least 50 epochs to create a clear dehazed image, but on the other hand, *UD-GAN*$_{HD}$ generates clear output with a model trained for less than 20 epochs. Intuitively, this is because the hue-disparity based attention highlights the hazy image regions which help the fast learning of the generator network for creating dehazed images.

- We ran multiple experiments using both attentive layers and observed that PSNR and SSIM values for *UD-GAN*$_{HD}$ values are always less than that of *UD-GAN*$_I$. We think this is because the skip connection of illumination-based attentive layers between first and last layers of the generator enforces the dehazed image to preserve the same relative brightness and therefore leading to better PSNR and SSIM values.

- Hue-Disparity based attention map, if used as it is, sometimes leads to strong artifacts in the dehazed images as shown in Fig. 4.3. Therefore, we do gaussian filtering on the attention map before providing it as an input to the generator.

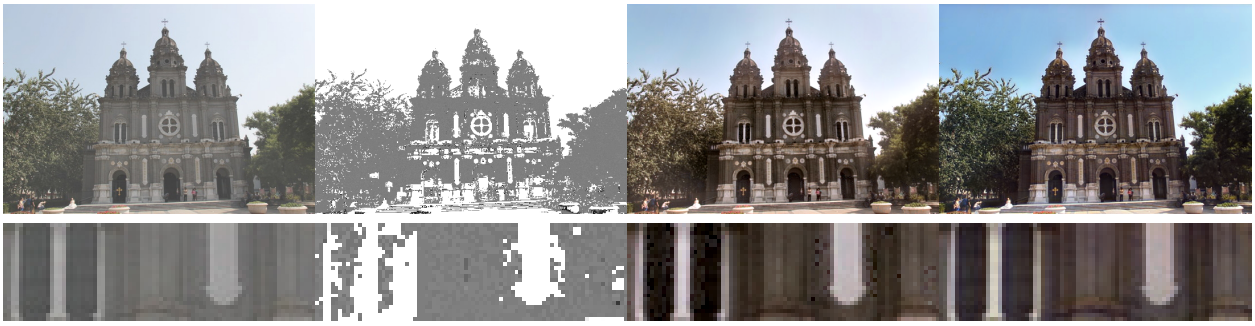

Figure 4.3: Gaussian blurring in UD-GAN$_I$. The first column contains the hazy image and the second column represents hue-disparity attention-map corresponding to the hazy image. Column 3 and 4 presents two versions of *UD-GAN$_I$* model – former version directly feeds the attention map to the UD-GAN generator, whereas later version first performs Gaussian blurring of the attention map. We can observe that the dehazed image generated from *UD-GAN$_I$* version which doesn't use Gaussian blurring suffers from serious artifacts

4.5 Quantitative Evaluation

In this section, we use three criterias to evaluate the performance of UD-GAN and compare it with existing state-of the-art dehazing methods [2] – DCP [8], CAP [23], DehazeNet [1], AOD-Net [10]. For the purpose of evaluation, we use *SOTS*, *HSTS*, and *RTTS-Annotated* datasets.

- ***Full Reference Metrics*** – Table 4.1 compares both **UD-GAN$_I$** and **UD-GAN$_{HD}$** against the state-of-the-art dehazing models[3]. These results are evaluated on the images where ground-truth images are available. Although PSNR values for **UD-GAN$_I$** are more comparable to the state-of-the-art methods than **UD-GAN$_{HD}$**, SSIM values are low for both versions of the UD-GAN models.

- ***No Reference Metrics***– Full-reference metrics require a ground-truth clear image against which a generated image will be compared. However, we lack that flexibility in most of the real-world settings. In this section, we use two popular no-reference image quality assessment (IQA) models [4]:

[2]The state-of-the-art results are from the dehazing *RESIDE* benchmark paper [2]

[3]The top-3 performances are highlighted using red, cyan and blue, respectively

[4]These metrics are calculated using the official implementation shared by the authors of BLIINDS-II and SSEQ (link)

i. *BLIINDS-II*– Blind Image Integrity Notator using DCT statistics [47]

ii. *SSEQ*– Spatial-Spectral Entropy-based Quality [48]

Different from PSNR and SSIM (higher values are better), both of the no-reference metrics values range from 0 (best) - 100 (worst). However, to compare our results against the *RESIDE* benchmark, we follow their suit and complement (reverse) these scores to make them consistent with the full-reference metrics.

