A Review of the Impacts of Defogging on Deep Learning-Based Object Detectors in Self-Driving Cars

Isaac Ogunrinde
Electrical and Computer Engineering Department
FAMU-FSU College of Engineering
Tallahassee, USA
isaac1.ogunrinde@famu.edu

Shonda Bernadin
Electrical and Computer Engineering Department
FAMU-FSU College of Engineering
Tallahassee, USA
bernadin@eng.famu.fsu.edu

Abstract—Autonomous Vehicle (AV) technologies are faced with several challenges under adverse weather conditions such as snow, fog, rain, sun glare, etc. Object detection under adverse weather conditions is one of the most critical issues facing autonomous driving. Several state-of-the-art Convolutional Neural Network (CNN) based object detection algorithms have been employed in autonomous vehicles and promising results have been established under favorable weather conditions. However, results from the literature show that the accuracy and performance of these CNN-based object detectors under adverse weather conditions tend to diminish rapidly. This problem continues to raise major concerns in the research and automotive community. In this paper, the foggy weather condition is our case study. The goal of this work is to investigate how defogging and restoring the quality of foggy images can improve the performance of CNN-based real-time object detectors. We employed a Cycle consistent Generative Adversarial Network (CycleGAN)-based image fog removal technique [1] to defog, improve the visibility and the quality of the foggy images. We train our YOLOv3 algorithm using the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) dataset [2]. Using the trained YOLOv3 network, we perform object detection on the original foggy images and restored images. We compare the performances of the object detector under no fog, moderate fog, and heavy fog conditions. Our results show that detection performance improved significantly under moderate fog and there was no significant improvement under heavy fog conditions.

Keywords— self-driving cars, adverse weather, fog, object detection, convolutional neural network, yolov3

I. INTRODUCTION

In 2018, according to the National Highway Traffic Safety Administration, over 36,000 people died from road accidents in the United States [3]. According to the World Health Organization, Global status report on road safety in 2018, about 1.35 million people die from road crashes annually [4]. A series of studies have indicated a growing need to provide cars with technologies that can mitigate driver error and negligence, and support drivers with physical and functional limitations [5].

Several Advanced Driver Assistant Systems (ADAS) have been developed in recent years with varying levels of autonomy to assist drivers. ADAS consists of safety features designed to avert potential accidents by alerting the driver of impending risk or seizing control of the vehicle in an emergency. ADAS such as intersection assistant system, lane change assistance, object detection, collision warning, electronic brake assistant system, Lane departure warning, etc., have been tested and implemented in research and commercial vehicles.

Today, most AV technologies appear to perform well under favorable weather but are faced with several challenges under adverse weather conditions. However, challenges regarding perception under unfavorable driving circumstances and/or inclement weathers remain. Such circumstances include the presence of snow, fog, haze, shadow, rainy road, extreme illumination into the camera. Many of the existing AV technologies and functionalities depend mainly on a group of sensors and/or camera systems. Under adverse weather conditions, the functionalities of sensors and cameras can severely be degraded, resulting in poor performance. Human drivers can visualize the environment with the eye, detect, and identify objects.

Many of these AV technologies used for perception, planning, and control, etc., are mainly tested under favorable weather conditions while some are tested under adverse weather conditions. Adverse weather conditions do not only degrade the performance of sensors and cameras used for perception but can increase the risk of traffic crashes and fatalities. Perception plays a substantial role in the object detection capability of AVs. Object detection is an important safety factor for both humans and AVs when navigating the road. Currently, object detection under adverse weather conditions (such as snow, fog, rain, sun glare, haze, etc.) is one of the most critical issues facing autonomous driving.

It is important to note that atmospheric phenomena including haze, fog, and mist occur as a result of suspended particles (such as dust, sand, water droplets, ice crystals, etc.) in the atmosphere. Meteorological studies show that all these phenomena mostly differ in their particle material, size, shape, and concentration [6, 7], however, their physical impacts on imaging are comparable [8]. Fog occurs when water droplets are suspended in the air, while haze occurs when air pollutants such as dust are suspended in the air. However, both phenomena can obscure visibility and decrease the contrast of an image. The quality of the image been captured by the cameras can seriously be degraded by fog. Degraded image quality can diminish the performances of image processing and object detection algorithms. The goal of this work is to investigate how defogging and restoring the quality of foggy images can improve the performance of CNN-based real-time object detectors. We employed a Cycle consistent Generative Adversarial Network (CycleGAN)-based image fog removal technique [1] to defog, improve the visibility and the quality of the foggy images. We train our YOLOv3 algorithm using the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) dataset. Using the trained YOLOv3 network, we perform object detection on the

SoutheastCon 2021 | 978-1-6654-0379-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/SOUTHEASTCON45413.2021.9401941

