Image Dehazing Using Deep Learning

A Project Report submitted to Manipal Academy of Higher Education in partial fulfilment of the requirements for the award of the degree of

of

BACHELOR OF TECHNOLOGY

in

Mechatronics Engineering

submitted by

ABU BAKR SIDDIQ (170929098)

RAZEEM AKTHAR (170929194)

Under the Guidance of

DR. KSHETRIMAYUM LOCHAN

Assistant Professor

Department of Mechatronics

Manipal Institute of Technology, Manipal

DEPARTMENT OF MECHATRONICS

MANIPAL INSTITUTE OF TECHNOLOGY MANIPAL - 576104

JULY, 2021

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

DEPARTMENT OF MECHATRONICS ENGINEERING

CERTIFICATE

This is to certify that the project titled **Image Dehazing using Deep Learning** is carried out by Abu Bakr Siddiq (170929098) and Razeem Akhtar (170929194) during February-May, 2021 and the project report is submitted to the Department of Mechatronics Engineering as part of requirement of B.Tech. (Mechatronics Engineering) project work evaluation.

DR. KSHETRIMAYUM LOCHAN
Assistant Professor
Department of Mechatronics
MIT Manipal

DR. MUNENDRA SINGH
Assistant Professor
Department of Mechatronics
MIT Manipal

DR. CHANDRASHEKHAR BHAT

Date: 1st August 2021

Place: MANIPAL

ACKNOWLEDGEMENTS

With immense honor, We submit our project thesis in the partial fulfillment of B.Tech Mechatronics Engineering. We hereby extend our sincere gratitude to each and every individual for their kind support and cooperation during this wonderful journey of our project.

We consider it our privilege to express voice of gratitude to all those who guided and inspired us in the completion of project work.

We are very grateful to Dr. Munendra Singh, our project mentor who helped us identify a domain to work on and shared his wealth of experience and knowledge with us. The project was mainly possible because of his insights and guidance, we dont know if the word thank you will do justice to the efforts he has put behind this project, but thank you sir.

We also take this opportunity to express deep sense of greatness to our project coordinator Dr. Kshetrimayum Lochan, Dept. of Mechatronics, for her valuable advice, expert guidance and constant encouragement at all stages of the project.

We are grateful to Dr. Chandrasekhar Bhat, HOD, Department of Mechatronics for his encouragement and support.

Lastly, we would like to thank all our family members who throughout our life have been giant pillars of support.

(ABU BAKR SIDDIQ & RAZEEM AKHTAR)

ABSTRACT

Many modern-day applications require a clear image to analyse and extract information, but generally, the presence of a turbid medium hinders their performance. This has given rise to the field of Image dehazing. The dehazing methods can be extended to applications such as image enhancement, denoising, color correction etc. The results can be also extended to depth correction models of an image which would aid in improving the performance of autonomous systems and make systems, like self-driving cars achieve level 5 automation thereby making it function in most weather conditions.

As a part of our study, we looked at the SOTA models and implemented them on TensorFlow framework. After implementing them we performed a comparative study to identify best features of each model. We also compared the models based on no reference metrics like naturalness factor, contrast quality as well as on haze specific metric like FADE.

We then developed a new dataset which was targeted to roads.The dataset was based on the atmospheric scattering model, taking the clear image and LIDAR sensor generated image that gives information about depth. We used this to generate hazy images from KITTI dataset. We had generated around 6000 images which constituted our training and validation set and 77 images which made up our test set.

We developed a novel network for dehazing called the ODD Net (Outdoor Depth based Dehazing). We firstly broke the problem into sub models to find atmospheric light and transmission map and combined these two models using atmospheric scattering model to generate the Dehazed image. The A-net was developed with help of LCA net and DCP paper and the T net was developed taking help of densedepth network.

Packages used: *TensorFlow, PyTorch, matplotlib, NumPy, OpenCV, ONNX*. Software's used: *Jupyter Lab, Google colab, MATLAB.*

Keywords: *Image dehazing, image denoising, Depth correction, KITTI, DenseDepth*

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

SOTA	State of the Art
CNN	Convolutional Neural Network
PSNR	Peak Signal To Noise Ratio
SSIM	Structural Similarity Index Measure
MS SSIM	Multi-Scale Structural Similarity Index Measure
DCP	Dark Channel Prior
HSV	Hue Saturation Value
RESIDE	Realistic Single Image DEhazing
BLIINDS	BLind Image Integrity Notator
MSE	Mean Square Error
MAE	Mean Absolute Error
LCA	Light Convolutional Autoencoder
GMAN	Generic Model-Agnostic Convolutional Neural Network
GCA	Gated Context Aggregation Network
FFA	Feature Fusion Attention
LOL	Low light paired dataset
NSS	Natural Scene Statistic
MVG	Multi-Variate Gaussian
NIQE	Naturalness Image Quality Evaluator
FADE	Fog Aware Density Evaluator
CEIQ	Contrast-changed Image Quality measure
ODD	Outdoor Depth-based Dehazing Network

NOMENCLATURE

$d(x)$	distance from scene point (Depth Map)
$\theta_0, \theta_1, \theta 2$	Coefficients in Color Attention Prior Model
$v(x)$	Pixel value in value channel in HSV
$s(x)$	Pixel value in Saturation channel in HSV
$\epsilon(x)$	Random Variable
P_i	clear image patch (Synthesizer model)
P_j	hazy image patch (Synthesizer model)
t	transmission map
A	Atmospheric light
$I(x)$	Observed Intensity (Hazy Image)
$J(x)$	Scene Radiance (Dehazed Image)
$\beta(x)$	Scattering coefficient

CHAPTER 1

INTRODUCTION

1.1 *General Discussion*

The quality of an image is critical to performance of many algorithms as well as human judgement. Today with the rise of air pollution there has been a rapid increase in particulate matter present around us, which causes increased atmospheric scattering thereby hindering the ability to see clearly. Single image dehazing models propose a way to combat this problem by reducing the haze density to a certain extent such that the depth and aesthetic information in the dehazed image is as close to a clear image retaining the naturalness and structure of image. Haze if entirely removed not only makes the image unnatural but also reduces the depth information that a clear image has, as there is always a need for certain amount of haze so that human eye can perceive depth. Most dehazing assume the atmospheric scattering model as the base model, so we also start with the same model. As described in many works the problem of dehazing is a non-trivial one especially when we try to improve the image quality with only one image which in most cases fails to provide enough information.

1.2 *Present Day Scenario*

The increasing research in image dehazing is well highlighted with the growing research in the field. The models for image dehazing proposed fall under three categories image enhancement, image fusion and image restoration. The most popular model for convention dehazing falls under image restoration which are based on a prior about hazy images, (1) assumes that Dark channel for hazy images is lighter that dark channel for haze free images. This prior based model is one of the ground breaking ones as it formed the basis of future work in the field. There have been other approaches based on polarization and filtering.

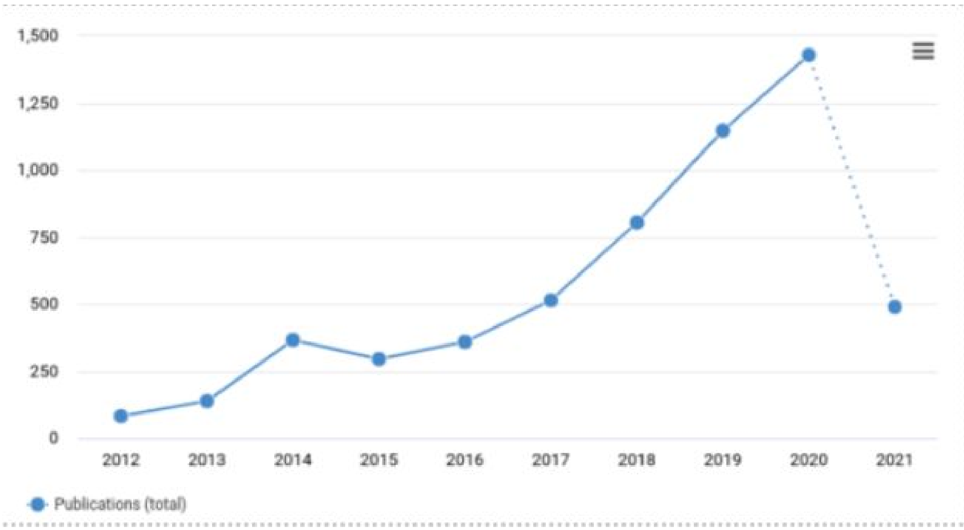


Figure 1.1: Growing Trend in Image dehazing

With the availability of better resources there was a boom in deep learning models which fall into the image enhancement category. These models could give satisfactory results but failed to consider the features specific to haze and therefore they were overshadowed by image restoration based deep network which considered the properties of haze. Some of the current SOTA models like DehazeNet, gated context attention network and feature fusion attention networks which incorporated haze specific features have given amazing performance with respect to metrics such as PSNR and SSIM.

1.3 *Motivation*

There have been catastrophic traffic accidents throughout the world that have caused a loss of life, which identify bad weather and limited visibility as their underlying cause according to current research and data analysis. The primary aim of this project is to create and train a model based on the new data set using Kitti data sets, which provide a range of methods and approaches to dehaze/defog a picture with as much precision as feasible.

Early in 2020 some of the statistics about air pollution by IQAir caught our attention. According to the report 22 out of 30 most polluted cities are in India. There have been many efforts to combat this problem, but we wanted to develop a solution which would help people in the short term till long term solution take effect.

Figure 1.2: Fog on roads of Delhi Hindustan Times

The image perfectly sums up the problem we are trying to address. There has already been a lot of work done in this area and the literature review section goes in depth, analyzing many popular models and highlighting their shortcoming. Most models assume a homogeneous atmospheric light model and we wanted to test this assumption in our work. We also wanted to strike a balance between computation resource, time and quality of output.

We looked at some of the popular websites for comparing the current models and found that there was no uniform comparison report available for comparison, websites such as paperswithcode compared models on the basis of PSNR and SSIM which is the most basic criteria. So, we through our work have tried to compare models based on haze specific metrics discussed in the methodology section (6). We also have tried to adopt best feature and metrics from field of image processing, low light enhancement and image dehazing. So the primary motivation is for self - driving cars and secondary is a solution for dehazing in a hazy/polluted area.

1.4 *Primary objective*

Research on self-driving cars has gained considerable attention in recent years, accelerating the development and adaption of these vehicles in the world today. However, before these cars become completely autonomous on public highways there are still big problems to resolve. These self-drive trucks have been trained to recognize things such as streets, stop signs and people using image recognition algorithms. These objects are, however, more difficult to identify under severe weather circumstances. As such, the research focuses mostly on all the poor weather associated with object recognition and picture segmentation. Imagine driving in a unfavorable weather to reduce your view of the environment, how are you driving without knowing what is to come? This project attempts therefore to tackle these difficulties in order to contribute to unfavorable weather. Image dehazing also plays an essential part in climatology, environment perception, monitoring and protection of wildlife, monitoring systems, object detection and identification.

1.5 *Significance of results*

The model developed on image dehazing can also be extended to image denoising, image demosaising and image de-raining. The algorithm will also have a impact on the field of computer vision in general as this model would improve the accuracy models in the field of image segmentation, object recognition and image enhancement. It is also of paramount significance in climatology, environment perception, monitoring and protection of wildlife, monitoring systems, object detection and identification.

1.6 *Shortcomings of previous papers*

Although the previous model has done a decent job in deconstructing the images without the haze and noise, the difficulty encountered by the model is that it is over-fitting with the data set and not well-suited for creating the transmission map. The SOTA models also have a major drawback since the computing resources needed are quite expensive, which cannot operate unless you have a top of the range resources. We have thus suggested that the present model be computer efficient, as well as has a better transmission map reconstruction, to address the constraints.

CHAPTER 2

LITERATURE REVIEW

The problem of dehazing has been studied extensively and a lot of literature has been published. The problem is a non-trivial one as the works that have been published try to predict many features like atmospheric light, depth maps and scattering coefficient from a pair of images or in some cases a single image. Some of the models try to solve this problem by using suitable priors while others use deep learning models which generally are many layers deep.

In this literature review go through firstly the conventional methods that have formed the basis for many works, then we study the different deep learning models that have been able to attain suitable results in terms of PSNR and SSIM, we also analyze the different datasets, metrics and losses that have assisted us in our project.

2.1 *Conventional approaches*

He et al.(1) proposes a dark channel prior, which consists of minimum intensities over 3 channels over a certain neighborhood. The assumption is that the dark channel for hazy images is brighter than the dark channel for non-haze images because the haze component is generally white, so it adds equal intensities to all the color channels. He et al.(1) also assumes that the hazing is homogeneous and uses 0.1% of brightest pixels in dark channel and compares them and the pixel having highest value in input image is considered as atmospheric light. The paper also suggests using soft matting to refine the transmission map and improve the results, the soft matting function increases complexity and the time required. In further works it has been shown by He et al. (2) that the guided filter if used instead of soft matting, improves the performance of the DCP algorithm.

Zhu et al.(3) is based on the color attention prior based approach. The paper observes that the density of haze increases with the increase in depth, and makes the assumption that depth is positively correlated to the concentration of depth which in turn depends upon the difference between value and saturation of the pixel in Hue Saturation Value (HSV) space. The paper proposes a linear model for modelling the scene depth

$$d(x) = \theta_0 + \theta_1 v(x) + \theta_2 s(x) + \varepsilon(x) \tag{2.1}$$

Where d(x) is depth map of a pixel, v(x) and s(x) are pixel values of value and saturation in HSV format. $\theta_0, \theta_1, \theta_2$ are the coefficients of the model. For dehazed image recovery the scattering coefficient is assumed constant and atmospheric light is assumed in th same way as done in (1).