Table 4.2 compares SSEQ and BLIINDS-II scores of UD-GAN against other dehazing methods[3]. Our model has better results than other state-of-the-art methods when using no-reference metrics. This implies that while UD-GAN doesn't perform well in converting the hazy image to its actual clear content, it does improve the overall quality of the image as proved by the no-reference metrics results.

- ***Performance of Object Detection on Dehazed Images*** – Another way we evaluated the performance of our approach is by performing object detection on the dehazed images. We use Faster-RCNN [49] model for comparing the performance with other state-of-the-art-methods. Table 4.3 lists mAP scores of object detection.

4.6 Qualitative Evaluation

Figure 4.4 shows that UD-GAN can recover a visually pleasing clear image from a hazy one. The table also compares our model output with other state-of-the-art methods. Figure 4.5 provides more examples of our model's performance under light and heavy haze conditions. As compared to light haze images, heavy haze images require more number of training epochs.

There are multiple cases where the generator fails to generate a visually pleasing image. Figure 4.6 shares few such examples. In all the examples shared, we can see that generator was able to recover the texture details under the haze very accurately, however the generated output has a lot of blue artifacts.

4.7 Stability of GAN Training

Achieving stability while training GANs is a difficult task. We observed in our experiments that with same number of epochs training, UD-GAN$_{HD}$ generates visually better dehazed results than UD-GAN$_{I}$. However, training of UD-GAN$_{HD}$ is highly unstable as compared to that of UD-GAN$_{I}$; the results generated in two consecutive epochs can vary hugely in visual quality. We experimented with RTTS dataset and observed that if we train the generator for a few more epochs, the visual quality of generated images can become better or worse.

	DCP [23]	CAP [23]	Dehaze-Net [1]	AOD-Net [10]	**UD-GAN**$_I$	**UD-GAN**$_{HD}$
SOTS - Synthetic						
PSNR	16.62	19.05	21.14	19.06	21.25	19.61
SSIM	0.8179	0.8364	0.8472	0.8504	0.7701	0.6931
HSTS-Synthetic						
PSNR	14.84	21.53	24.48	20.55	21.32	19.27
SSIM	0.7609	0.8726	0.9153	0.8973	0.7836	0.6894

Table 4.1: Full-Reference Evaluation Results on Dehazed Images from SOTS and HSTS-Real Datasets. Results of DCP, CAP, Dehaze-Net, AOD-Net are reprinted from [2] for the purposes of comparison.

	DCP [23]	CAP [23]	Dehaze-Net [1]	AOD-Net [10]	**UD-GAN**$_I$	**UD-GAN**$_{HD}$
SOTS - Synthetic						
SSEQ	64.94	64.69	65.46	67.65	82.47	79.87
BLIINDS-II	74.71	73.41	71.71	79.02	85.93	94.15
HSTS-Synthetic						
SSEQ	86.15	85.32	86.01	86.75	83.10	81.71
BLIINDS-II	90.70	85.75	87.15	87.5	89.85	95.55
HSTS-Real World						
SSEQ	68.65	67.67	68.34	70.05	82.92	85.44
BLIINDS-II	69.35	63.55	60.35	74.75	90.95	94.90

Table 4.2: No-Reference Evaluation Results on Dehazed Images from SOTS, HSTS-Real and HSTS-Synthetic Dataset. Results of DCP, CAP, Dehaze-Net, AOD-Net are reprinted from [2] for the purposes of comparison.