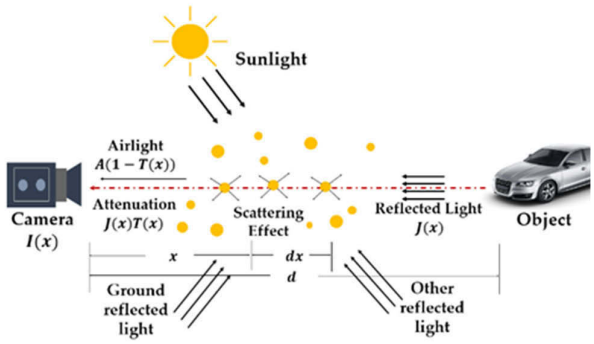


Fig. 1. An atmospheric scattering phenomenon of foggy imaging model

original foggy images and restored images. We compare the performances of the object detector under no fog, moderate fog, and heavy fog conditions. Our results show that detection performance improved significantly under moderate fog and there was no significant improvement under heavy fog conditions. This paper is organized as follows. In section II, we present the background of the research. In section III, we present a CycleGAN-based image fog removal technique. In section IV, we present the YOLOv3 network. In section IV, we present our results and discussion. Also, we summarize this work and provided direction for future work.

II. Background

A. Fog Imaging Model

Figure 1 illustrates the physical atmospheric scattering model under foggy weather. The physical atmospheric scattering model is made up of the attenuation factor, transmission model, and the airlight model. Under a foggy condition, the transmission model consists of atmospheric scattering that attenuates the light for imaging. As a result, the object textures and edge details of the target image can become degraded. Under foggy weather conditions, reflected light from the target object crosses attenuation and interference before it gets to the camera. Within the airtight model, the atmosphere scatters the light rays from the sun before been transmitted to the imaging camera. However, the transmitted lights, rather than been the scene light from the object in the image, consist of fog components that obscure the objects in the image.

Koschmieder [9] proposed the haze image model expressed in equation (1):

$$I(x) = J(x)t(x) + A[1 - t(x)] \quad (1)$$

where $I(x)$ represents the observed foggy image by the imaging equipment (camera), $J(x)$ represents the scene radiance image also known as the clean image recovered, $t(x)$ represents the transmission map, A is denoted as the airlight vector and it is homogeneous for every pixel in the image. $J(x)t(x)$ represents the attenuation factor, $A\,[1 - t(x)]$ represents the atmospheric components. A, t, and J are the unknown parameters of a foggy single input image I. The atmospheric light A and transmission t can be estimated to obtain the restored image (recovered image) J using equation (2).

$$\hat{J}(x) = (\hat{I}\,(x) - \hat{A}[1 - \hat{t}(x)])/\hat{t}(x) \quad (2)$$

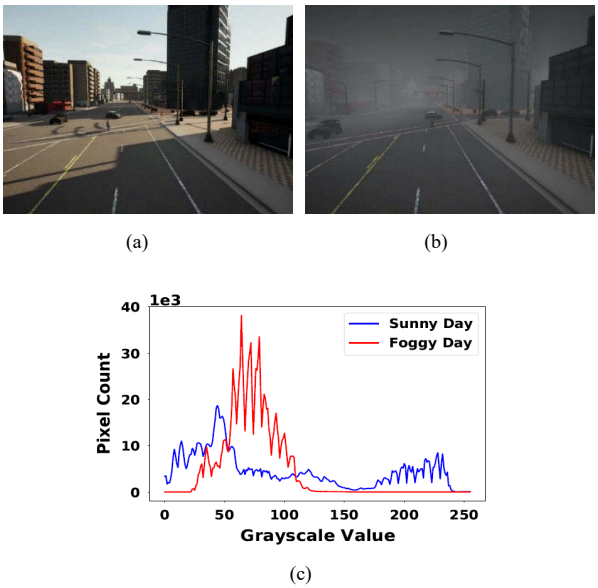


Fig. 2. The contrast between the gray scale of sunny day and foggy day images (a) sunny day image, (b) foggy day image, and (c) image showing the grayscale of both sunny day and fog day images.

B. Effect of Fog on Autonomous Driving

Fog is an atmospheric phenomenon made up of water droplets and ice crystals that are suspended in the air. The presence of fog can create an unsafe driving condition when navigating the road. Fog can impede visibility to the human eye and seriously degrade the quality of the image been captured in a machine such as AVs. Degraded image quality can diminish the performances of image processing and object detection algorithms. Under light foggy conditions, visibility can be lower than 1000 meters [10] and under thick fog, visibility can reduce to 50 meters or less [11].

Figure 2 represents the contrast between the grayscale of a sunny day and foggy day images. The color and feature information included in an image can substantially be revealed via grayscale. In object detection, feature information contained in the image can be used for label classification. From Figure 2, the grayscale of the sunny day image spreads from 0 to about 250. Nevertheless, the grayscale of the foggy day image is extremely concentrated 25 to about 110. Noise in form of spikes (approximately 40, 000-pixel counts) caused by the presence of fog is evident on foggy days. Thus, the presence of fog in an image can significantly change the feature information of an image and can negatively impact object detection. Also, in an environment without fog, frequency components are found to have a broad spectrum while frequency components are clustered at zero frequency in a foggy condition. Under fog, the smooth edges of an image are defined by low frequencies, and the sharp edges are formed by both high and low frequencies [12].