The above prior based methods are successful to a certain degree in recovering the dehaze image but both suffer from some notable drawbacks. (1) generally over enhances the contrast of image, it also suffers from halo effects. (3) suffers from color distortion and background noise. Both the approaches are based on prior which may generally be useful but may not always hold true. This shortcoming pave way for approaches based on deep learning models.

2.2 *Deep learning approaches*

The major advantage with deep learning approach is that the model can itself come up with suitable priors based on the training methods and datasets.

2.2.1 *Dataset*

Datasets play a critical part in the performance of deep learning models.

For this reason, we looked at Tang et al. (4). The paper's aim is to identify the best features combination for dehazing, the paper also looks at features such as multi-scale dark channel, multi-scale local max contrast, hue disparity and multi-scale local max saturation. The paper observers that dark channel is positively correlated to amount of haze and color saturation is negatively correlated. The most interesting part of the work is that it proposes a way to develop a synthetic hazy image as for a scene it is very difficult to have a haze free and hazy pair. The synthesizer assumes that depth is locally constant. The synthesizer takes a non hazy patch (P_i), a transmission t and atmospheric light A, and outputs the hazy image patch (P_j) according to the equation:

$$P_i = tP_j + (1 - t)A \tag{2.2}$$

Here the transmission map is assumed as a random value between 0.3 and 0.7 thereby helping us generate image with variable amount of haze. For dehazing the paper proposes firstly a method for robust preprocessing based on adaptive atmospheric light, and then this is passed throughout the random forest regressor which is trained on the generated dataset. The results of the paper further go on to show that the synthetic dataset is very effective in training the model.

In deep learning models that are generally trained for hazing of images there is generally a limitation on the quantity of data, which affects the model's performance. B Li. et al. (5) aims at providing a benchmark for single image dehazing. The paper introduces Realistic single image dehazing dataset (RESIDE), which is a large-scale synthetic dataset which originally had 15,000 images and was later updated and the RESIDE: V0 has around 500,000 images focused at both objective and subject evaluations. B Li. et al. (5) further observers that the conventional metrics like PSNR, SSIM fail to satisfy the needs of a dehazing model when testing, so the paper compares some of the dehazing algorithms based on PSNR, SSIM and some of the popular No reference metrics such as the blind image integrity notator (BLIINDS) which is based on discrete cosine transform. The paper finally suggests the use of perceptual loss for improving the training process.

The scientific community continues to be lacking a reference dataset to judge objectively and quantitatively the performance of projected dehazing ways. The few datasets that are presently thought-about, each for assessment and training of learning-based dehazing techniques, completely deem artificial hazy pictures, to deal with this limitation, the paper Ancuti et al. [6] introduces the primary out of doors scenes information (named O-HAZE) which has both real hazy and corresponding haze-free pictures. The hazy pictures are captured in presence of real haze, generated by haze machines, and it contains forty-five different outdoor images portraying a similar visual content recorded in haze-free and hazy conditions, below similar illumination parameters.

Y. Zhang et al. (19) proposes the Haze Realistic Dataset for dehazing algorithm benchmarking. The author investigates haze simulation using clear images and depth maps. To simulate haze, sRGB images are first converted to LRGB. Then an atmospheric light term of 0.76 is used, which, according to the paper, is the value of atmospheric light at which the image becomes vivid. The paper also suggests using a visual range term, which is converted to a scattering coefficient . The hazy image is created by combining these two terms and the clear image.The paper then compares many models trained at NYU and Middlebury, analyzing various metrics such as SSIM and comparing performance on the HazeRD dataset. The paper falls short in that it proposes the dataset for testing rather than training, and the number of images in the dataset is also extremely limited. Also, the algorithms compared in the paper are old models based on priors, and the deep learning models used for benchmarking, such as dehazenet, are quite simple. The metrics reflect this shortcoming, as the maximum SSIM in the comparative study is only 0.70, which is quite low by modern standards.

Geiger et al. (20) proposes a novel dataset for driving related tasks. We particularly choose Kitti-depth subset of the dataset for our study. The KITTI-Depth dataset provides depth maps of LiDAR clouds which have been matched with the stereo cameras' depth estimates. With only 5% of the pixels accessible, the depth pictures are quite little and the remainder is lost. The dataset contains 86K training, 7K validation , and 1K test images on the server that have no ground truth access. In order to build new challenge-oriented computer vision benchmarks, the paper uses autonomic driving platform Annieway, for tasks such as: stereo, optical flow, visual odometry, identification of 3D objects, and 3D tracking. have fitted a typical station car for this purpose with two high-resolution video cameras in colour and in grey. A Velodyne laser scanner and a GPS tracking device offer precise ground truth. Each picture can be seen in up to 15 vehicles and 30 footpaths. The objective of the work is to eliminate this distortion and supplement current standards by providing the community with new real-world benchmarks.

The evaluation of the work presented on the cvlibs KITTI dataset website. It includes about 93,000 profile maps with raw LiDaR scans and RGB pictures in line with KITTi's raw data. The dataset will allow the formation of complicated, profound learning models to complete and forecast a single image depth, given the vast number of training data. It offers unpublished depth maps for manually picked pictures to act as benchmarks for the two challenges. With the KITTI dataset, the ground clearance for different activities such as stereo viewing and the flow assessment was acquired with a 64-layer laser scanner with a 10 Hz rotating frequency. Scanners with just 32 or 16 layers and higher frequencies are commonly utilized for industrial applications, such as independent driving. This leads to extremely limited predictions in depth. Assessing the Sparse Convolutionary Network and a number of basic guidelines for the different levels of sparsity for newly listed KITTI sub-sets to evaluate the impact of extremely scarce information. Educate all networks using all laser readings and assess their performance using random dropout in order to change the input density.

2.2.2 *Deep Learning Networks*

We have looked at some of the most popular as well as distinct models.

Pavan et al. (7) has proposed a light convolution network which emphasis on factors such as time and computational efficiency for real-time purposes and faster network. The paper boasts an encoder-decoder network with the modern CNN with no atmospheric scattering assumptions. With the trade-off between network complexity-image quality in the LCA neural network, it can be used in any light processing device. With ideal elimination of haze with the help of a light convolutional neural network for quicker results on the current standard datasets, this gives a proportional result with respect to image quality compared to current revolutionary and modern dehazing models with significantly less computational resources. The paper suggests the mean squared error (MSE) as its loss function which is optimal and helps with the variance between the high contrast pixels and also helps with faster and smoother convergence to the local minima with the gradient descent. Although there are contrast problems in the paper with no means to reconstruct the color fidelity and intensity. The downside of this network is its inaccurate representation and reconstruction on low-light and thick hazed images and the shortcomings and flaws are overcome with the help of other models which will be discussed briefly in the next section.

B. Cai et al. (8) proposes a dehazenet which offers good trade between computational complexity and dehazing output, and hence making this model a viable and optimal starting point for dehazing application. The paper suggests a deep learning architecture based on Convolutional Neural Networks (CNN) with parallel networks with the help of concatenation of layers for feature extraction with lesser trainable parameter. The paper basically aims at generating a transmission map through 4 steps namely feature extraction, multi scale mapping with kernel of size 3,5,7 and then finding the local extremum and finally applying nonlinear regression. The paper is delineated to train the model to learn and assume the prior atmospheric assumptions. The paper proposes a nonlinear activation function called Bilateral Rectified Linear Unit (BReLU), which helps the model to incorporate the best features of Rectified linear Unit(ReLU) and maintain the transmission map in range [0,1]. The paper achieves prime performance over current models yet keeping it optimal and ease of access. The paper [8] limits itself to the atmospheric scattering model, as it can also be learned in a deeper neural network. The network due to its simplicity fails to perform as well as the other state of the art models but with around 800 citations it has been a base for many of the newer deep learning models.

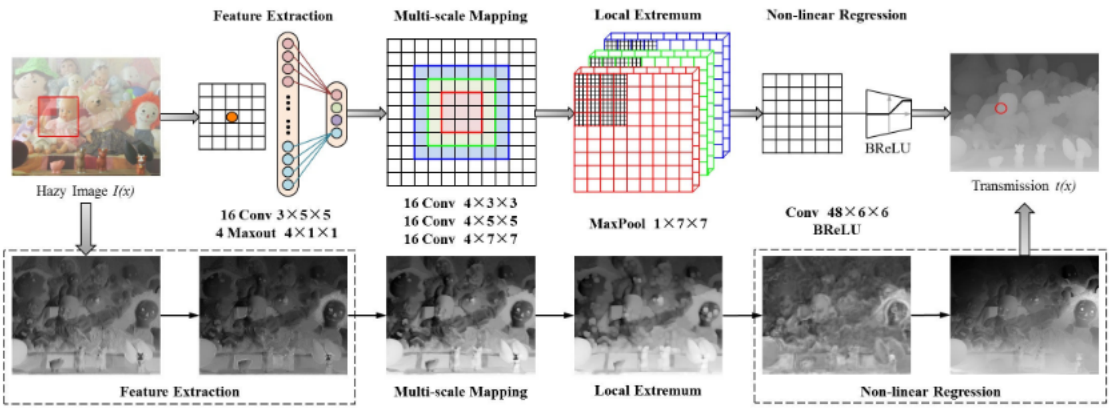


Figure 2.1: Dehazenet Architecture

B. Cai et al. (8)

Liu et al. (9) proposes a Generic Model-Agnostic Convolutional Neural Network (GMAN) for Single Image Dehazing. The paper does not estimate any atmospheric scattering model and only depends on trainable parameters and hence it learns complex hazy features and structures. The paper has numerous advantages over the Gated Fusion Network in terms of optimization, input size versatility, efficiency and complexity of the network. The paper suggests that the network is tailored for reconstructing the hazy image which is more generic and could be used for broader application. It also uses image by upscaling and downscaling by a factor of 2. As it is a new type of network there is not any concrete papers and research on this network.

Chen et al. (10) proposes a Gated Context Aggregation (GCA) Network for Image Dehazing and Deraining. Unlike the traditional computer vision papers which leverages traditional mathematical or image priors as the restoration constraints, e.g., DCP, the paper uses an gated context aggregation network to directly restore the final haze-free image. The paper proposes the modern smoothed dilation technique with dilated convolution in the network which takes up few to little trainable parameters and also uses a gated sub-network which blends features from multiple scales of the same image which helps the network remove the hazy elements from the image and also artifacts. The paper surpasses previous state-of-the-art dehazing methods by a large margin both quantitatively and qualitatively. The paper boasts the dilated convolutional layer, which braces the exponential expansion of the responsive field without loss of resolution and maintaining fidelity. The paper is computationally heavy, can be improved with custom loss function and can be optimized and made efficient for ease of access.

Y. Lee et al. (11) proposes the popular U-net based architecture for dehazing that employs squeeze and excitation blocks. The paper incorporates the dilation principle to increase the receptive field in the encoder part of the network. The paper uses squeeze blocks to provide channel attention and nearest neighbour upscaling in the decode network to improve output image quality. The model uses a sum of mean square error and perceptual loss as loss function for training. When implementing we found the model to exhibit colour distortion and dullness (like a light image).

Ayush et al. (12) proposes a pyramid convolution which dehazes in multiple layers of scales. The Uses a U-Net block which extracts new and complex features from the input image. The paper proposes a deep learning method which makes it as a generalized model resulting in it to be more versatile compared to other models. The whole of the algorithm is an extensive experimentation. It uses a combination of many losses. This model uses a pyramid convolution technique. The downside is reconstruction of the colors.

Qin et al. (13) proposes an end-to-end feature fusion attention network. The surpasses all the previous methods by a very high margin. It supersedes the previous algorithm with its powerful advantage of superior result in thick haze regions. Includes a FA module which mixes both channel attention and pixel attention mechanism which helps it focus more on the thick haze part. Local residual learning and feature attention (FA) each have a separate block which helps the model to process the image in the thin haze region through many skip connections.

Attention-based feature fusion (FFA) structure, it can process thin hazed regions by running it through multiple deep layers which helps it outperform every other feature fusion method. But one of the key drawbacks of the model is that it is quite computationally heavy.