	DCP [23]	CAP [23]	Dehaze-Net [1]	AOD-Net [10]	**UD-GAN**$_I$	**UD-GAN**$_{HD}$
mAP	40.58	39.63	40.54	37.47	44.08	45.62
Person	61.54	61.29	61.40	61.22	63.15	66.15
Bicycle	40.77	40.48	40.68	40.33	49.63	50.25
Car	42.15	41.52	41.74	35.13	46.99	46.92
Bus	24.18	24.74	25.20	20.56	23.23	24.70
Motorbike	34.25	30.10	33.70	30.09	37.74	40.07

Table 4.3: Detection Results on dehazed images obtained using UD-GAN$_I$ and UD-GAN$_{HD}$. Results of DCP, CAP, Dehaze-Net, AOD-Net are reprinted from [2] for the purposes of comparison.

Figure 4.4: Qualitative results of single image dehazing on **real-world** hazy images.

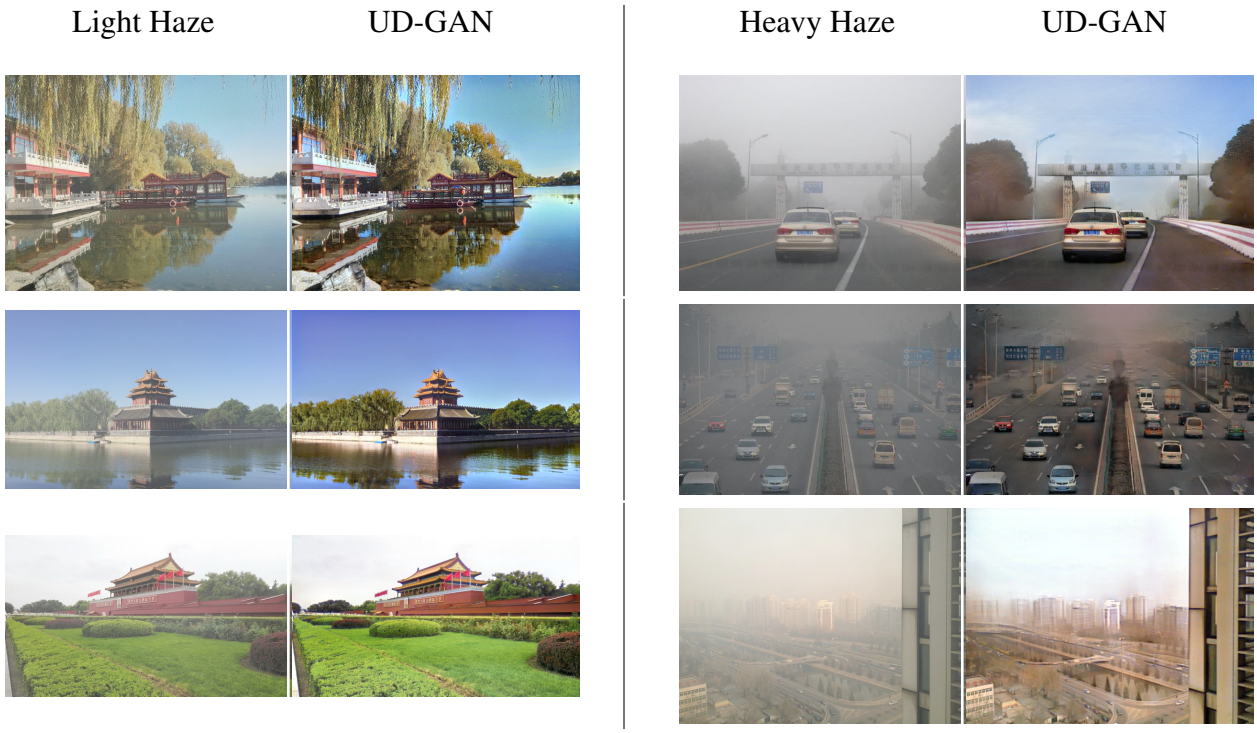

Figure 4.5: Left column: light hazy images, Right column: heavy hazy images

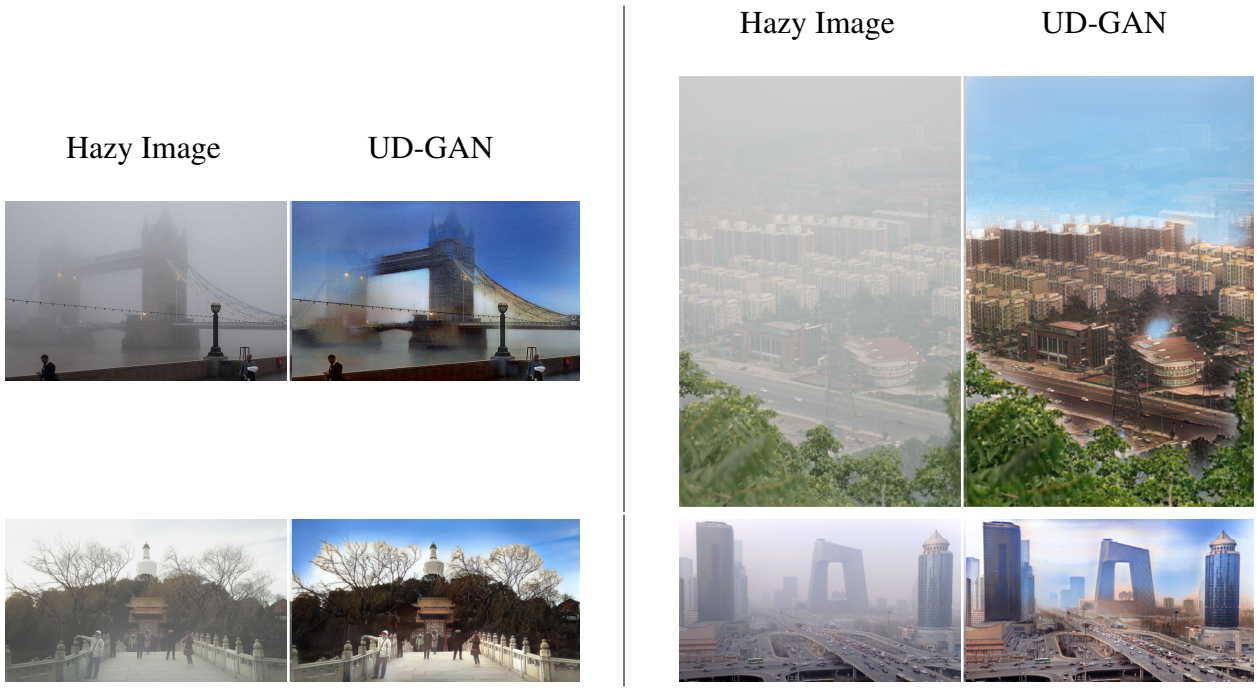

Figure 4.6: Failure Cases: Blue artifacts appear in the output images specifically under heavy haze conditions

5. CONCLUSION AND FUTURE WORK

The proposed UD-GAN model in this thesis work introduces a novel way to perform dehazing in an unsupervised manner. Adding attention to the generator network as well as the perceptual loss helps in regularizing the adversarial training process. Moreover, our model is independent of the physical scattering model and can overcome some of its common failure scenarios. Experimental results show that our model beats the state-of-the-art methods in both no-reference and object detection evaluation criteria. This architecture can be extended to other image restoration methods as well - such as image deraining, denoising and others. For applying it to another problem, we only need to prepare an attention map corresponding to the problem.

In future work, we would like to tweak the UD-GAN's architecture to avoid the blue artifacts in the final dehazed output and to achieve good results for SSIM and PSNR. Moreover, current training of GANs is unstable as the outputs vary widely as we train for more epochs. We can introduce changes in UD-GAN methodology to make the training more stable.

REFERENCES

[1] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *CoRR*, vol. abs/1601.07661, 2016.

[2] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2019.

[3] Wikipedia contributors, "Haze — Wikipedia, the free encyclopedia." `https://en.wikipedia.org/w/index.php?title=Haze&oldid=878378777`, 2019. [Online; accessed 14-January-2019].

[4] "How autonomous vehicles will navigate bad weather remains foggy," *Forbes*, Nov 2016.

[5] "Can fog affect the systems that detect cars at red lights?," *Oregonlive*, Dec 2015.

[6] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[7] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, pp. 72:1–72:9, Aug. 2008.

[8] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 2341–2353, Dec 2011.

[9] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *European Conference on Computer Vision*, 2016.