Hence, the contrast of an image tends to diminish while pattern edge recognition of the image becomes extremely difficult under fog weather conditions [12, 13]. When compared to other adverse weather conditions, fog negatively impacts the detection capabilities of sensing devices most. Because the extinction and backscattering coefficients generated by fog (5×10^{-3} to 1.5×10^{-2}) is greater than those of rain and snow ($< 10^{-3}$) [12, 14]. In a study conducted by Anik Das in [15], driving situations on foggy days were

compared to sunny days. The author observed that the likelihood of lane deviation from standard deviation under foggy conditions was greater than that of favorable conditions.

C. Image Restoration Method Based on Single Images

From literature image restoration method based on a single image can be divided into two main categories: (i) priori-based (ii) learning-based.

Priori-based: Single image fog removal based on priori-based methods is also known as the hand-crafted technique. Usually, the priori-based methods take advantage of the feature information obtained from the natural image to estimate the transmission map.

Tan et al [16] proposed the automated defogging algorithm. The method presented in [16] made use of two fundamental remarks. First, the contrast property present in a clean image is often higher than that of a foggy image. Second, in minute local surfaces of the image, the airlight changes easily. In their work, Tan et al. converted the input image into white color utilizing the white balance operation. Furthermore, the authors developed the airlight model using the Markov random field. One can maximize the local contrast of the recovered image to determine the airlight. With no human interaction, the automated defogging algorithm is capable of improving the visibility of a foggy image automatically. Knowing the peak intensity of the input image can help to determine the atmospheric light A. Despite the automatic defogging capability of this method, it exhibits some limitations. One of the limitations is color distortion in recovered images because the method does not consider color restoration when enhancing images. Another shortcoming is the presence of the halo effect in recovered images.

In [17], He et al. proposed the dark-channel prior (DCP) method to address the drawbacks mentioned in the methods discussed above. The dark-channel priori method has demonstrated a remarkable capability to restore outdoor images. In their work, He et al. examined a great number of clean outdoor images. The authors observed that oftentimes large portion of these clean outdoor images has a channel of pixels excluding those of the sky area and white area. The aim of the dark channel prior theory proposed in [17] was to determine the image restoration transmission map using the min operation in the local area. The recovered image (or restored image) contains block artifacts (also known as halo artifacts) as a result of the min filtering utilized in the local area of the dark channel image [18]. To address the problem of halo artifacts in the recovered image, first, the authors chose the local area in the dark channel image containing the highest 0.1% clearest pixels. Second, the authors represented atmospheric light A with the pixel which had the maximum intensity of the original foggy image. Thus, the recovered image J was obtained in equation (3) below [17]:

where t is the transmission using soft matting and t_0 represents a modest constant value that helps to avoid zero denominators.

$$J = \frac{I(x) - A}{\max\left(t(x), t_0\right)} + A \tag{3}$$

Despite the outstanding performance of the dark channel prior, the theory is faced with a host of limitations. First, the dark channel prior can be inefficient when operating on images with large sky areas, large white areas, or dense fog and inhomogeneous fog [18]. Another shortcoming is the soft matting used for estimating the transmission. The soft matting process can be time-taking and it can be impracticable in real-world applications.

Learning-Based: Recently, several single image defogging algorithms based on learning-based methods have been proposed in the literature. Learning-based methods primarily employ CNN-based or Generative Adversarial Networks (GANs)-based algorithms to recover fog-free images. In [19], Tang et al. suggested a learning-based algorithm that improves the accuracy of estimating the transmission map and trained the proposed algorithm using random forest. In [20], Mai et al. discovered a substantially linear relationship between the RGB color feature of hazy images and scene depth. Using back-propagation, the authors formed the inherent correlation between the color feature of the hazy image and the scene depth to restore the scene depth.

In [21], Cai et al. suggested dehazenet that improves transmission and learn various characteristics of color in a foggy image (which include color fading, maximum contrast, dark primary color, etc.) using CNN. Ren et al [22], proposed multi-scale convolutional neural networks (MSCNN) made up of two sub-networks namely coarse-scale and fine-scale for estimating transmission map. The purpose of the coarse-scale network is to determine the transmission map while the fine-scale network optimizes the transmission locally. Li et al [23], proposed a dehazing learning-based method that redevelops the atmospheric scattering model to generates clean images from hazy images.

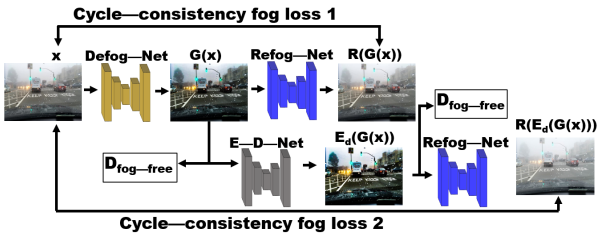
Goodfellow et al. [24], suggested GANs to synthesize natural images by efficiently learning the probability distribution of the training datasets (images). In their work, Goodfellow et al. implemented the concept of two-player min-max game optimization to simultaneously train both the generative and discriminative models G and D respectively. The authors believe that representing both models as multilayer perceptrons is the most simplified way to implement the adversarial modeling framework. GANs aims to train generative model G to produce samples from training distribution of the dataset in a way that the synthesized samples are identical to real distribution by the discriminator D.