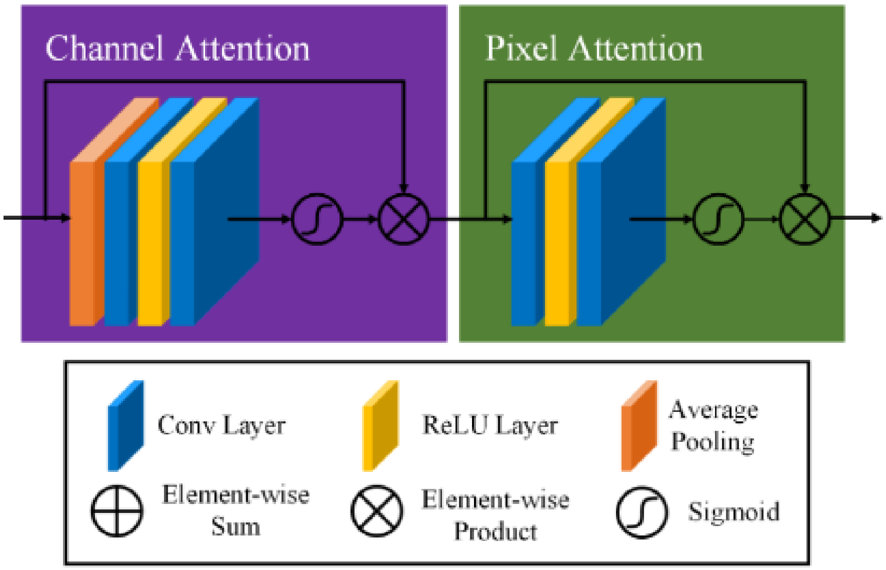


Figure 2.2: Feature Attention Blocks

Qin et al. (13)

C. Wei et al. (14) proposes a deep retinex decomposition model for low light enhancement. The model decomposes a image into reflectance and illuminance based on the retinex theory, and reconstructs the image and later uses image enhancement operations. C. Wei et al. [14] uses a data driven approach, and built a large dataset with paired low and normal light images named Low light paired dataset (LOL). The model takes input as both low light image and normal image during the training stage and only low light image during the testing stage. The paper uses a combination of 3 losses while training namely reconstruction loss, invariable reflectance loss and illumination smoothness loss. The results of the paper are quite good therefore we used this model skeleton when developing our model as described in methodology section (6).

Ibraheem et al. (21) proposes a simple model that lays emphasis on incremental model design. The paper firstly proposes a Densedepth network to generate a dense depth map monocularly I.e. from a single image. The model leverages the popular auto encoder model in specific a U-net type architecture to Generate the depth maps. The encoder part is the popular

Densenet169 model which is pretrained on Image net dataset. This encoder model is analyzed in Theoretical background subsection in detail. The decoder part of the network is made up sampling blocks which are composed of Up-scaling, concatenation, and convolution layers.

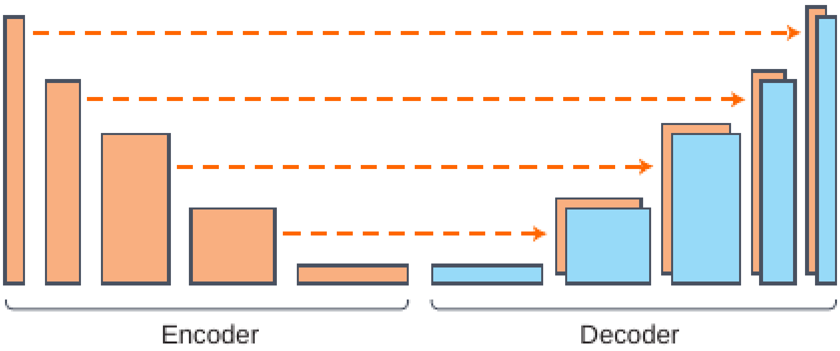


Figure 2.3: Dense Depth network

Ibraheem et al. (21)

The network was trained on NYU v2 as well as on KITTI dataset. The model output is at half the input models resolution, so for KITTI dataset images of input size 352X1216 the output is of size 176X608. The model is trained on a NVIDEA TITAN 12 Gb GPU for 50 epochs on KITTI dataset with size of 26k images, with optimizer as ADAM. The loss function used is a sum of three different loss functions SSIM, L1 loss and L1 loss on image gradients. For the KITTI dataset the average relative error is 0.093 and the SSIM is around 0.96 which makes the model stand out for task of monocular depth prediction.

2.2.3 *Losses*

Most of the networks we looked at in the above section, used either Mean square error (L2) or Mean absolute error (L1) as their loss function for computation of gradients for back propagation.

H. Zhao et al. (15) looks into loss functions which are specific to the tasks of image restoration for applications such as denoising, deblurring, super resolution and de-mosaicking. The paper looks into 4 different loss function candidates. Firstly, at Mean square error (MSE) which is one of the most common loss functions. MSE correlates poorly to image quality and human perception. It then looks at Mean absolute error (MAE) whose performance is better than MSE but still sub optimal. Then it looks at multi scale structural similarity (MS SSIM) , as SSIM with sigma value for the gaussian filter if used fails to preserve the local structure. The finally proposes a weighted combination of both MAE and MS SSIM.

2.2.4 *Metrics*

Metrics play a huge part in deciding how well a model performs, most of the models described earlier look at PSNR and SSIM as a metric to decide the quality of results. We in our work have tried to go one step further and have studied No reference image quality evaluators.

Mittal et al. (16) proposes a no reference opinion unaware, distortion unaware image quality assessment model. The model proposed by the paper derives quality aware around 36 quality aware features from natural scene statistics on a local patch. The optimal size of the local patch can vary from 32x32 to 160x160 as observed by the paper. The paper uses a corpus of 125 images selected from Berkley image segmentation dataset. A simple model of NSS is obtained by fitting the features to a multi-variate Gaussian (MVG). The difference between MVG of test image and that of the natural images gives the value of the Naturalness image quality index.

K. Choi et al. (17) propose a Fog aware density evaluator which is dependent on natural scene statistic like NQIE, but it only considers fog specific features. The paper considers a total of 12 feature like sharpness, colorfulness, dark channel prior etc. and fits a Multi variate Gaussian to the features which are derived on a local patch. The Mahalanobis-like distance between the haze free images and hazy images gives the density of haze. The paper further shows that this metric is linearly related to mean score of humans thereby making it a very import and suitable metric in the field of dehazing.

Y Liu et al. (18) proposes a novel no reference image quality assessment algorithm for contrast distorted images like hazy images. Contrast distortion is a general type of distortion found in many images and these hinders the quality of the images, so to quantify the quality of the image with the amount of contrast distortion present in an image. The paper proposes spatial characteristics, image histogram, visual perception characteristics and chrominance to quantify the quality. All the above four mentioned properties are extracted and then passed through a Support Vector Regression (SVR) to integrate a regression line to extract information. The downside of this method is its time complexity as it runs twice as slow as PSNR and SSIM.

2.2.5 Summary

In the above section we have gone through firstly the conventional models of dehazing like He et. al (1) which depend on priors and although they give satisfactory results the prior hold true in specific conditions only. We then went on to review different works in deep learning for the task of dehazing. We found abundance of deep learning for this task and have picked out some networks like FFA net (11), GCA net (10) etc. The networks we selected for the review are the current standout one taken from paper with code website on different datasets. We also selected the most popular dataset called RESIDE (5) and also looked at some of the other datasets. In the HazeRD (19) we looked at ways of generating hazy images. We then finally went through different losses, and metrics. In the metrics section we laid a emphasis on no reference and haze specific metrics (15; 16; 17).

After the review we found that there is an absence of deep learning models with theoretical backing, most of the models which we went through were developed for general purpose and on training also performed well for dehazing. We also found that most of the papers did a comparative study of their architecture of their model vs. Other models using full reference metrics which is quite a old approach, this does not give information about the model's ability to remove haze, or its performance on new data which might not a ground truth to compare to. Lastly, we found that no datasets in dehazing we application or task driven most of them were very generalized, taking the RESIDE (5) as an example, it has two divisions which cater to outdoor and indoor haze, but both the sub datasets are not driven by any application or tasks. Through our work we hope to work on those gaps, so that the field of dehazing can go to the next level.

2.3 Theoretical Background

2.3.1 *Atmospheric scattering model*

The haze present in a image is due to two features:

1. Direct Attenuation: results from decay in the medium
2. Air light: occurs due to the scattering of light in the surrounds

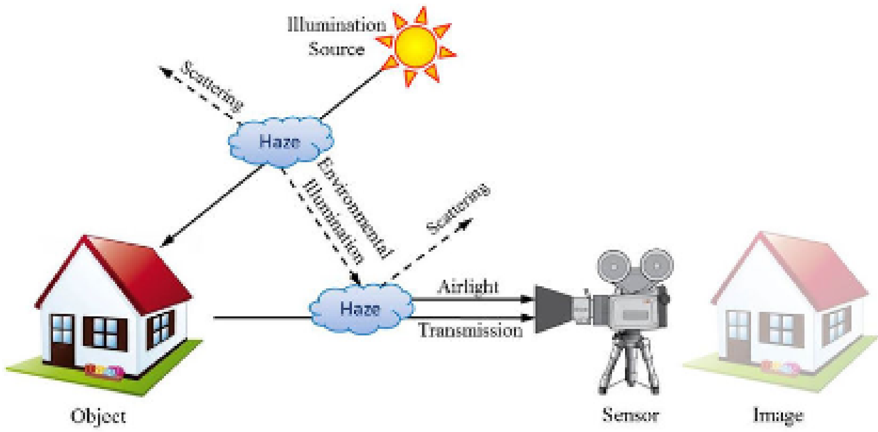


Figure 2.4: Atmosphere Scattering Model

B. Cai (8)

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{2.3}$$

Where I the hazy image captured, J is the dehazed image (scene Radiance), A is the global atmospheric light, and t is the transmission of the medium i.e. the portion of the light that is unscattered.

$$t(x) = e^{-\beta.d(x)} \tag{2.4}$$

Where β is the scattering coefficient of the atmosphere and d(x) is distance from the scene point It indicates that the scene radiance is attenuated exponentially with the scene depth. The model assumes homogeneous scattering of light and a constant scattering coefficient.

2.3.2 *Convolutional neural networks*

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning method for extracting features from images and classifying or reconstructing them. Convolutional neural networks are made up of many layers of artificial neurons. Artificial neurons are mathematical representations of their biological counterparts. Traditionally, the first layers seek out simple features and patterns. As the layers are added, it learns more complex features that allow it to comprehend the images in the same way that we do. The Pooling layer is similar to the Conv layer, but it aids in reducing the dimensionality of an image in order to reduce computation.The computing power required to process the data is reduced as a result of dimensionality reduction. There are two types of pooling: average pooling and maximum pooling. Max Pooling returns the maximum value from the image's Kernel-covered region. Average Pooling, on the other hand, returns the average of all the values from the Kernel's portion of the image. Because of the reduction in dimensionality, the process of learning an image becomes more effective. In terms of performance, CNN outperforms other traditional ML algorithms such as SVM and the XGBoost classifier. ALEXNET pioneered the CNN trend with only 8 layers.

Typically, the first layer extracts basic elements such as horizontal or diagonal rims. This output is sent to the next layer, which detects more complicated characteristics such as edges or combination edges. As we expand our network, even more difficult characteristics, such as objects and people, may be recognised.

CNN-related hyperparameters

- Number of Hidden Layers and Units: Having a large number of layers with regularisation will result in better performance. Less would result in an underfit.
- Network Weight Initialization: It is best to use different weight initialization distributions for different types of activation functions in a layer.
- Dropout: This is a type of regularisation in which a specified percentage of neurons in a specific layer are dropped at random in order to increase each neuron's reliance on learning new features.
- Activation function: These are used to highlight non-linearity in the model using a specific continuous and discrete mathematical function.For example, ReLu, Sigmoid, Tanh, and so on.

- Learning Rate: A higher learning rate takes larger steps during back propagation to reach minima, whereas a lower learning rate takes smaller steps. Ideally, we would like to use a larger learning rate at first and gradually reduce it to converge at the minima with less hindrance. This is referred to as Declining Learning Rate.
- Momentum: This parameter assists the model in taking a step forward so that it does not oscillate between two points.
- Batch size: The number of subsamples that must be trained before the model can be back propagated.
- The number of epochs is the number of times the training process is repeated.

2.3.3 VGG NET

The first ImageNet VGG19 network for ILSVRC-2014 was built, with an error rate of 7.3%. This was a year-end accomplishment. VGG19 is simple to understand, quick, and simple. It has 5 layers, each with 2 to 4 layers. The network is made up of 5 layer stacks. They are followed by categorization layers that are fully linked.

This is based on a simple concept: convolution layers function as "feature extraction." They function as perceptual perception fields, identifying patterns and geometric forms of increasing complexity, and completely linked layers function as a classical perceptron.

In applications such as super-resolution, weights from different levels of VGG19 are removed in order to use VGG19 for content loss (also called perceptual loss). The network concentrates more details and patterns as you go deeper. Based on the content loss, we can almost flawlessly rework the image using early convolution layers (conv1-conv3). As networks begin to catch more information, quality begins to deteriorate in the fourth and fifth layers.

The VGG Loss is a type of content loss introduced in the Real Time and Super-resolution Style Transfer and Perceptual Loss Framework. VGG Loss strives for perceptual similarity; it is an alternative to pixel-like loss. The loss of VGG is determined by the active layers of ReLU in the 19-layer VGG network. With the n-th convolution, before the n-th maxpooling layer, specifying the feature map within the provided VGG19 network. As a result, the VGG loss is defined as the Euclidean distance between a reconstructed image representation and a reference image.

2.3.4 DenseNet

A DenseNet is a form of neural network which uses dense links between different layers, through Complex blocks. All layers are connected to one another directly (with corresponding functional map sizes). Each layer gets additional context from all previous levels and translates its own feature maps into all following layers to protect the feed forward nature. Each layer of DenseNet receives new information from all previous layers and transmits its own functional charts to all levels in the future. Focusing is utilized. The ”collective knowledge” is obtained in each layer from all previous levels.

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

Figure 2.5: Densenet Architecture

There are four Dense Blocks for each of the architectures with different layers. The DenseNet-121 for instance contains four layers in the DenseNet-121 and the DenseNet-169 in the 4 dense blocks [6, 12, 32, 32]. A 7x7 stride2 Conv layer, followed by a 3x3 stride-2 MaxPooling layer, comprised the initial element of the Architecture DenseNet. Following the fourth dense block, a classification layer allows for the characterization maps for the categorization of all layers of the network.