[10] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "An all-in-one network for dehazing and beyond," *CoRR*, vol. abs/1707.06543, 2017.

[11] D. Engin, A. Genç, and H. K. Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," *CoRR*, vol. abs/1805.05308, 2018.

[12] A. Golts, D. Freedman, and M. Elad, "Unsupervised single image dehazing using dark channel prior loss," *CoRR*, vol. abs/1812.07051, 2018.

[13] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *CoRR*, vol. abs/1703.10593, 2017.

[14] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *arxiv*, 2016.

[15] E. J. McCartney, *Optics of the atmosphere: Scattering by molecules and particles*. 1976.

[16] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, vol. 1, pp. 598–605 vol.1, June 2000.

[17] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 713–724, June 2003.

[18] J. A. Stark, "Adaptive image contrast enhancement using generalizations of histogram equalization," *IEEE Transactions on Image Processing*, vol. 9, pp. 889–896, May 2000.

[19] T. K. Kim, J. K. Paik, and B. S. Kang, "Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering," *IEEE Transactions on Consumer Electronics*, vol. 44, pp. 82–87, Feb 1998.

[20] T. Zhang, C. Shao, and X. Wang, "Atmospheric scattering-based multiple images fog removal," in *2011 4th International Congress on Image and Signal Processing*, vol. 1, pp. 108–112, Oct 2011.

[21] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–I, Dec 2001.

[22] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski, "Deep photo: Model-based photograph enhancement and viewing," *ACM Trans. Graph.*, vol. 27, pp. 116:1–116:10, Dec. 2008.

[23] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on Image Processing*, vol. 24, pp. 3522–3533, Nov 2015.

[24] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *2013 IEEE International Conference on Computer Vision*, pp. 617–624, Dec 2013.

[25] Z. Li, P. Tan, R. T. Tan, D. Zou, S. Z. Zhou, and L. Cheong, "Simultaneous video defogging and stereo reconstruction," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4988–4997, June 2015.

[26] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *CVPR*, 2018.

[27] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated Fusion Network for Single Image Dehazing," 2018.

[28] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, 2014.

[29] E. L. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," *CoRR*, vol. abs/1506.05751, 2015.

[30] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *CoRR*, vol. abs/1511.06434, 2015.

[31] Y. Du and X. Li, "Perceptually Optimized Generative Adversarial Network for Single Image Dehazing," pp. 1–10, 2018.

[32] H. Zhang, V. Sindagi, and V. M. Patel, "Joint Transmission Map Estimation and Dehazing using Deep Networks," no. 3, pp. 1–11, 2017.

[33] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," *CoRR*, vol. abs/1711.07064, 2017.

[34] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CoRR*, vol. abs/1609.04802, 2016.

[35] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," *CoRR*, vol. abs/1711.10098, 2017.

[36] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," *CoRR*, vol. abs/1703.05192, 2017.

[37] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," *CoRR*, vol. abs/1704.02510, 2017.

[38] T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," *CoRR*, vol. abs/1711.11585, 2017.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.

[40] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," *arXiv preprint arXiv:1805.08318*, 2018.

[41] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert, "A fast semi-inverse approach to detect and remove the haze from a single image," in *ACCV*, 2010.

[42] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard GAN," *CoRR*, vol. abs/1807.00734, 2018.

[43] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, and Z. Wang, "Multi-class generative adversarial networks with the L2 loss function," *CoRR*, vol. abs/1611.04076, 2016.

[44] J. Johnson, A. Alahi, and F. Li, "Perceptual losses for real-time style transfer and super-resolution," *CoRR*, vol. abs/1603.08155, 2016.

[45] P. K. Nathan Silberman, Derek Hoiem and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *ECCV*, 2012.

[46] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 1, pp. I–I, June 2003.

[47] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Transactions on Image Processing*, vol. 21, pp. 3339–3352, Aug 2012.

[48] L. Liu, B. Liu, H. Huang, and A. Bovik, "No-reference image quality assessment based on spatial and spectral entropy," *Signal Processing: Image Communication*, vol. 29, 09 2014.

[49] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015.