To learn an efficient generator G with the aim to fool the learned discriminator D, such that the discriminator D is adequately efficient to identify fake images from real images, alternatively updating the models G and D is inevitable. Consider a real image x and a random noise z, the goal of GAN is to learn a mapping function that produces output image y using the using an adversarial loss expressed in equation (4) [24]:

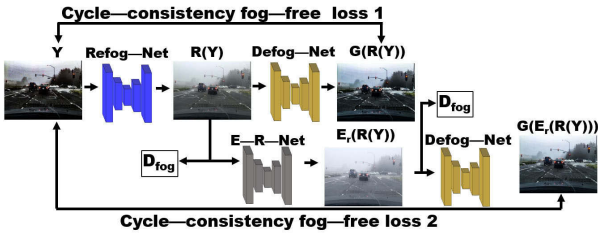
$$L_{GAN} = \min_{G}\max_{D} \mathrm{E}_{x\sim p_{\mathrm{data}}(x)}[\log D(x)] + \mathrm{E}_{z\sim p_z(z)}\left[\log\left(1 - D\big(G(z)\big)\right)\right] \tag{4}$$

In contrast to Generative Stochastic Networks [25] which generate samples using a Markov chain, GANs employs standard gradient descent methods [24].

At the earlier stage, GANs was faced with several limitations. One of the major shortcomings of GANs is that they can become unstable during training. This unstable behavior can



(a)



(b)

Fig. 3. The architecture of cycledefog2refog: (a) Defog architecture, Refog architecture [1].

cause artifacts in synthesized images. In [26], Radford et al suggested Deep Convolutional GANs that provide a set of constraints to solve the problem of instability. Another shortcoming of GANs is the lack of control on the kinds of data the generator produces especially under unconditioned generative models. To address this issue, Mirza et al. [27], included supplementary conditional variables that ensure efficient and stable learning of the generative model.

In [28], Isola et al proposed a method that employed conditional adversarial networks to enhance image-to-image translation. Karacan et al. [29] proposed a deep GAN method to synthesize natural outdoor images under varying conditions.

With a large amount of training data, learning-based techniques can learn to map a foggy image to a fog-free image. The learning-based techniques largely rely on fog to fog-free paired image datasets to train their networks. This implies that for every foggy image there is a corresponding fog-free image (ground truth image) of the same scene. In reality, because of varying contrast and light intensity throughout the day, there is a limited amount of fog-free images that can correspond to a foggy image of the same scene.

More recently, learning-based solutions that can operate under unpaired image supervision and are independent of ground truth images have been developed [1, 30, 31]. Zhu et al proposed a Cycle-consistency GAN (CycleGAN)-based method [31] which used unpaired image-to-image translation to recover a clean image. CycleGAN uses "two-cycle consistency losses that capture the intuition that if we translate from one domain to the other and back again we should arrive at where we started" [31].Consider individual image x from

$$L_{cyc}(G,F) = \mathrm{E}_{x\sim p_{\mathrm{data}}(x)}[\| F(G(x)-x)\|_1] + \mathrm{E}_{y\sim p_{data}(y)}[\| G(F(y)-y)\|_1] \quad (5)$$

domain X, the image translation cycle should be able to recover the original image x, this refers to as the forward cycle-consistency loss. The forward cycle-consistency loss is such that $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ [31]. The same concept applies to individual images y from domain Y, this is refer to as backward cycle-consistency loss. The backward cycle-consistency loss is such that $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$. The cycle consistency loss is expressed as follows in equation (5) [31].

Engin et al [32] employed cycle-consistency and VGG perceptual losses to recover fog-free images. A noteworthy benefit of using CycleGAN is that it does not require paired fog-to-fog-free images to train the defog algorithm. Nevertheless, because of one stage mapping strategy, CycleGAN-based methods are susceptible to color distortion, low contrast, and loss of texture information after the removal of fog from a foggy image.

D. CNN-Based Object Detectors

The state-of-the-art CNN-based models used for object detection can be classified into (i) two-stage detectors and, (ii) one-stage detectors [33]. The two-stage detection algorithms perform objection detection in two stages. First, the Region Proposal Network (RPN) is employed to suggest object bounding boxes for the candidate targets. Second, the Region of Interest Pooling operation (RoI Pool) is employed for feature extraction to predict and identify the location and class of the targeted objects [34]. The two-stage object detectors include Region-Convolutional Neural Network (R-CNN) [35], Fast R-CNN [36], Faster R-CNN [37], Mask R-CNN [38], region-based fully convolutional network (R-FCN) [39], feature pyramid networks (FPN) [40], etc. Girshick et al [36] made significant contributions in the field of object detection and classification. The authors were the first to successfully employ deep learning in object detection tasks.

However, in one stage object detectors, the RPN is not required to generate proposed boxes for targets. Instead, the one-stage object detectors immediately predict the location and class of the targets from the input image. The one-stage object detectors are end-to-end algorithms and they include Single Shot Detection (SSD) [41], YOLO (You Only Look Once) [42], YOLOv2 (YOLO 9000) [43], YOLOv3 [44], and deeply supervised object detectors (DSOD) [45], etc. Redmon et al [42-44] proposed YOLO which can extract features from an input image and immediately predict bounding boxes and the class of the target. One-stage object detectors have a faster speed of object detection than the two-stage object detectors and can be implemented in real-time.