Furthermore, the bottle Neck layers are the convergence processes inside the individual design. Bottleneck layers are used to reduce the complexity and dimension of the model. This indicates that 1x1 conv lowers the number of input channels, whereas 3x3 conv executes the transformer version of the input with less than the input channels. As the network may be smaller and compact, i.e. the number of channels can be less as each layer gets functional mappings for all previous levels. For each layer, the growth rate k is the added number of channels. Therefore, computer efficiency and memory efficiency are better.

CHAPTER 3

METHODOLOGY

3.1 Introduction

We started our working trying to implement all the aforementioned papers and compare them using Full reference as well as no reference metrics. We also considered factors like training time, number of blocks and the ability of the model to be extended while analysing each model. We then looked at models from different domains which could be extended to image dehazing, and decided to use the Deep retinex network as base of our model. We then broke the task of image dehazing into sub-task and designed a network for each task and combined the networks using the popular atmospheric scattering model. Lastly we have presented a trained the model on a new dataset named hazy-KITTI.

3.2 Research Gap

After review of various works in they field we found that the embryonic field of single image dehazing has a few missing pieces which we have highlighted below:

- Comparative study in most of the papers considers just traditional full reference metrics like PSNR and SSIM, even when there is a wide array of no reference and haze specific metrics which would be ideal method of comparing the models.
- Absence of a application oriented dataset, as most of the datasets available are very general and just aim to perform the task of dehazing outdoors or indoors. Just to illustrate this, in RESIDE outdoors dataset there are images of buildings from a very high point as well as pictures of roads and people from very near.
- There seems to be an absence of any mathematical modelling in the models, even the best performing models such as FFA are very generalized, as they rely on good resources for training and using very specialized training techniques rather than emphasizing on a good architecture.

3.3 Problem Definition

As stated earlier, we have taken up the project to combat the visibility issues faced by people and machines, as clear images not only help humans avoid accidents but also help in applications such as local path planning and object recognition.

When we look at the available approaches and there seems to a clear lack of ranking system, this has prompted us to take up an intensive comparison study on the present SOTA models. We further looked at developing a model which although being domain specific could also be extended to many computer vision algorithms like depth correction,image deraining, image denoising, processing before object detection and segmentation.

3.4 Objectives

- Compare various SOTA dehazing algorithms and compare them using common performance metrics.
- To analyse the performance of SOTA models on no reference and haze specific metrics.
- Analyse different areas of application for the algorithm and suggest the most suitable area of use.
- Develop a dehazing model that readily integrates with the software stack for autonomous cars.
- Develop a model that has sound theoretical backing, so that not only dehazed image but other haze related components can be extracted.
- Create a dataset which is driven towards dehazing images of roads.
- Perform a comparative study where novel model is compared to other SOTA models and bring out the novelty of the new model.

3.5 Work Plan

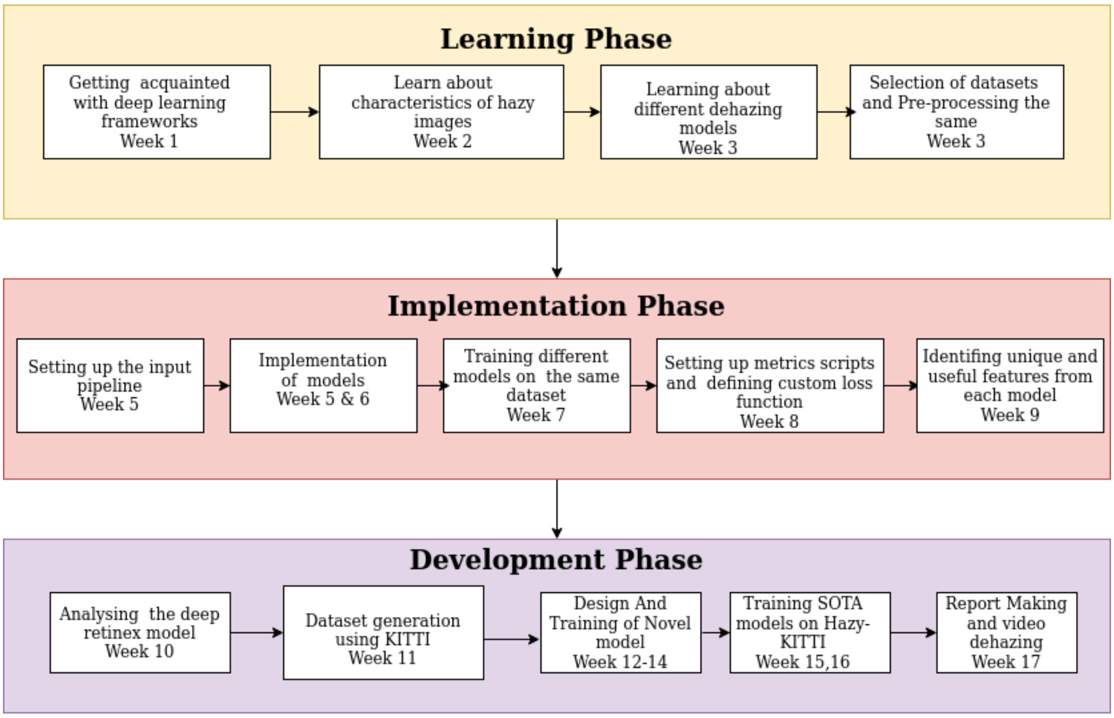


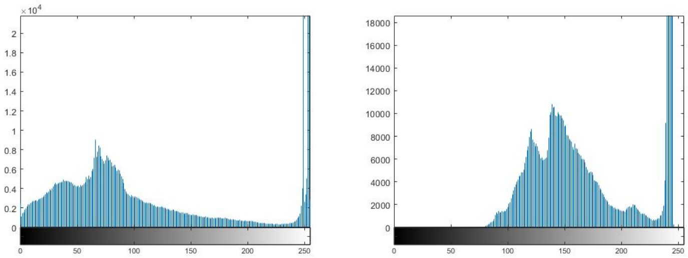
Figure 3.1: Project schedule

3.6 Learning phase

We started out by learning the basics of Deep Learning, we concentrated on getting an understanding of the concepts in deep learning, and learning about the various models like LeNet, ResNet, VGG net etc. We began by building some basic sequential models in TensorFlow. We also learnt the basic modules in TensorFlow and read the documentation to get acquainted with the framework. The initial phase of the project was dedicated to learning and getting hands on experience in the field of deep learning.

After understanding the diverse ways of defining a deep network architecture in we selected the model sub classing method as it gives us more control over the low-level parameters than the Functional API and the sequential API. Also, the Model subclass API is similar to the way we define models in Pytorch, so we could use this to easily convert models from Pytorch to TensorFlow. We then looked at distinctive features of a hazy image and started comparing the

(a) (b)



(c) (d)

Figure 3.2: Analysis of Haze (a) Clear image (b) Hazy image (c) Pixel intensity histogram of clear image (d) Pixel intensity histogram of Hazy image

histograms of pixel intensities as well as many other features to get an answer to the question, what makes an image hazy.

3.7 Comparison of state-of-the-art models

We then identified the best models that are used for dehazing an image. A comprehensive analysis of these papers is available in the literature review section. We looked at a wide range of methods, starting with conventional methods such as Dark channel prior to the recent complex deep learning methods like Feature fusion attention network. To gain an in-depth understanding of each of the models we implemented all of them in TensorFlow, on the reside dataset. Most of the models were implemented from scratch taking help from the respective papers and the available implementation which was in Pytorch. We firstly calculated PSNR and SSIM for the model from a test set of 64 images. Then we looked at calculating No reference metrics like CQIE, NIQE and we further looked at haze specific metric like FADE. This was a missing in many of the papers we have reviewed in the literature review section, as most paper provided comparison on only Referential metrics like PSNR. Non-Referential metrics are a great choice since there is generally an absence of ground truth when dehazing the image. In the table detailed in the results section we have compared both Referential and non-Referential metrics for each model.

Through the implementation of SOTA models above we gained valuable insights into each of the model, and we got an answer to the most important question, what makes each model a viable choice for dehazing. We also tried combining various models and there features to come up architectures that would incorporate best features of each model. We looked at implementing features such as dilations and attention layers in the models and trained the models to study the improvement.

3.8 Ideation phase

We then looked at models from domains like low light enhancement, denoising, super resolution, depth prediction for inspiration. At this point we came across an interesting model named Deep Retinex, the brief review of the same can be found in literature review section. The model proposed a novel architecture for low light enhancement. The model first broke down the image into two parts, reflectance and illuminance. Then the model trained networks for each of the tasks separately, and used a combine net to combine the two nets to generate the final image output. The main features we picked out from this paper was to keep our network modular and to decompose our large dehazing tasks into smaller tasks mathematically and train a network for each of the tasks.

As pointed out in the research gap section, one of the vital missing pieces in all the literature is absence of a dataset that was focused on a particular application. One of the leading datasets was the RESIDE dataset which focuses on the task of dehazing but the models trained on this dataset cannot be applied to outdoor dehazing applications. We looked at the NYU dataset, which emphasizes on indoor dehazing and looked at developing a similar one for outdoors. As we have already stated one of the main motivations behind this work was to develop a model that help combat accidents that occur due to haze. So, we looked at mainly two dataset, KITTI and Cityscapes dataset. Cityscape's dataset already had a sub dataset with foggy image and ground truth, but the dataset relied on stereoscopic techniques, I.e., there needs to be 2 cameras for getting the depth information. We wanted to develop a better dataset which used LiDAR sensor for getting the depth map, as LiDAR sensor has better performance under different weather conditions as well as being more accurate.

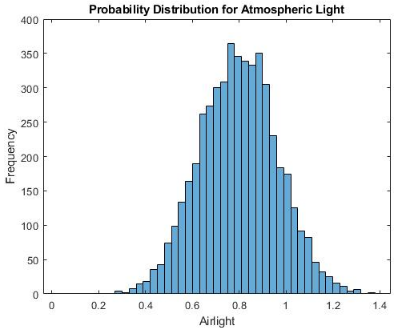
3.9 Synthesizing a Dataset

We chose the KITTI dataset as a base for developing our dataset, the dataset is large and is used in training many SOTA models for different applications, we have done a review of the dataset in the Literature review section. We picked out the depth prediction subset of the dataset, as shown by the atmospheric scattering model, a hazy image can be constructed by 2 components, a ground truth image and the depth map of the respective image. The synthesis of dataset was done with the help of code from the HAZERD paper.

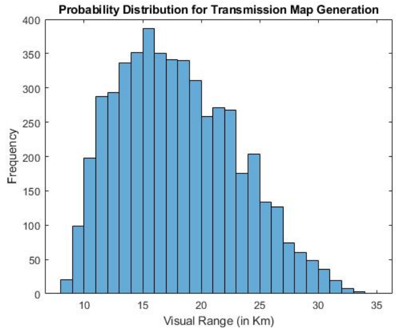
```
%Code snippet to generate hazy image
I = srgb2lrgb(I0);
transmission = exp(bsxfun(@times,-beta,d));
Ic = bsxfun(@times,transmission,I)+bsxfun(@times,1-transmission,A);
I2 = lrgb2srgb(Ic);
```

Above is the program that is used to generate hazy images; where I0 is the ground truth image, beta is the scattering coefficient, d is depth map; A is atmospheric light and Ic is the hazy image generated.

The HazeRD paper suggests setting the atmospheric light to 0.76 and deriving beta from more intuitive term called visual range that indicates how far we can see clearly in haze. For the task of dataset generation, we derived the atmospheric light variable from a uniform distribution and the visual range variable from a skewed Gaussian distribution, The skew in the Gaussian distribution was introduced so that the dataset would lay greater emphasis on dense haze. These



(a)



(b)

Figure 3.3: Probability distribution of selection of (a) Atmospheric Light (b) Visual range

two random distributions were designed after testing the images on visual as well as on other aspects such as naturalness, fog aware features.

The KITTI dataset has ground truth Images captured on 1.4 Mega pixel Point grey flea 2 cameras, and the LiDAR images were captured using a Velodyne HDL-64E sensor. The LiDAR sensor images were converted to dense depth maps using impainting algorithms available in the NYU V2 toolbox. We choose one thousand images from the Kitti dataset, and for each image there were 5 values of visual range that were chosen from the random distribution described above, also a random value for atmospheric light is also chosen from a uniform distribution. We created 5,000 images for the training task ahead, and we produced 77 images from a separate subset of KITTI for the test.

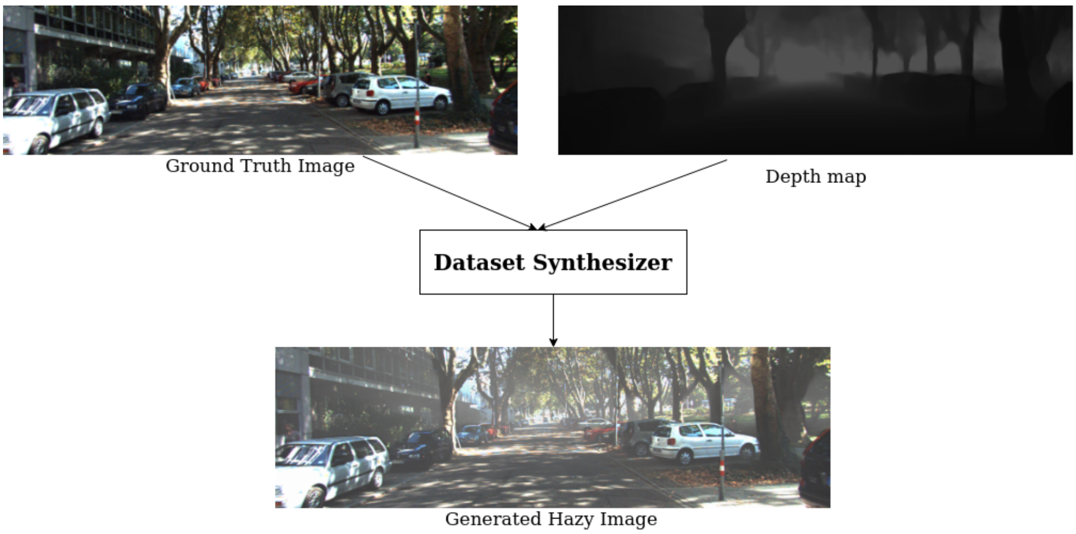


Figure 3.4: Dataset Synthesis Using HazeRD

3.10 Training and performance of different models on Haze

3.10.1 Dark Channel prior

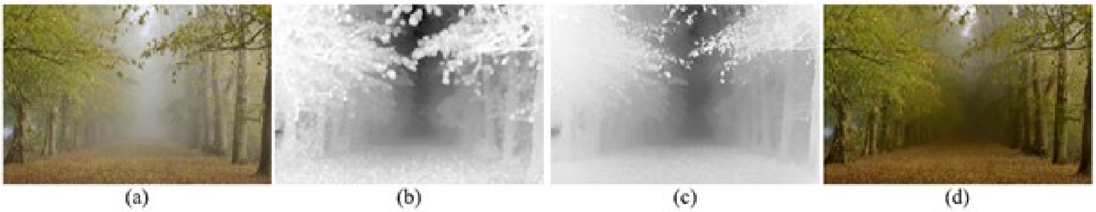


Figure 3.5: DCP process (a) input image (b) transmission map estimation (c) refined transmission map (d)Dehaze Image

He et. al (1)

There are 4 steps in dehazing a image using DCP;

1. Calculating the dark channel of a patch; I.e. finding the minimum of pixel intensities in a neighborhood across the three channels.
2. Calculating patch transmission, using the aerial perspective factor and assuming the parameters atmospheric light and transmission is constant in a neighbourhood.
3. Performing soft matting to refine the transmission map.
4. Estimating the atmospheric light which is determined by choosing the 0.1% brightest pixel in the dark channel and among these pixel picking the brightest corresponding pixel in the hazy image.
5. Finally the scene radiance is recovered using the Atmospheric scattering model.

After analysis of results it is found that DCP model has many drawbacks; firstly as the model is based on priors it does not perform well on sky images and as the algorithm relies entirely on priors it fails to retain the Naturalness of the image and most of the times it over enhances the image leading to high contrast or dark images.

3.10.2 Dehazenet

Dehazenet is a model that provides a considerable improvement from the conventional DCP method, it uses the same method to atmospheric light but uses a deep learning approach to find transmission map. The model is both light weight and quick. The in-depth analysis of dehazenet is discussed in literature review section. There are 4 steps in getting the dehazed image from a hazy image:

1. The first step is feature extraction which is done using 16 convolution filters with kernel size 5, then a max-out layer is used to get haze relevant features.
2. The next layer we implemented was using multi scale kernels I.e. convolutions layers with different kernel size. We implemented it using 3 different kernel size 3,5,7. The outputs of all this are concatenated.
3. The next layer we implement a max pool layer to get local extremum, this helps achieve spacial in-variance.
4. Finally a non linear regression layer is implemented using 48 filters with kernel size 6, and a ReLU layers is used as activation. A bounded ReLU is used as the value of transmission ranges between 0 to 1 as it is a fraction.

For the implementation we first used transfer learning to get the wights from a pretrained model in pytorch framework on RESIDE dataset, and then countinued training the model on hazy-KITTI dataset. We then trained the model for around 50 epochs taking 30% random subsamples of dataset at a time, and the model converged to give a mean square error of 0.0086.

3.10.3 Light Convolution Autoencoder

Light convolution Autoencoder (LCA) net is a light network which focuses on being faster rather than performance. The model is a sequential one with first part of the model acting like an encoder i.e. the image is reduced to a feature vector and then a decoder network is used to recover the haze free image from the feature vector.

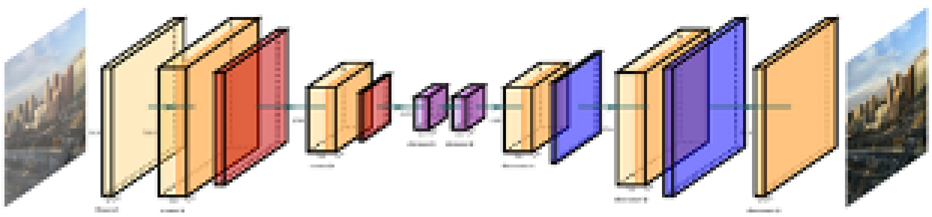


Figure 3.6: LCA NET

Pavan et. al (7)

Layer	Output Shape	Trainable parameters
Input_layer	(Batch,352,1216,3)	0
Conv2d_1 (encoder)	(Batch,352,1216,50)	1400
Average_pool_1 (encoder)	(Batch,176,608,50)	0
Cov2d_2 (encoder)	(Batch,176,608,50)	22550
Average_pool_2 (encoder)	(Batch,88,304,50)	0
Dense_1 (encoder)	(Batch,88,304,10)	510
Dense_2 (encoder)	(Batch,88,304,10)	110
Conv2d_Transpose_1 (decoder)	(Batch,88,304,50)	4550
UpSample2d_1 (decoder)	(Batch,176,608,50)	0
Conv2d_Transpose_2 (decoder)	(Batch,176,608,50)	22550
UpSample2d_2 (decoder)	(Batch,352,1216,50)	0
Conv2d_Transpose_3 (decoder)	(Batch,352,1216,3)	1353

Table 3.1: LCA net architecture

The model has around 53 thousand trainable parameters. We could not find an implementation of the paper so we designed the layer from scratch, and as the model was simple it was implemented using functional API. After Training the model for 50 epochs taking a 30% random subset of Hazy Kitti dataset, we achieved a Mean square error of 0.0056 and the model had converged giving very decent results for a simple model. The recovered image was still showing a small amount of haze and the image also showed blurring effect. After a through analysis we could conclude that because of incorrect architecture I.e. attempting to recover the image from a feature vector rather than taking advantage of the mathematical modeling of haze led to the model's low performance.

3.10.4 U-Net

U net consists of two parts, the first part which is on the left is the contracting part, which consists of convolution layers and secondly the expansion part which up-samples the output of intermediate layers using a combination of convolution and transpose convolutions to recover back the final image. U net is used for a wide range of applications from segmentation to super resolution. For dehazing task, we picked out a special variant of U-net called Contextualized Attentive U-net. The U-net is based on two principles,

1. Squeeze and Excitation module: The main principle behind this block is first a squeeze action is applied I.e., with the help of convolution filters and max pool the size of the

of the image is reduced and this is concatenated to the expand side of the network after applying a suitable activation function. This is done to calibrate the channel attention of feature maps.

2. Dilated convolutions: The main aim of dilated convolutions is to capture a larger field than traditional convolutions, so basically the receptive field of the kernel is increased. In this model we use dilations in the bottle neck layer. The concept of parallelized dilation is used which has two branches first is a direct connection between the encoder part of unet to decoder part and the next one consists of a set of parallel dilation blocks with different dilation rates.

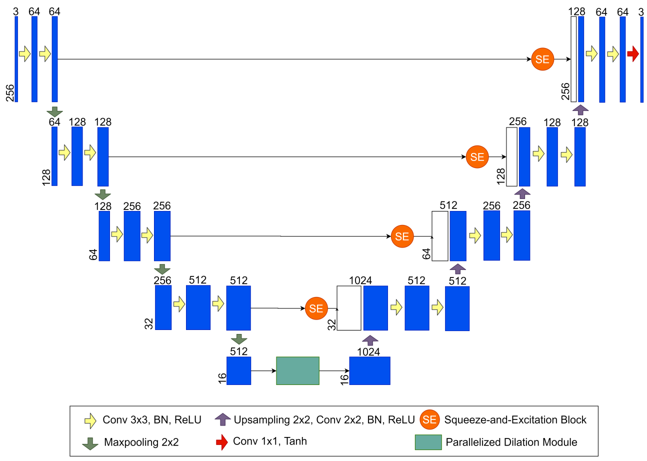


Figure 3.7: U-net architecture

Lee et. al (11)

For training this model on hazy Kitti dataset we applied transfer learning I.e., we started with the weights of model already trained till convergence on RESIDE indoor training set. This model was trained on hazy Kitti dataset with mean square error and perceptual loss as the loss functions. After training for 20 epochs the model gave a mean square error of 0.0016.

3.10.5 Generic Model-Agnostic Convolutional Neural Network

1. The Generic Model-Agnostic Convolutional Neural Network (GMAN) does not use atmospheric scattering model and only depends on trainable parameters and hence it learns complex hazy features and structures.

2. The Generic Model-Agnostic Convolutional Neural Network (GMAN) has numerous advantages over the Gated Fusion Network in terms of optimization, input size versatility, efficiency and complexity of the network.

3. The Generic Model-Agnostic Convolutional Neural Network (GMAN) network is tailored for reconstructing the hazy image which is more generic and could be used for broader application. It also uses image by upscaling and downscaling by a factor of 2.
4. As it is a new type of network there is not enough work and research on this network.

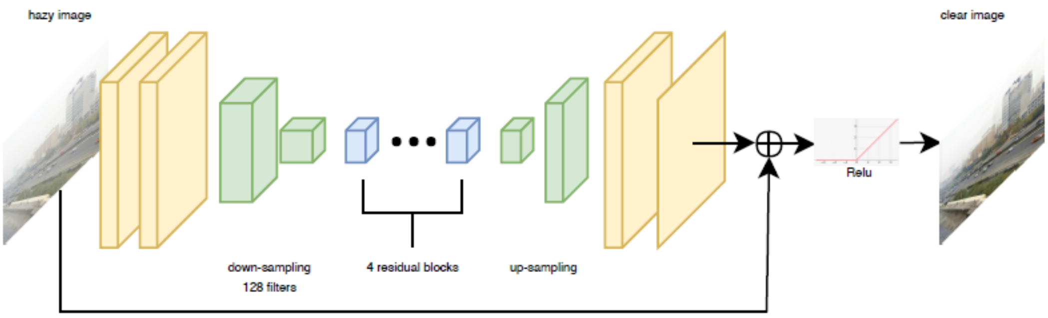


Figure 3.8: GMAN architecture

Liu et. al (9)

We started training GMAN nets from the pretrained weights on RESIDE dataset. After training the model for 50 epochs we could get a MSE loss of 0.0035. Also we found that the generated image lacks the colours of the ground truth image.

3.10.6 Gated Context Aggregation Network

Unlike the traditional computer vision models which leverages traditional mathematical or image priors as the restoration constraints, e.g., DCP, the Gated Context Aggregation (GCA) uses a gated context aggregation network to directly restore the final haze-free image.

1. The Gated Context Aggregation (GCA) proposes the modern smoothed dilation technique with dilated convolution in the network which takes up few to little trainable parameters and uses a gated sub-network which blends features from multiple scales of the same image which helps the network remove the hazy elements from the image and artifacts.
2. The Gated Context Aggregation (GCA) surpasses previous state-of-the-art dehazing methods by a large margin both quantitatively and qualitatively.
3. The Gated Context Aggregation (GCA) uses the dilated convolutional layer, which braces the exponential expansion of the responsive field without loss of resolution and maintaining fidelity.

4. The Gated Context Aggregation (GCA) is computationally heavy, can be improved with custom loss function and can be optimized and made efficient for ease of access.

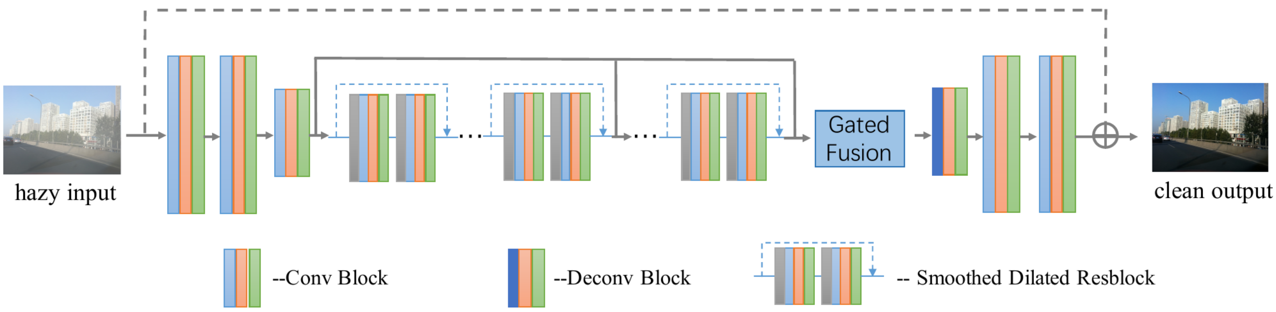


Figure 3.9: GCA architecture

Chen et. al (10)

After training the model for 80 epochs on the hazy Kitti dataset we could achieve a mean square error of 0.0013. One of the noticeable drawbacks of this model is that sometimes we found black patches in the output image. After a complete analysis we found that this might have happened due to dilation rates which might not be chosen with any theoretical backing.