Several studies [12, 46-48] have analyzed the impact of adverse weather conditions on state-of-the-art CNN-based object detection algorithms. Results from these studies have shown that the performance of the object detectors can diminish rapidly under adverse weather conditions. For instance, Liu et al [46] conducted a study that analyzed how perception in foggy conditions impacts the detection recall. The authors presented a visual imaging model to help understand the influence of fog on perception and implemented the Faster R-CNN. Experimental results in [46] show that detection recall of 91.55% (sunny), 85.21% (light fog), 72.54%~64.79% (moderate fog), and less than 57.75% (heavy fog). Nonetheless, the performance of object detection algorithms under moderate and heavy fog conditions still needs improvement.

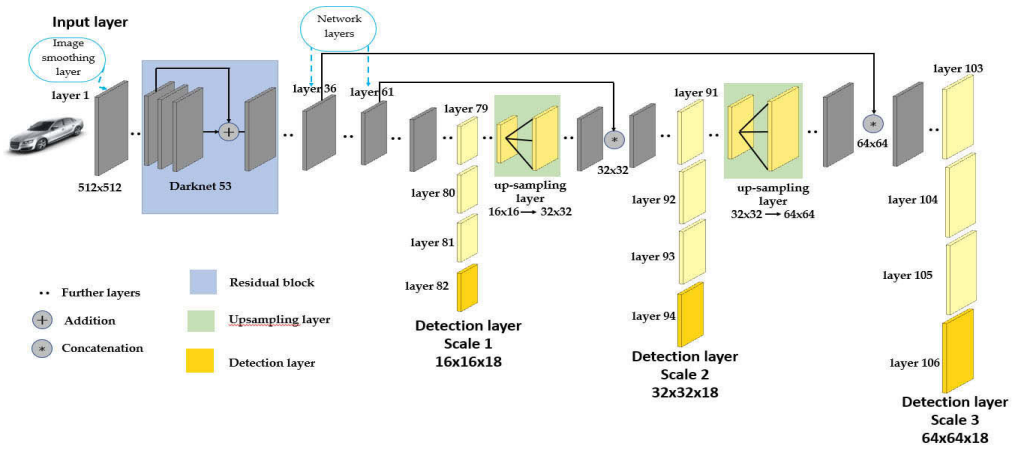


Fig. 4. YOLOv3 network architecture

III. Enhanced Cycle Consistent Adversarial Networks (Cycle-Defog2Refog Network)

In this section, we have adopted one of the most recently proposed CycleGAN based methods proposed by Liu et al [1] to recover clean images from foggy images. Liu et al proposed an end-to-end single image fog called Cycle-Defog2Refog network removal technique using enhanced cycle consistent adversarial networks. Instead of the one-stage mapping strategy used in traditional cycleGAN, Liu et al proposed the use of a two-stage mapping strategy. The architectures of the Cycle-Defog2Refog network are made up of two parts, the defog and refog architectures illustrated in Figures 3(a) and 3(b) respectively.

From Figure 3(a) X denotes the input foggy image, Defog-Net generator is denoted by G, and $G(X)$ is the recovered fog-free image. The Enhancer-Defog-Net(E-D-Net) generator which enhances the recovered fog-free is denoted by E_d. The adversarial discriminator $D_{fogfree}$ is employed to differentiate between the actual fog-free image and the recovered image after defogging. The defog architecture employs both the refog-net (R) and an enhancer-defog-net (E_d) with the aim to restrict the defogging mapping function with two consistency fog loss functions and an adversarial discriminator $D_{fogfree}$ [1].

From Figure 3(b), the clear image is represented by Y, Refog-Net generator is represented by R, and the synthetic foggy image is represented by $R(Y)$. The E_r which enhances the synthetic image is represents the generator Enhancer-Refog-Net(E-R-Net). The adversarial discriminator D_{fog} is employed to categorize the actual foggy image and the synthesized foggy image. The defog architecture employs both the defog-net (G) and an enhancer-refog-net(E_r) with the aim to control the refogging mapping function with two consistency fog-free loss functions and an adversarial discriminator D_{fog} [1].

IV. YOLO V3 Algorithm

YOLOv3 [44] network shown in Figure 4 is an enhanced version of YOLOv2 with multi-label classification capabilities. The multi-label classification capabilities enable YOLOv3 to accommodate more complex datasets with numerous overlapping targets [34]. YOLO splits the input image into multiple grids and employs three separate scale feature maps for predicting the bounding box of targets. The input dimensions are 16×16, 32×32, and 64×64. The purpose of the grid cell is to detect objects captured in its center, predict the bounding boxes, their confidence score, and the target class.

To perform object detection at 3 different scales, YOLOv3 employs 1 x 1 detection kernels and are implemented on feature maps of three different sizes placed at three different positions within the network. Output tensors from those detection layers have the same widths and heights as their inputs, but depth, which is the detection kernel is defined as $1\times1\times(B\times(5+C))$. Where B = number of bounding boxes; "5" is for the 4 bounding box coordinates and one object confidence, and C = the number of classes. Unlike YOLOv2 that uses Darknet-19 for feature extraction, in YOLOv3, a deeper and robust network called darknet-53 is used for feature extraction.

V. Result and Discussion

A. *Single Image Defogging Using Cycle-Defog2Refog Network*

We resized the training images to 512 × 512. Similarly, the testing images were resized to 512×512 for defogging. Nonetheless, for performance evaluation, we resized the defogged image back to its original size. The generator and discriminator were trained using ADAM optimizer (learning rate $= 2 \times 10^{-4}$, batch size $= 1$). We trained the cycle-defog2refog using TensorFlow on an NVidia GeForce RTX 2070 with Max-Q Design graphic processing unit.

In this experiment, we trained the cycle-defog2refog algorithm using the RESIDE dataset (which includes ITS and SOTS datasets) [49]. The ITS dataset consists of 100,000 synthetic indoor foggy images. The SOTS dataset includes 500 indoor foggy and outdoor foggy images each, their clear image ground truth. After training the cycle-defog2refog network, we implement the defogging network on the driving dataset. We feed the BDD100k deepdrive dataset into the network. It is important to note that the aim of defogging the

foggy images is to investigate whether defogging significantly

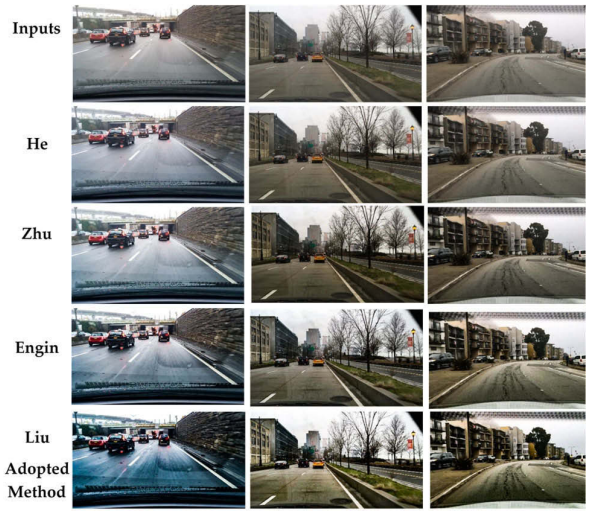


Fig. 5. Qualitative comparison of several methods on BDD100K data (moderate fog)

improves the performance of object detection algorithms in foggy conditions.

To measure the quality of the generated images we employed two evaluation standards. (i) Peak Signal-to-Noise Ratio (PSNR), (ii) Structural Similarity Index (SSIM). PSNR helps to the measure quality between the original image and the resulting image. A higher PSNR value implies a lower reconstruction error and an efficient reconstruction algorithm [1]. The measure of similarity between two different images. SSIM approximates perceptual image degradation using structural information change in the content of images [50].

Figure 5 illustrates defogged results for BDD100K moderate fog image data. The first three methods suffer some trace of artifacts and color distortion, low contrast. Both Zhu's and Engin's methods have a similar result with less trace of color detection than those of He's. Liu's method outperformed and generated clearer images than the other 3 methods. Liu's method has the highest SSIM and PSNR values of 0.9184 and 23.0672 respectively as shown in Table I.

Figure 6 illustrate defogged results for BDD100K heavy fog image. None of the four techniques were able to remove the heavy fog effectively. He's method has the least performance with halo-artifacts in the result. Although Liu's method has the highest performance among the four methods, yet the generated images were not as clear as expected. The SSIM and PSNR result (0.6104 and 15.0917 respectively) for Liu's method in Table II. Because the atmospheric degradation model in Liu's method inaccurately describes the fog map. This implies that a better atmospheric degradation model is required to effectively remove the fog on heavy fog images.

B. Object Detection Results Using YOLOv3

To train the YOLOv3, we use the KITTI dataset with 11040 train images, and 1380 validation images. We initialized the weight used in the network to COCO dataset [51].

TABLE I. AVERAGE PSNR AND SSIM OF DEFOGGED RESULT ON BDD100K DATA (MEDIUM FOG)

Metric	He [17]	Zhu [31]	Engin [32]	Liu [1]
SSIM	0.2971	0.7615	0.8052	0.9184
PSNR	10.2647	18.5241	20.6183	23.0672

TABLE II. AVERAGE PSNR AND SSIM OF DEFOGGED RESULT ON BDD100K DATA (HEAVY FOG)

Metric	He [17]	Zhu [31]	Engin [32]	Liu [1]
SSIM	0.1973	0.3725	0.3806	0.6104
PSNR	8.5218	11.3645	12.6581	15.0917

TABLE III. DETECTION PERFORMANCE OF YOLOV3 ON BDD100K DATASET.

Weather	Original Images		Defogged Images	
	Recall	*Precision*	*Recall*	*Precision*
No Fog	97.12%	98.84%	-	-
Moderate Fog	71.25%	71.67%	75.43%	76.31%
Heavy Fog	59.61%	60.98%	62.02 %	62.74%

Fig. 6. Qualitative comparison of several methods on BDD100K data (heavy fog)

With Python programming, we trained the YOLOv3 network using Pytorch framework on a computer with the following: Graphics card - Nvidia GeForce RTX 2070 with Max-Q Design; RAM - 16 gigabytes of memory; CPU - Intel Core 17-8570H 2.2 GHz 6 cores. Figure 7 illustrates the average loss curve of the trained network such that the training stops when there is no change decreasing trend.

We employed Recall and Precision to evaluate the performance of the object detection network on no fog, medium fog, and heavy fog conditions. Recall which is also referred to as sensitivity denotes the ratio of relevant instances that have been retrieved to the overall amount of relevant instances. Precision is categorized as positive predictive values (PPV) [47]. It denotes the proportion of positive results that are true positive. Figure 8 illustrates the detection results on real-world (BDD100K) data which include no fog images, and the defogged images (moderate and heavy fog). In Table

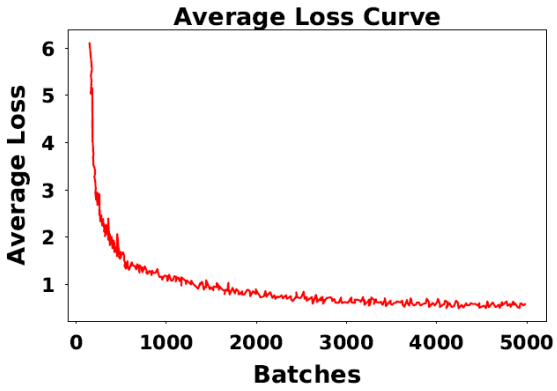


Fig. 7. Average loss curve

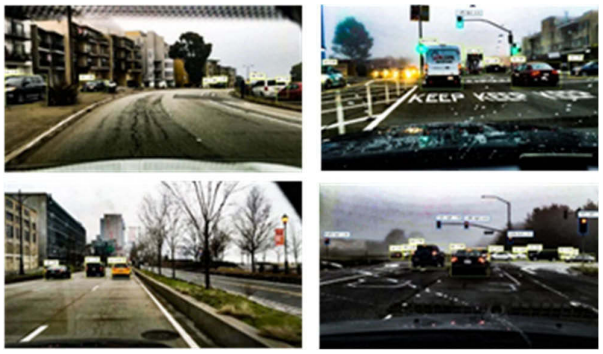


Fig. 8. Detection results on real world (BDD100K) data (a)no fog , (b) defogged (moderate fog) images, (c) defogged (heavy fog) images

III, we present detection recall and precision on the original images and defogged images (BDD100K data set). Under no fog images, the detection algorithm performed well with a detection recall of 97.12% and detection precision of 98.84%. However, detection recall and precision on moderate fog images improved by 5% after they were defogged. For heavy fog image, defogging the heavy fog images had a negligible improvement on the detection performance.

VI. Conclusion

Here, we have presented a brief review on how defogging and restoring the quality of foggy images can improve the performance of CNN-based real-time object detectors. The restoration of medium fog images using learning-based method significantly outperformed priori-based methods. As a result, we achieved a significant increase in detection recall and precision of the YOLOv3 object detector.

However, we observed that both the priori and learning based defogging algorithms including Liu's (recently proposed) struggled to restore heavy fog images. This is because the atmospheric degradation model in inaccurately described the fog map. Therefore, the recovered images contain halo-artifacts, color distortion, low contrast, and were not clear. This contributed to the low performance of the object detection algorithm on the heavy fog data.

This implies that a defogging algorithm with a better atmospheric degradation model is required to effectively restore heavy fog images and thus improve detection capabilities of object detectors under heavy fog condition.

References

[1] W. Liu, X. Hou, J. Duan, and G. J. I. T. o. I. P. Qiu, "End-to-end single image fog removal using enhanced cycle consistent adversarial networks," vol. 29, pp. 7819-7833, 2020.

[2] A. Geiger, P. Lenz, C. Stiller, and R. J. T. I. J. o. R. R. Urtasun, "Vision meets robotics: The kitti dataset," vol. 32, no. 11, pp. 1231-1237, 2013.

[3] NHTSA., "2018 Fatal Motor Vehicle Crashes: Overview," National Highway Traffic Safety Administration., Washington, DC2019.

[4] W. H. Organization, "Global status report on road safety 2018: Summary," World Health Organization2018.

[5] R. J. J. I. r. Davidse, "Older drivers and ADAS: Which systems improve road safety?," vol. 30, no. 1, pp. 6-20, 2006.

[6] K. HE, "Single Image Haze Removal Using Dark Channel Prior," Dissertation, The Chinese University of Hong Kong, 2011.

[7] "World Meteorological Organization. Manual on Codes," vol. volume I.2. , 2010.

[8] S. G. Narasimhan and S. K. J. I. j. o. c. v. Nayar, "Vision and the atmosphere," vol. 48, no. 3, pp. 233-254, 2002.

[9] H. J. B. z. P. d. f. A. Koschmieder, "Theorie der horizontalen Sichtweite," pp. 33-53, 1924.

[10] *Federal Meteorological Handbook Number 1: Chapter 8-Present Weather* (no., no. 2). Office of the Federal Coordinator for Meteorology, 2005.

[11] R. Younis and N. Bastaki, "Accelerated Fog Removal from Real Images for Car Detection," in *2017 9th IEEE-GCC Conference and Exhibition (GCCCE)*, 2017, pp. 1-6: IEEE.

[12] S. Zang, M. Ding, D. Smith, P. Tyler, T. Rakotoarivelo, and M. A. Kaafar, "The Impact of Adverse Weather Conditions on Autonomous Vehicles: How Rain, Snow, Fog, and Hail Affect the Performance of a Self-Driving Car," *IEEE Vehicular Technology Magazine,* vol. 14, no. 2, pp. 103-111, 2019.

[13] C. Dannheim, C. Icking, M. Mader, P. J. J. S. I. C. o. C. I. Sallis, Communication Systems, and Networks, "Weather Detection in Vehicles by Means of Camera and LIDAR Systems," pp. 186-191, 2014.

[14] L. Hespel, N. Riviere, T. Huet, B. Tanguy, and R. Ceolato, *Performance evaluation of laser scanners through the atmosphere with adverse condition* (SPIE Security + Defence). SPIE, 2011.

[15] A. Das, A. Ghasemzadeh, and M. M. J. J. o. s. r. Ahmed, "Analyzing the effect of fog weather conditions on driver lane-keeping performance using the SHRP2 naturalistic driving study data," vol. 68, pp. 71-80, 2019.

[16] R. T. Tan, "Visibility in bad weather from a single image," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8: IEEE.

[17] K. He, J. Sun, X. J. I. t. o. p. a. Tang, and m. intelligence, "Single image haze removal using dark channel prior," vol. 33, no. 12, pp. 2341-2353, 2010.

[18] Y. Xu, J. Wen, L. Fei, and Z. J. I. A. Zhang, "Review of video and image defogging algorithms and related studies on image restoration and enhancement," vol. 4, pp. 165-188, 2015.

[19] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image

dehazing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 2995-3000.

[20] J. Mai, Q. Zhu, D. Wu, Y. Xie, and L. Wang, "Back propagation neural network dehazing," in *2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014)*, 2014, pp. 1433-1438: IEEE.

[21] B. Cai, X. Xu, K. Jia, C. Qing, and D. J. I. T. o. I. P. Tao, "Dehazenet: An end-to-end system for single image haze removal," vol. 25, no. 11, pp. 5187-5198, 2016.

[22] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *European conference on computer vision*, 2016, pp. 154-169: Springer.

[23] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4770-4778.

[24] I. J. a. p. a. Goodfellow, "NIPS 2016 tutorial: Generative adversarial networks," 2016.

[25] J.-L. Starck, M. Elad, and D. L. J. I. t. o. i. p. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," vol. 14, no. 10, pp. 1570-1582, 2005.

[26] A. Radford, L. Metz, and S. J. a. p. a. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015.

[27] M. Mirza and S. J. a. p. a. Osindero, "Conditional generative adversarial nets," 2014.

[28] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125-1134.

[29] L. Karacan, Z. Akata, A. Erdem, and E. J. a. p. a. Erdem, "Learning to generate images of outdoor scenes from attributes and semantic layouts," 2016.

[30] Z. Anvari and V. J. a. p. a. Athitsos, "Dehaze-GLCGAN: Unpaired Single Image De-hazing via Adversarial Training," 2020.

[31] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223-2232.

[32] D. Engin, A. Genç, and H. Kemal Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 825-833.

[33] Z. Zhao, P. Zheng, S. Xu, and X. Wu, "Object Detection With Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 30, no. 11, pp. 3212-3232, 2019.

[34] L. Jiao *et al.*, "A survey of deep learning-based object detection," vol. 7, pp. 128837-128868, 2019.

[35] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.

[36] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440-1448.

[37] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 6, pp. 1137-1149, 2017.

[38] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980-2988.

[39] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," in *Advances in neural information processing systems*, 2016, pp. 379-387.

[40] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117-2125.

[41] W. Liu *et al.*, "Ssd: Single shot multibox detector," in *European conference on computer vision*, 2016, pp. 21-37: Springer.

[42] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[43] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263-7271.

[44] J. Redmon and A. J. a. p. a. Farhadi, "Yolov3: An incremental improvement," 2018.

[45] Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, and X. Xue, "Dsod: Learning deeply supervised object detectors from scratch," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1919-1927.

[46] Z. Liu, Y. He, C. Wang, and R. J. S. Song, "Analysis of the influence of foggy weather environment on the detection effect of machine vision obstacles," vol. 20, no. 2, p. 349, 2020.

[47] Y. Hamzeh, Z. El-Shair, and S. A. Rawashdeh, "Effect of Adherent Rain on Vision-Based Object Detection Algorithms," 2020. Available: https://doi.org/10.4271/2020-01-0104

[48] S. Hasirlioglu and A. Riener, "Challenges in Object Detection Under Rainy Weather Conditions," Cham, 2019, pp. 53-65: Springer International Publishing.

[49] B. Li *et al.*, "Reside: A benchmark for single image dehazing," vol. 1, 2017.

[50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. J. I. t. o. i. p. Simoncelli, "Image quality assessment: from error visibility to structural similarity," vol. 13, no. 4, pp. 600-612, 2004.

[51] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *European conference on computer vision*, 2014, pp. 740-755: Springer.