3.10.7 Feature fusion attention network

Feature fusion attention network is one of the current best nets in the field of image dehazing. We have already looked at the network at detail in the literature review section, here we discuss the network from a implementation standpoint. The network architecture has 4 key features which are pixel attention, channel attention, blocks, and residual connection of groups. When we first tried to implement the network on TensorFlow framework we hit a local minimum, we tried to get around this model using random weights as direct transfer of weight tensors from Pytorch had proved to be unsuccessful, as there were some of the layers in Pytorch which we could not implement in TensorFlow. The next solution we tried to implement is converting the model from Py torch to TensorFlow using ONNX. We were able to accomplish this task, but the model converted could only be used for inference not for training.

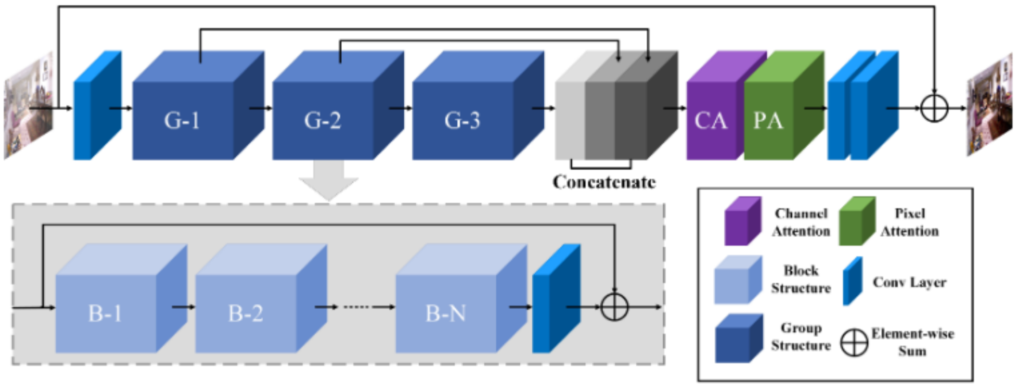


Figure 3.10: FFA architecture

Qin et. al (13)

Finally unable to get a functioning implementation of FFA model on TensorFlow we turned to the authors implementation on Py torch, which was trained on RESIDE. When we started training the model on hazy Kitti dataset we ran into memory overflow on GPU, we solved this problem by reducing the number of blocks from 19 to 14, and we took random patches of image rather than full image to avoid the memory overflow issues. Finally, after running the model for around 1000 steps; which were trained taking pretrained weights as initial weights; we could achieve a mean square error of 0.0014.

3.11 Outdoor Depth-based Dehazing Network (Novel Model)

While designing our model, we hoped to implement the best features from each each of the model we analyzed in the literature review section. We firstly went with the approach suggested in deep retinex network, which is breaking the problem into many smaller models, so keeping atmospheric scattering model in mind we subdivided the problem into 2 smaller problems. The first one was finding the atmospheric light and the second one was generation of a transmission map. After iterating through many other approaches, we had decided to use this approach as it had the theoretical backing behind the architecture.

The approach was made possible as during the dataset synthesis part we had kept a track of atmospheric light by appending it to the end of the file name of the generated hazy image.

We also stored the transmission maps generated for each visual range parameter of the image thereby enabling us to train a network to generate transmission map.

3.11.1 Atmospheric Light Net (A-net)

1. Taking up the inspiration from LCA-net for calculating atmosphere light we created this simple end-to-end network.
2. It was trained for 20 epochs and loss function used was MSE. Giving a validation loss of 0.0191 this model outperformed all our previous models.
3. It applies dilated convolution and batch normalization for faster convergences and a feature learning which is simple but effective.
4. To keep the continuous features without any hindrance in the output since it is also a continuous value, we used ReLU to our advantage.
5. Compression of the model is taken care by the pooling layers which takes the maximum value of the window and compresses into half every pooling layer it passes through.

Layer	Output Shape	Parameters
input	(Batch,352,1216,3)	0
max_pooling2d	(Batch,88,304,3)	0
conv2d (Conv2D)	(Batch,84,300,32)	896
max_pooling2d_1	(Batch,21,75,32)	0
conv2d_1	(Batch,19,73,16)	4624
max_pooling2d_2	(Batch,4,18,16)	0
conv2d_2	(Batch,2,16,8)	1160
flatten	(Batch,256)	0
dense	(Batch,1)	257

Table 3.2: A_net Architecture

This model was developed using trial and error method after iterating over several architectures. The final model was developed keeping the features of the atmospheric light in dcp paper; in specific we kept in mind that atmospheric light in DCP paper is maximum pixel intensity of both dark channel and input image, so we tried to incorporate this feature by using convolution layers followed by max pool layers and embedded dilation layers to increase the receptive field. When we had tried to generate hazy image from ground truth images, we had noticed that a change of 0.1 does not significantly affect the image generated. So, once we achieved a mean square error of 0.0191, we stopped the training as it was good enough to proceed, also the combined model would be trained later so a MSE of 0.0191 was acceptable.

3.11.2 Transmission map net (Tnet)

Tnet reflects how much information from object goes through to the camera. This is a generally a value between 0 to 1, where 0 reflects a very dense haze and 1 reflects a clear image. One of the earliest approaches for generating a transmission map was the dehazenet which used multiscale convolutions and nonlinear regression to calculate transmission map. After doing an initial training on the dehazenet we started searching for better models from different application. As we have seen in the theoretical background section, we know that the transmission map is related to the depth map of the image, So we started to look at depth prediction models that could be used for this application. We looked at Monocular depth estimation models on KITTI dataset as they could be readily absorbed into our workflow. While searching for the model we the below requirements in mind:

- simple to code from scratch and easy to train.
- Have a sound theoretical backing.
- Should generate a refined depth map.
- Should be Agile I.e. it should be ready to integrate with future changes such that it can support incremental design workflow.

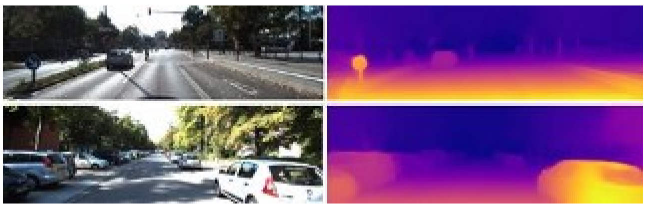


Figure 3.11: Densedepth outputs

After looking at different models like Monodepth, Monodepth2, Adabins and densedepth we finalized densedepth as the base model. Densedepth model was chosen because it could be readily modified to make it suitable for other application as the architecture was simple and generalized. The other models such as Adabins and Monodepth were focused on accurate depth map generation and that the reason they were not a good fit for our use case. As mentioned earlier, the depth map is directly proportional to the transmission map, therefore we took the pretrained densedepth network made few tweaks like adding bounded ReLU activation function

Figure 3.12: Tnet Outputs

as transmission map values are between 0 and 1, also we added a upscaling block as densedepth give output at half of the input image resolution but we require output to be of the same size. We also tested by adding pixel and channel attention block, but these did not improve the model so this was dropped later. We also experimented with different versions of encoders in the encoder part of the densedepth net and found that densenet 169 was better than densenet 201. We also found that efficientNet B4 also gave a satisfactory performance.

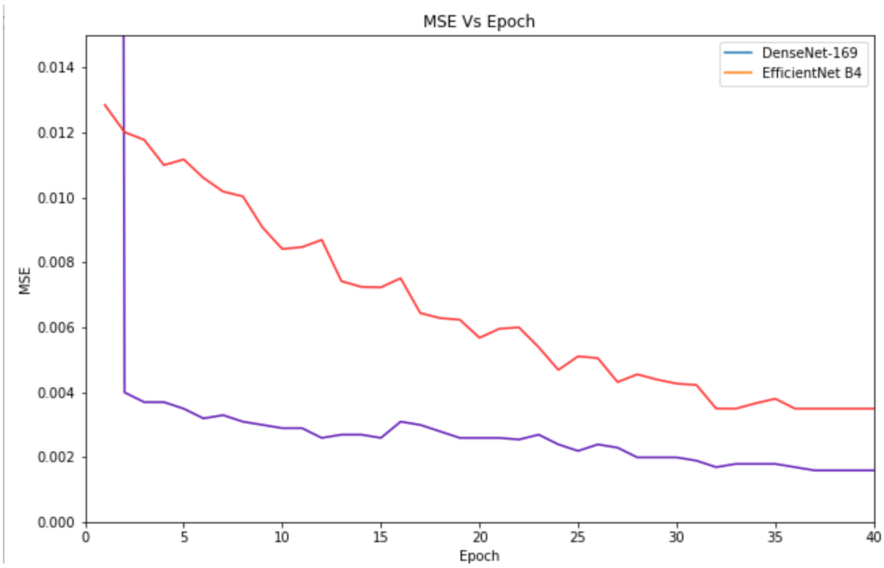


Figure 3.13: Tnet Training

3.11.3 Combination Net

For designing the combination net I.e. the net that would take the outputs of anet and T net then after combining them produces a final image we had two ideas in mind

1. The first approach we tried was a simple one which was basically taking the outputs from the two nets and using the atmospheric scattering model formula combining the two intermediate outputs to generate final output.
2. The second approach was a very ambitious one, after combining the terms using the atmospheric scattering model, we denoise the network by taking inspiration from a simple denoising network.

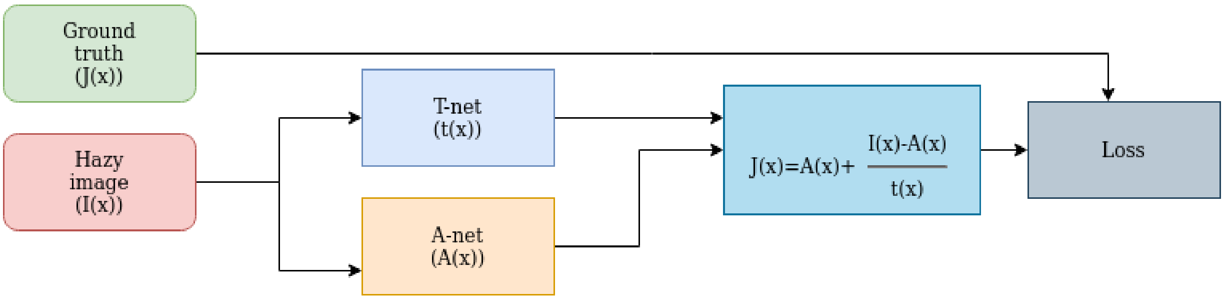


Figure 3.14: ODD net Architecture

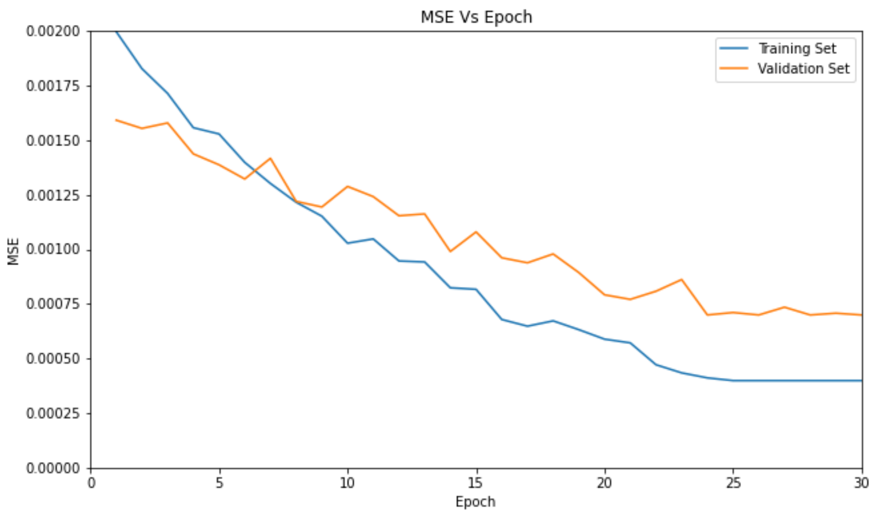


Figure 3.15: ODD net Training

Following the first approach, we combined the already trained A net ad T net models using the Atmospheric scattering model and trained the total architecture (ODD net Outdoor depth-based dehazing model) for a further of 30 epochs, by periodically reducing the learning rate from 1x10e-4 to 1x10e-9. We also used combination of MSE and VGG loss as the loss function. After training we were able to achieve a MSE error of 0.0004. We also started to work on the second approach but the size of the network made training difficult, as colab provides only a GPU of size 16 GB and the training was taking a long time, but we hoped to continue and try out this approach as a part of our future study.

3.12 Tools

The tools used for the current project are: NumPy, Matplotlib, TensorFlow

NumPy is numerical computational library which is fast and compact with the help of its vectorization and broadcasting. Matplotlib is visualizing tool to display the image outputs and inputs. TensorFlow is an open-source library which is used for machine learning and deep learning. TensorFlow was developed by the Google rain team for internal Google use. We chose TensorFlow over PyTorch is because of the flexibility with Google Colab. In TensorFlow framework a model can be defined in 3 ways: sequential, functional API and model subclassing. We mostly defined our models using sub-classing as it gives us lower-level control. Most models implemented in PyTorch can be converted into TensorFlow model subclass. Google colab lets us run programs on a jupyter notebook kind interface online. It also lets up access GPU and TPU to speed up training.

3.13 Assumptions

We though our work wanted to focus on visibility outdoors as we were motivated to take up the problem due to the visibility loss through rise in air pollution. So, our main focus was on roads therefore our dataset was based on the kitti dataset which is developed by processed feed from cameras and LiDAR sensors placed on top of the car. During the process of synthesis of dataset we deliberately biased the dataset towards dense haze so that our results could be visually checked. Lastly we based our model on atmospheric scattering model as it has a sound theoretical backing. As we have considered the atmospheric scattering model for our work, the underlying assumptions are that depth is directly proportional to the amount of haze and atmospheric light is constant through out the image as outdoors there is generally one source of light that dominates.

3.14 Summary

In the methodology section we have first described the research gap in the current literature, highlighting the lack of theoretical backing behind the current SOTA models and absence of an application focused dataset. Keeping the research gap in mind we formulated our problem definition and objectives. Then after a flow chat describing the workflow, we have detailed our work.

Project was divided into 3 phases, in the first phase we dedicated most our time to learning the architectures and deep learning frameworks, and then when through the available literature. In the second stage of the project, we trained the SOTA models on RESIDE dataset and performed a comparative study based on reference, no reference and haze specific parameters.

In the last stage of the project, we developed a new model named Outdoor Depth-based Dehazing, which relies on a deep net implementation to find atmospheric light and a pre-trained densedepth model with some modifications to generate transmission map, finally the two models are combined using the atmospheric scattering model to generate a dehazed image.

CHAPTER 4

RESULT AND DISCUSSION

The project as explained before can broadly be divide into 2 steps, the first one was comparative study on RESIDE dataset. We have implemented 7 models, with each model proposing a quite unique architecture and training approach.

Metrics	**DCP**	**LCA**	**DehazeNet**	**GMAN**	**U-net**	**GCA**	**FFA**
PSNR	12.21	17.07	17.23	14.80	19.28	20.13	20.67
SSIM	0.61	0.65	0.66	0.64	0.73	0.77	0.79
FADE	0.38	0.95	0.50	0.65	0.68	0.91	1.24
NQIE	2.85	4.42	2.93	2.70	3.71	2.7	2.67
CEIQ	3.19	3.27	3.33	3.22	3.4	3.22	3.42

Table 4.1: Comparison Study of dehazing models on RESIDE Dataset

For the study firstly each model was developed and trained in tensorflow framework, and each model was trained from scratch, as implementation and trained weights for most of them were only available in pytorch, and consideration of trained weights for only some models would lead to bias towards those models. Each model was trained on a subset of RESIDE outdoor dataset which consisted of around 2000 images and trained each model for 30 epochs. After the study we validated that FFA net is better then other SOTA models.We also validated the fact that DCP even after being just a prior based algorithm, is one of the best conventional method of dehazing. After the comparative study we also found that along with full reference metrics like PSNR, no reference metrics like naturalness index and haze specific metrics such as fog density are vital for in-depth analysis in the field of dehazing. We also found that our work had few shortcomings, like the dataset was quite small and the performance of the models obtained was not good enough for any dehazing application.

In the second part of the project, we first developed the hazy-KITTI dataset, whose quality we calculated using naturalness index and human opinion scores. The dataset had was a synthesized one, but the metrics indicated it would very closely replicated real world scenarios. As the dataset was generated from LiDAR sensor images, it also had some drawbacks like inconsistent haze near object edges.

After the dataset synthesis process we went on to retrain all the models previously consider on the hazy-KITTI dataset. This time the dataset size was around 6000 images, and while training we made use of transfer learning. Special care was taken so that each model was trained till convergence, we kept the number of maximum epochs for each model at 50, most models converged before that limit. Also effort was taken to make the metric as close as possible to the ones the authors were able to achieve in their respective works.

After the study we could get close performance values for most of the models, this makes our work one of the first to rigorously compare various dehazing models. After the comparative study we found that FFA net was truly a standout net even when a mini version was implemented by us. We also reaffirmed the fact that GMAN, GCA and U-NET are all good variants and call be used depending on requirements and constraints. Lastly, we were able to design a net based on strong theoretical backing, which outperformed the other networks in most of the metrics.

Metrics	Hazy	LCA	Dehazenet	DCP	GMAN	GCA	FFA	U-NET	Ours
PSNR	15.66	18.32	20.87	16.65	24.64	22.94	27.45	22.56	31.35
SSIM	0.73	0.69	0.69	0.62	0.89	0.90	0.90	0.86	0.95
FADE	1.12	0.63	0.62	0.36	0.43	0.41	0.39	0.57	0.45
NQIE	2.75	4.36	3.92	2.53	3.09	3.21	2.54	2.73	2.79
CQIE	3.14	3.02	3.22	2.77	3.23	3.11	3.23	3.19	3.27
BLIINDS2	6.23	40.80	23.40	21.87	16.13	3.97	9.54	8.88	11.86

Table 4.2: Comparison Study of dehazing models on Hazy-KITTI dataset

Our work is not just a dehazing net, it also gives information about atmospheric light which is directly linked to the intensity of light and in the intermediate layers it also generates a transmission map, which has been derived from a trained depth net. Therefore it can be used to identify haze density and also in further study it can readily be incorporated in the study of depth correction in hazy weather.

(a) Ground Truth

(b) Hazy image

(c) DCP

(d) LCA

(e) Dehazenet

(f) GMAN

(g) GCA

(h) U-Net

(i) FFA

(j) Ours

Figure 4.1: Comparision of outputs of various models on Hazy-KITTI dataset

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Summary

Through this project we have firstly provided a through comparison study of the SOTA models, as discussed before we have considered not only traditional metrics such as PSNR and SSIM but also considered no reference and haze specific metrics thereby making our study one of the first efforts to understand the SOTA models through the perspective of Naturalness, Fog factor etc. The comparative study was performed on common datasets and particular care was taken while training each model, so that it could reflect its novelty in the comparative study. Also, it was made sure that each model converges on training set thereby considering the best trained model for comparison and avoiding bias towards any model.

As already discussed, there was a lack of a task driven application-oriented dataset, so we came out with a new dataset which was based on KITTI dataset. Taking help of the methods previous authors had used to generate datasets, we generated a dataset whose size was around 6000 images from KITTI dataset. This dataset was designed with skew towards dense haze as we wanted to make our results visually verifiable.

The last and the most important contribution of this project is the ODD net, which is a novel architecture built by us. The model makes use of a lightweight net for atmospheric light estimation and a cleverly designed net which is a derivative of the densedepth net for generation of transmission map. These two networks are combined using the atmospheric scattering model to generate the haze free image.

5.2 Future scope of work

Though lot of time and effort was put into this project, there are some areas which require additional work and could be taken up in the future. One of the main tasks is deployment of the model and making the model adaptable to different input image sizes. This would enable the project to be integrated with the self-driving car software stack. It would also expose the shortcomings of the model thereby enabling the betterment of the model.

The next area where additional work can be done is dataset synthesis, although we have made our dataset as close as possible to real world scenarios, still there is scope to achieve better naturalness in the generated images. Also, the current impainting algorithms which are used to convert sparse LiDAR models to dense depth maps are still lacking the dense depth maps that we are currently able to generate using Kinect sensor indoors.

Lastly, with the rise of autonomous navigation depth maps have been a regular feature in most of the navigation and path planning algorithms. This project has taken an initial step in the field of depth correction, but there is still a lot to be done. Haze affects the sensor's ability to provide correct depth information therefore there is a need of a depth correction algorithm that would be an intermediate between the depth sensor outputs and the depth maps synthesis.

5.3 Future Work Proposed

- Train the denoise layers of the net, by incorporating random sampling of image
- Write a wiki page for the model to help people understand the model.
- Set up a access based system for access of internet community to the dataset through AWS S3 buckets.
- Deploy the model on the cloud using the concept of FaaS and allow users to send a stream of hazy images stream as input and get dehazed images as output.
- Add a input size in-variance feature to the model so that it has the ability to work with inputs from different cameras.
- Submit the work as a research article.

REFERENCES

[1] K. He, J. Sun and X. Tang, ”Single Image Haze Removal Using Dark Channel Prior,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 12, pp. 2341-2353, Dec. 2011, doi: 10.1109/TPAMI.2010.168.

[2] K. He, J. Sun, and X. Tang. Guided image filtering. In ECCV, pages 1–14. 2010.

[3] Q. Zhu, J. Mai and L. Shao, ”A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior,” in IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3522-3533, Nov. 2015, doi: 10.1109/TIP.2015.2446191.

[4] K. Tang, J. Yang and J. Wang, ”Investigating Haze-Relevant Features in a Learning Framework for Image Dehazing,” 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 2995-3002, doi: 10.1109/CVPR.2014.383.

[5] B. Li et al., ”Benchmarking Single-Image Dehazing and Beyond,” in IEEE Transactions on Image Processing, vol. 28, no. 1, pp. 492-505, Jan. 2019, doi: 10.1109/TIP.2018.2867951.

[6] C. O. Ancuti, C. Ancuti, R. Timofte and C. De Vleeschouwer, ”O-HAZE: A Dehazing Benchmark with Real Hazy and Haze-Free Outdoor Images,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, pp. 867-8678, doi: 10.1109/CVPRW.2018.00119.

[7] A. Pavan Bennur, Adithya Gaggar, Mohit S S, Shylaja. (2020). LCA-Net: Light Convolutional Autoencoder for Image Dehazing.

[8] B. Cai, X. Xu, K. Jia, C. Qing and D. Tao, ”DehazeNet: An End-to-End System for Single Image Haze Removal,” in IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5187-5198, Nov. 2016, doi: 10.1109/TIP.2016.2598681.

[9] Z. Liu, B. Xiao, M. Alrabeiah, K. Wang and J. Chen, ”Single Image Dehazing with a

Generic Model-Agnostic Convolutional Neural Network," in IEEE Signal Processing Letters, vol. 26, no. 6, pp. 833-837, June 2019, doi: 10.1109/LSP.2019.2910403.

[10] Chen, D, He, M, Fan, Q, Liao, J, Zhang, L, Hou, D, Yuan, L Hua, G 2019, Gated context aggregation network for image dehazing and deraining. in Proceedings - 2019 IEEE Winter Conference on Applications of Computer Vision, WACV 2019., 8658661, Proceedings - IEEE Winter Conference on Applications of Computer Vision, WACV , IEEE, pp. 1375-1383, 19th IEEE Winter Conference on Applications of Computer Vision (WACV 2019), Waikoloa Village, United States, 7/01/19.

[11] Y. -W. Lee, L. -K. Wong and J. See, "Image Dehazing With Contextualized Attentive U-NET," 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 2020, pp. 1068-1072, doi: 10.1109/ICIP40778.2020.9190725.

[12] Ayush Singh, Ajay Bhave, Dilip K. Prasad, "Single image dehazing for a variety of haze scenarios using back projected pyramid network," arXiv:2008.06713 [eess.IV]

[13] Qin, Xu & Wang, Zhilin & Bai, Yuanchao & Xie, Xiaodong & Jia, Huizhu. (2019). FFA-Net: Feature Fusion Attention Network for Single Image Dehazing.

[14] Wang, Wenjing & Yang, Wenhan & Liu, Jiaying. (2018). Deep Retinex Decomposition for Low-Light Enhancement.

[15] H. Zhao, O. Gallo, I. Frosio and J. Kautz, "Loss Functions for Image Restoration With Neural Networks," in IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 47-57, March 2017, doi: 10.1109/TCI.2016.2644865.

[16] Mittal, Anish Moorthy, Anush & Bovik, Alan & Chen, Chang-Wen & Chatzimisios, Periklis & Dagiuklas, Tasos & Atzori, Luigi. (2015). No-Reference Approaches to Image and Video Quality Assessment. 10.1002/9781118736135.ch5.

[17] L. K. Choi, J. You and A. C. Bovik, "Referenceless Prediction of Perceptual Fog Density and Perceptual Image Defogging," in IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3888-3901, Nov. 2015, doi: 10.1109/TIP.2015.2456502.

[18] Y. Liu and X. Li, ”No-Reference Quality Assessment for Contrast-Distorted Images,” in IEEE Access, vol. 8, pp. 84105-84115, 2020, doi: 10.1109/ACCESS.2020.2991842.

[19] Y. Zhang, L. Ding and G. Sharma, ”HazeRD: An outdoor scene dataset and benchmark for single image dehazing,” 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 3205-3209, doi: 10.1109/ICIP.2017.8296874.

[20] Geiger A, Lenz P, Stiller C, Urtasun R. Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research. 2013;32(11):1231-1237. doi:10.1177/0278364913491297

[21] I. Alhashim and P. Wonka, “High quality monocular depth estima-tion via transfer learning,” 2018,arXiv:1812.11941. [Online]. Available:http://arxiv.org/abs/1812.11941

[22] Clément Godard, Oisin Mac Aodha and Gabriel J. Brostow, ”Digging into self-supervised monocular depth estimation”, CoRR, 2018.

[23] G. Huang, Z. Liu, L. v. d. Maaten and K. Q. Weinberger, ”Densely connected convolutional networks”, Proc. Comput. Vision Pattern Recognit., pp. 2261-2269, 2017.

[24] https://github.com/cxtalk/DehazeZoo

[25] https://github.com/onnx/onnx

ANNEXURE I

Program Outcomes Mapping of project (NBA)

PO & PSO Mapping

Student Details: Abu Bakr Siddiq (170929098) & Razeem Akhtar(170929194)

Note: Use a ✓mark if you have addressed that PO in your report.

Table I.1: PO mapping

PO	**✓mark**	**Pg. No**	**Topics & Section No**	**Guides Observation**
PO1	✓	16-20, 26-38, 41	2.3, 3.9, 3.10, 3.11, 4	
PO2	✓	5-15, 21, 24, 28-33, 39	2.1, 2.2, 3.2, 3.7, 3.10, 3.13	
PO3	✓	25, 26, 34-38	3.8, 3.9, 3.11	
PO4	✓	15, 24, 26, 28-33, 41, 44-45	2.2.5, 3.7, 3.9, 3.10, 4, 5	
PO5	✓	6-20, 24, 28-38, 39	2.2, 2.3, 3.7, 3.10, 3.11, 3.12	
PO6	✓	2, 3, 22, 26, 39	1.3, 1.4, 3.3, 3.9, 3.13	
PO7				
PO8				
PO9	✓	23	3.5	
PO10	✓	1-4, 21-23, 40, 41-45	1, 3.1-3.5, 3.14, 4, 5	
PO11	✓	3, 22, 23	1.4, 3.3, 3.4, 3.5	
PO12	✓	22, 23, 25, 45	3.5, 3.6, 3.8, 5.2, 5.3	

Table I.2: PSO mapping

PSO	**✓mark**	**Pg. No**	**Topics & Section No**	**Guides Observation**
PSO1	✓	26-27, 28-33, 34-38	3.9, 3.10, 3.11	
PSO2	✓	16-20, 22, 34-38	2.3, 3,4, 3.11	
PSO3	✓			

Signature of Student: Name and Signature of Guide:

Date:

ANNEXURE II

Program Learning Outcomes (LO) during project period (IET)

LO Mapping

Student Details: Abu Bakr Siddiq (170929098) & Razeem Akhtar(170929194)

Note: Use a ✓mark if you have addressed that LO in your report.

Table II.1: Learning Outcome Mapping

LO	**✓mark**	**Pg. No**	**Topics & Section No**	**Guides Observation**
C1	✓	16-20, 26-38, 41	2.3, 3.9, 3.10, 3.11, 4	
C2	✓	5-15, 21, 24, 28-33, 39	2.1, 2.2, 3.2, 3.7, 3.10, 3.13	
C3	✓	16-20, 24-38, 41-43	2.3, 3.7, 3.8, 3.9, 3.10,3.11, 4	
C4	✓	5-15, 21, 28-33, 41-43	2.1, 2.2, 3.2, 3.10, 4	
C5				
C6	✓	16-20, 26-27, 34-38, 41-43	2.3, 3.9, 3.11, 4	
C7	✓	2, 3, 26-27, 39	1.3, 1.4, 3.9, 3.13	
C8				
C9				
C10				
C11				
C12				
C13	✓	6-15, 16-20, 28-33, 39	2.2, 2.3 3.2, 3.10, 3.12, 3.13	
C14				
C15	✓	3, 22, 23	1.4, 3.3, 3.4, 3.5	
C16	✓	23	3.5	
C17	✓	1-4, 21-23, 40, 41-45	1, 3.1-3.5, 3.14, 4, 5	
C18	✓	22, 23, 25, 45	3.5, 3.6, 3.8, 5.2, 5.3	

Signature of Student: Name and Signature of Guide:

Date:

ANNEXURE III

Questions and Answers

Answer the following questions with relevant to your Practice School work.

1. Explain the steps you considered to investigate and define the problem in your project work

(C4, evaluate level)

Answer: Below are the steps we took for defining the problem

- Identify a broad field of challenge:
  - Efficiency or performance issues.
  - Model processes that can be enhanced
  - Researchers' area of concern
  - Difficulties faced by Users.
- Learn more about the problem:
  - In the context, background, specificity and significance of what has been already known about the problem
  - choose the particular element your study will cover.
- Constraints and Exclusions
- Acceptance Criteria
- Assumptions taken
- Finally, by assessing and evaluating the work.

2. Discuss the science, mathematics, statistics, engineering principles and other basic technology you identified for design (Mechanical, Electronic, Physics, Chemistry, Automation) in your project work. (C1, C2, C3, Application, Analysis, Evaluation of Science and Mathematics in the project)

Answer: Artificial intelligence, statistics and computer vision were the main fields of study and research in the project. Other small component comprises atmospheric light dispersion, fog dispersion and methods for optimization.

3. Have you considered the Environmental and Sustainability limitations in your project work? (C7, evaluate)

Answer: We have looked at air pollution and its affects at driving, and proposed a way to combat it for the task of driving.

4. Have you considered ethical, health, safety, security, and risk issues; intellectual property; codes of practice and standards while addressing these issues in your project work? If so, Explain in detail. (C5, create)

Answer: We have used the Do not Repeat yourself approach (DRY) while programming and made our code modular, also effort has been taken to contribute to open source community.

5. What were the esthetical issues faced and how it was addressed in your project in the design phase? (C5, analysis)

Answer: There were no esthetical issues faced to address them.

6. Were there any health issues considered during design process? How it was addressed in your project in the design phase? (C5, create)

Answer: There were not any health or related issues regarding this project as the project was solely focused on the Neural Network model development.

7. What were the safety, security and risk issues considered in the design stage? (C10, create)

Answer: The model was skewed towards dense haze for safety purposes.

8. Have you come across intellectual property issues in the project phase? (C5, create)

Answer: There were no intellectual property issues faced in this project.

9. What are the codes of conduct and standards you needed to use in design phase and in other phases of your project as well? (It may include codes of practice and standards for safety, security, health, risk) Explain the legal issues, ISO standards, IEC standards, etc. (C8, evaluate)

Answer: Since our project was being carried out in the online mode, the only codes of conduct were punctuality in submission of reports and completion of the assigned work.

10. What were the professional ethics needed to be followed in general while you are doing the project? (C8, evaluate)

Answer: Some of the professional ethics needed to be followed in generally during the project are: accountability, honesty, loyalty and respects for others.

11. Do you think ethics and professionalism needs to be paid attention by students during study? If, yes, explain how it can be inculcated/introduced/implemented? (C8, evaluate)

Answer: Ethics and professionalism are an integral part for any work environment. Implementing ethics and professionalism will manifest the respect for fellow members and their work and most importantly respect for each other's time which is of paramount significance. We inculcated this by following a calendar/schedule for each week, honesty, respecting each other and receiving suggestions and feedback from our mentor.

12. Do you think environmental and sustainability limitations; ethical, health, safety, security, and risk issues; intellectual property; codes of practice and standards are sufficiently covered in the courses you have studied in your curriculum? (C8, evaluate)

Answer: Environmental and sustainability limitations; ethical, health, safety, security, and risk issues; intellectual property; codes of practice and standards were sufficiently covered by the Department of Humanities.

13. Have you gone through online classes, or a crash course in which you are familiarized with intellectual property rights as well as risk issues in professional environment? (C8, evaluate)

Answer: No, we have not gone through online classes, or a crash course in which you are familiarized with intellectual property rights as well as risk issues in professional environment.

14. In the beginning of your project did you evaluate environmental effects and sustainability factors in your work? (C7, evaluate)

Answer: Our project doesn't affect the Environment and any of its resources.

15. Did you addressed the limitations of your project work and have you improved the results through continuous improvements in your project work? (C5, create)

Answer: Yes, the limitation has been addressed and the model was developed by following the Software Development Life Cycle (SDLC) process with agile methodology for continuous improvement over time.

16. How did you plan your project, deadlines, maintaining dairy of each stage and improved the quality of the project? (C14, understand)

Answer: Implementing ethics and professionalism will manifest the respect for fellow members and their work and most importantly respect for each other's time which is of paramount significance. We inculcated this by following a calendar/schedule for each week, honesty, respecting each other and receiving suggestions and feedback from our mentor.

17. Are you aware of the ethical clearance when you work in the field of health/medical applications.? (C8, evaluate)

Answer: No, I am not aware of the ethical clearance when you work in the field of health/medical applications.

18. Did you adopt any quantitative technique for any engineering activity related to your project? (C3, evaluate)

Answer: The project was solely focused on the loss and how to reduce it by tuning the model, hence there was no need of quantitative techniques

19. What were the elements of your project work which addresses sustainable development and were you able to apply quantitative techniques to analyze and achieve your project goals? (C7, evaluate)

Answer: There aren't any elements of our project work which addresses sustainable development and quantitative techniques were not adopted in this project.

20. Did your project need the understanding of relevant legal requirements governing engineering activities you carried out as a part of your project work? Explain in detail. (C8, evaluate)

Answer: No, this project did not have any requirements for knowledge on legal requirements for engineering activities.

21. What are the legal, ethical practices you followed while working on project? (C8, evaluate)

Answer: Most of the work was carried out online and hence the only thing that mattered was punctuality in submission.

22. Are you sure that you abide IPR/copy right issues? (C15, apply)

Answer: Yes, this project does abide by the IPR/Copyright issues

23. What online course you attended to improve your communication skills, Report writing, Oral presentation, Software used for writing report? (C17, apply)

Answer: There was no online course taken up by us to improve our communication skills, Report writing, Oral presentation, Software used for writing report.

24. In your project, was it needed to tackle risk issues, including health & safety, environmental and commercial risk, and of risk assessment and risk management techniques? Explain in detail. (C5, create)

Answer: There were no risk issues, including health safety, environmental and commercial risk, and of risk assessment faced by this project.

25. How is the organization addressing a fire accident/human safety when working with machines? (C9, evaluate)

Answer: We are not aware of the fire accident/human safety precautions carried out at the organization.

26. Process of teamwork. How each of you are involved in the team? What part the work is addressed by you.? (C16, evaluate)

Answer: We split the work for each model and kept up with each other by using GitHub and regular meetings.

27. Have you filed patent, IPR, or published your work? Give more details. (C17, evaluate)

Answer: We have not filed for a patent, IPR or published this work

28. How you documented the literature review, your analysis on their results, discussion with the guide and team members, provide the documents on weekly basis. Put as one chapter in final report. (C4, evaluate).

Answer: Since the project was being carried out on online mode, we used GitHub to the fullest purpose by making a calendar, the issues faced and needed improvement. GitHub saves all this data in chronological order and hence it was very helpful.

29. Have you sensitized about inclusion and diversity in the team? If yes, what are the diversification in the team in terms of religion, gender, ethnicity, etc? (C11, apply).

Answer: Our primary focus on creating a good dehazing model, so we did not sensitize about inclusion and diversity in the team.

30. How were you able to keep yourself updated with the technology? How you incorporated advanced technology in your project. (C18, lifelong learning)

Answer: We were up to date with the current technologies with the help of internet, and we also included some of the latest technology in our project.

31. Which are the laboratory skills you found applicable to your project? Explain. (C12, apply)

Answer: Computer Vision labs we went through were applicable to this project.

ANNEXURE IV

Project Classification

Student Details: Abu Bakr Siddiq (170929098) & Razeem Akhtar(170929194)

Note: Use a ✓mark to specify under which domain your practice school work falls into.

Table IV.1: classification based on project domain classification

Domain	**✓mark**
Product	
Application	✓
Review	✓
Research	✓
Management	

Note: Use a ✓mark to specify Societal impacts you considered during your practice school.

Table IV.2: classification based on societal consideration

Social Impact	**✓mark**
Ethics	✓
Safety	✓
Environmental	✓
Commercial	
Social	

Signature of Student: Name and Signature of Guide:

Date:

ANNEXURE V

Student Details

Table V.1: Student Details

Student Details			
Student Name	Abu Bakr Siddiq		
Register Number	170929098	Section / Roll No	32
Email Address	mailabubakrsiddiq@gmail.com	Phone No (M)	8494945140
Project Details			
Title	Image Dehazing using Deep Learning		
Start Date	08/02/2021	End Date	28/05/2021
Organization Details			
Guide Name	Dr. Kshetrimayum Lochan		
Designation	Assistant Professor		
Full contact address with pin code	Dept. of Mechatronics Engineering, Manipal Institute of Technology, Manipal – 576 104, Karnataka, INDIA		
Email address	lochan.k@manipal.edu	Phone No (M)	+91-7002525059
Co-Guide Details			
Co-guide Name	Dr. Munendra Singh		
Full contact address with pin code	Dept. of Mechatronics Engineering, Manipal Institute of Technology, Manipal – 576 104, Karnataka, INDIA		
Email address	munendra.singh@manipal.edu		

Signature of Student: Name and Signature of Guide:
Date:

Table V.2: Student Details

Student Details			
Student Name	Razeem Akthar		
Register Number	170929194	Section / Roll No	32
Email Address	razeemakthar79@gmail.com	Phone No (M)	9739230355
Project Details			
Title	Image Dehazing using Deep Learning		
Start Date	08/02/2021	End Date	28/05/2021
Organization Details			
Guide Name	Dr. Kshetrimayum Lochan		
Designation	Assistant Professor		
Full contact address with pin code	Dept. of Mechatronics Engineering, Manipal Institute of Technology, Manipal – 576 104, Karnataka, INDIA		
Email address	lochan.k@manipal.edu	Phone No (M)	+91-7002525059
Co-Guide Details			
Co-guide Name	Dr. Munendra Singh		
Full contact address with pin code	Dept. of Mechatronics Engineering, Manipal Institute of Technology, Manipal – 576 104, Karnataka, INDIA		
Email address	munendra.singh@manipal.edu		

Signature of Student: Name and Signature of Guide:
Date: