Advanced Multiple Linear Regression Based Dark Channel Prior Applied on Dehazing Image and Generating Synthetic Haze

Binghan Li[*,1], Yindong Hua[2], Mi Lu[1]

[1]*Department of Electrical & Computer Engineering, Texas A & M University, College Station, TX, 77840, USA*

[2]*Department of Electrical & Computer Engineering, Stony Brook University, Stony Brook, NY, 11794, USA*

A R T I C L E   I N F O

A B S T R A C T

Haze removal is an extremely challenging task, and object detection in the hazy environment has recently gained much attention due to the popularity of autonomous driving and traffic surveillance. In this work, the authors propose a multiple linear regression haze removal model based on a widely adopted dehazing algorithm named Dark Channel Prior. Training this model with a synthetic hazy dataset, the proposed model can reduce the unanticipated deviations generated from the rough estimations of transmission map and atmospheric light in Dark Channel Prior. To increase object detection accuracy in the hazy environment, the authors further present an algorithm to build a synthetic hazy COCO training dataset by generating the artificial haze to the MS COCO training dataset. The experimental results demonstrate that the proposed model obtains higher image quality and shares more similarity with ground truth images than most conventional pixel-based dehazing algorithms and neural network based haze-removal models. The authors also evaluate the mean average precision of Mask R-CNN when training the network with synthetic hazy COCO training dataset and preprocessing test hazy dataset by removing the haze with the proposed dehazing model. It turns out that both approaches can increase the object detection accuracy significantly and outperform most existing object detection models over hazy images.

1   Introduction

This paper is an extension of work initially presented in conference name [1]. Computer vision has recently played a major role in broad applications on urban traffic, such as autonomous and assisted driving, traffic surveillance, and security maintenance. However, the existence of haze, mist, dust, and fumes can severely degrade the visibility of images captured outside. Haze generates reduced contrasts, fainted surfaces, and color distortion to outdoor scenes, which will inevitably complicate many advanced computer vision tasks, including object classification and segmentation. Since the depth information of haze is non-linear and dependent over a global scene, haze removal becomes a challenging task. Most computer vision algorithms are designed based upon haze-free input images. They benefit a lot from haze removal, making it a highly desired task in computational photography and computer vision applications.

Many algorithms have been proposed to restore clear images from hazy images. Polarization-based methods presented in [2]–[3] analyze the polarization effects of atmospheric scattering and re-move haze through as few as two images with different degrees of polarization. Depth-based approaches introduced in [4], [5] evaluate depth information upon some assumptions or priors, then estimate transmission map $t(x)$ and atmospheric light A from it. Some recent CNN-based haze-removal models proposed in [6]–[7] are built upon various powerful CNNs to self-learn transmission map $t(x)$ directly from large-scale image datasets. Among effective conventional dehazing algorithms, Dark Channel Prior (DCP) [8] is generally accepted due to its novel prior and outstanding performance. In most non-sky patches of the haze-free image, at least one color channel contains dark pixels with extremely low intensity, which is primarily generated by the air light. However, the estimation on the medium transmission $t(x)$ and atmospheric light A is not precise, especially when the scene object is inherently similar to the air light over a large local region and no shadow is cast on it. And the restored image looks unnaturally dark when there is a sky region with sunlight.

In this paper, the authors propose a novel Multiple Linear Regression Dark Channel Prior based model (MLDCP). Trained with

[*]Corresponding Author: Binghan Li, Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX, 77840, USA, Email: libinghan1994@outlook.com,

the training dataset in REalistic Single Image DEhazing (RESIDE) [9], the MLDCP model can optimize the rough estimation of transmission map $t(x)$ and atmospheric light A by self-learning. The authors show experimentally on RESIDE test dataset that their model achieves the highest SSIM and PSNR values (two important full-reference metrics) compared with DCP and some other well-known state-of-the-art dehazing algorithms and CNN-based architectures. The authors further evaluate the effect on object detection in the hazy environment when dehazing the test images by their MLDCP model. Besides, the experimental results demonstrate that MLDCP not only enhances the performance of object detection in the hazy environment but also outperforms most dehazing algorithms on this task with higher detection accuracy.

The authors also present a straightforward and flexible algorithm to generate synthetic haze to any existing image datasets, inspired by a reversed MLDCP model. The authors aim to enhance object detection performance in the hazy environment by utilizing synthetic hazy images as training datasets. In the experiment, this algorithm is applied to MS COCO training dataset [10] by adding synthetic haze to the images and build a new Hazy-COCO training dataset. The authors evaluate the mean average precision (mAP) of Mask R-CNN [11], a widely adopted object detection and segmentation model, by training the network with the Hazy-COCO training dataset. The experimental results indicate that it leads to an impressive improvement when preprocessing training datasets with the inverse MLDCP algorithm.

2 Related Work

2.1 Overview of Dehazing Algorithms

2.1.1 Background Knowledge

In computer vision, the widely used atmospheric scattering model to describe the generation of a hazy image is as follows:

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

where $I(x)$ is the observed intensity (hazy image), $J(x)$ is the scene radiance (haze-free image), $t(x)$ is the medium transmission map, and A is the atmospheric light. The first term $J(x)t(x)$ is called attenuation and the second term $A(1 - t(x))$ is called airlight [12].

The medium transmission map $t(x)$ describes the portion of the light that is not scattered and reaches the camera [8]. When the atmosphere is homogeneous, the transmission matrix $t(x)$ can be defined as:

$$t(x) = e^{-\beta d(x)} \qquad (2)$$

where β is the scattering coefficient of the atmosphere, and $d(x)$ is the scene depth representing the distance between the object and camera.

Most state-of-the-art single image dehazing algorithms exploit the atmospheric scattering model (1) and estimate the transmission matrix $t(x)$ and the atmospheric light A in either physically grounded or data-driven ways. Then the haze-free images $J(x)$ can be recovered by computing the reformulation of (1):

$$J(x) = \frac{1}{t(x)}I(x) - A\frac{1}{t(x)} + A \qquad (3)$$

2.1.2 Conventional Single Image Dehazing Algorithms

Haze removal is a challenging task due to the non-linear and dependent depth maps over a global scene in hazy images. Many efforts have been made to tackle this challenge by exploiting natural images priors and depth statistics. Most conventional dehazing algorithms focus on predicting two critical parameters, medium transmission matrix $t(x)$ and global atmospheric light A, which are necessary to recover haze-free images via computing (3). In [12], an automated method is proposed based on the observation that the contrast of a haze-free image is higher than that of a hazy image. Furthermore, a Markov Random Fields (MRFs) framework is implemented to estimate the atmospheric light A by maximizing the local contrast of a hazy image. The output results are visually impressive but may not be physically valid. Assuming that the transmission and surface shading is uncorrelated in local areas, the authors eliminate the scattered light by locally estimating the optical transmission map of the scene with constant constraints in [13]. Despite its compelling results, it may fail in the cases with heavy haze and lead to the inaccurate estimation of color and depth maps.

A widely recognized single image dehazing algorithm called Dark Channel Prior (DCP) is proposed in [8], which can estimate the transmission map $t(x)$ more reliably. A regular pattern is found that in most non-sky patches of haze-free images, at least one color channel (dark channel) has some pixels whose intensity is very low and even close to zero. Then this pixel-based observation can be formally described by defining the dark channel J^{dark} as:

$$J^{dark}(x) = \min_{y \in \Omega(x)} (\min_c J^c(y)) \approx 0 \qquad (4)$$

where c indicates RGB color channels and y refers to the pixel in a local patch $\Omega(x)$ centered at x. Adding minimum operators to both sides of the transformation of (1):

$$\min_{y \in \Omega(x)} (\min_c \frac{I^c(y)}{A^c}) = \widetilde{t} \min_{y \in \Omega(x)} (\min_c \frac{J^c(y)}{A^c}) + 1 - \widetilde{t} \qquad (5)$$

Transmission map $\widetilde{t}$ can be put outside of the minimum operators based on the fact that $\widetilde{t}$ is a constant in the patch.

Since the dark channel of a haze-free image can be approximately taken as 0, the multiplicative term in (5) can be eliminated by adding (4). Then transmission map $\widetilde{t}$ can be predicted by:

$$\widetilde{t} = 1 - \omega \min_{y \in \Omega(x)} (\min_c \frac{I^c(y)}{A^c}) \qquad (6)$$

The additional parameter ω is a constant parameter that optionally controls the degree of haze removal. Even in a haze-free image, the haze still exists among distant objects. A small amount of haze will keep the vision perceptual natural with the sense of depth. The dehazing parameter determines how much haze will be removed.

In case that the recovered scene radiance $J(x)$ is prone to noise when transmission map $t(x)$ is extremely low, DCP restricts $t(x)$ by a lower bound t_0, which is set to 0.1 in [8]:

$$\widetilde{t} = \max(t(x), t_0) \qquad (7)$$

As for the estimation of atmospheric light A, it is defined as the color of the most haze-opaque regions in [12], which refers to the brightest pixels in a hazy image. However, this assumption only

applies when there is no sunlight in local regions. This limitation is optimized in [8] by considering the sunlight and adopting the dark channel to detect the most haze-opaque. DCP picks the top 0.1 percent brightest pixels in the dark channel, among which the pixels with the highest intensity in an input image are selected as the atmospheric light.

2.1.3 Limitations of DCP

DCP has some limitations that it may fail to accurately estimate transmission map $t(x)$ and atmospheric light A, when object surfaces are essentially analogous to the air light over a local scene without any projected shadows. Although DCP in [8] takes sunlight into consideration, the influence of sunlight is still tremendous when there is strong sunlight in the sky region. It will underestimate the transmission map of these objects and overestimate the haze layer. Thus the brightness of the restored image is darker than the real-world haze-free image. The authors compare a group of hazy images and recovered images from DCP, the color distortion in the sky region can be observed obviously in Figure 1:

(a) Real-world Hazy image	(b) Dehazed image via DCP
(c) Real-world Hazy image	(d) Dehazed image via DCP

Figure 1: Limitations of DCP

Additionally, the constant parameter ω in (6) is fixed to 0.95 in [8] without any changes corresponding to different haze distributions. Excessive haze removal will create color distortion, and images with insufficient haze removal will remain blurred. In order to solve this problem, the authors propose a dehazing parameter adaptive method in [14] based on DCP that estimates dehazing parameter ω locally instead of globally. It can automatically adjust the value of ω according to the distribution of haze. DCP in [8] implements soft matting to optimize the estimation of transmission map $t(x)$, which is further enhanced to be more accurate and ef-

ficient by utilizing an explicit image filter called *guided filter* in [15]. In [16], the authors point out that traditional DCP has not fully exploited its potential and will generate undesirable artifacts due to inappropriate assumptions or operations. Then it introduces a novel method that estimates transmission map $t(x)$ by energy minimization. The energy function combines DCP with piecewise smoothness and obtains an outstanding performance compared to conventional pixel-based dehazing algorithms. Several attempts have been made in [17], [18] to process the color distortion and optimize the restoration in some bright regions, such as the sky and reflective surfaces.

2.1.4 Overview of CNN-based Dehazing Algorithms

In recent years, neural network has made significant progress in numerous computer vision tasks [19], [20] and natural language processing tasks [21], [22]. Various Convolutional Neural Networks (CNNs) are designed to obtain a more accurate estimation of transmission matrix $t(x)$ by self-learning the mapping between hazy images and corresponding transmission maps, which outperform most conventional dehazing algorithms. DehazeNet in [23] is an end-to-end system built upon a deep convolutional neural network whose layers are specially designed to embody established priors in haze removal. Furthermore, a novel non-linear activation function is executed to improve the quality of output recovered images. A multi-scale deep neural network is proposed in [7] to dehaze a single image, which consists of a coarse-scale stage that roughly predicts the transmission matrix $t(x)$ over a global view, and a fine-scale stage that refines the rough estimation locally. In [24], the DCP energy function is defined as the loss function in a fully-convolutional dilated residual network. Feeding the network with real-world outdoor images, it minimizes the loss function completely unsupervised during the training process. A light-weight CNN model called All-in-One Dehazing Network (AOD-Net) in [6] is designed based on a reformulated atmospheric scattering model. It generates the recovered images directly and can be widely embedded into other deep CNN models to enhance the performances of some high-level tasks over hazy images.

2.1.5 Dehazing Benchmark Datasets and Metrics

Traditional haze-removal algorithms used to evaluate and compare dehazing performances by merely presenting a group of hazy images and dehazed images restored from various dehazing algorithms. The enhancement of dehazing performance is expected to be observed from the visual comparison of images. However, it is not convincing to prove that a new dehazing algorithm outperforms other algorithms only from human eyes perception. Two widely adopted image metrics to evaluate and compare single image dehazing algorithms are PSNR and SSIM [25]. PSNR refers to the peak-signal-to-noise ratio, which is generally applied to evaluate the image quality. SSIM refers to the structural similarity index measure, a well-known metric to measure the similarity between two images. Since it is generally impossible to capture the same visual scene with and without haze, while all other environmental conditions stay identical, it is incredibly challenging to measure the SSIM value between a hazy image and its haze-free ground truth

image. Therefore, recent efforts have been made to create synthetic hazy images from haze-free images based on the depth information.

In [26], the authors build two sets of images without haze and with synthetic haze from both real-world camera captured scenes and synthetic scenes to evaluate the performance of the proposed visibility enhancement algorithm. They utilize the software to generate 66 synthetic images built upon a physically-based road environment. By obtaining the depth information, four different fog types are added to 10 camera images and finally create a dataset with over 400 synthetic images in total. A fog simulation is proposed in [27] by simulating the underlying mechanism of hazy image formation (1) and utilizing the standard optical model for daytime haze. The fog simulation pipeline is leveraged to add synthetic fog to urban scenes images in Cityscapes dataset [28] and generate a Foggy Cityscapes dataset. Foggy Cityscapes dataset consists of 550 refined high-quality synthetic foggy images with detailed semantic annotations and additional 20000 synthetic foggy images without sufficient annotations.

The REalistic Single Image DEhazing (RESIDE) dataset [9] is the first large-scale dataset for benchmarking single image dehazing algorithms, and it includes both indoor and outdoor hazy images. RESIDE dataset also contains a large-scale synthetic training set and two sets designed respectively for objective and subjective quality evaluations. Moreover, in the supplementary RESIDE-β set, they add annotations and object bounding boxes to an evaluation set consisting of 4322 real-world hazy images, which can be utilized to test the performance of object detection in the hazy environment. A rich variety of criteria beyond PSNR and SSIM is also provided in [9] to evaluate the performance of dehazing algorithms, including full-reference metrics, no-reference metrics, subjective evaluation, and task-driven evaluation. However, the criteria are only practically applicable to the global performance of haze-removal. It cannot embody the difference locally between two images, failing to judge if our MLDCP model outperforms traditional DCP specifically in the bright region. Therefore, the authors compare dehazing performances experimentally with both recovered images' visual quality and two pivotal metrics PSNR and SSIM.

2.2 Overview of Object Detection Models

Object detection is the combination of object classification and localization, which can both recognize and localize all object instances of specific categories in an image. Due to its close relationship with image and video analysis, object detection has been widely applied in various computer vision tasks, especially in autonomous driving, traffic surveillance, and some other smart city fields.

Fast R-CNN refers to Fast Region-based Convolutional Network [29]. The network is fed with an image and a set of object proposals and outputs a convolutional feature map. A region of interest (RoI) pooling layer is proposed to extract a fixed-length feature vector and feed it into a sequence of fully connected layers. The output layers contain a softmax layer that estimates the softmax probability over K object classes plus a background class, and another layer with offset values that refine the bounding box positions of an object. Fast R-CNN overcomes the disadvantages of R-CNN and SPPnet [30] with a higher detection accuracy as well as faster training and test speed. Later in [31], a Region Proposal Network

(RPN) is added into Fast R-CNN generating a Faster R-CNN. The RPN aims to simultaneously propose candidate object bounding boxes and corresponding scores at each position. It then implements RoIPool in Fast R-CNN to extract features from each candidate box and perform classification as well as bounding-box regression. Both RPN and Fast R-CNN are trained independently but share the same convolutional layers. While achieving state-of-the-art object detection accuracy, the unified network can further increase the speed significantly.

Mask R-CNN [11] extends the Faster R-CNN by an output branch with a binary mask for each RoI, in parallel to a branch for bounding-box recognition and classification. However, the quantization of RoIPool in Faster R-CNN has a negative effect on predicting binary masks. To remove this harsh quantization, Mask R-CNN replaces RoIPool with a RoIAligh layer that aligns the extracted features with the input properly. It can efficiently detect object instances in an image and simultaneously generate a pixel-accurate segmentation mask for each instance. Due to its compelling performance and influential architecture, Mask R-CNN is widely used as a solid baseline to exploit more object detection tasks.

2.3 Domain Adaption Methods

Domain adaptation is a novel strategy that can be utilized to advance object detection models [32]–[33]. And it has been proved effective, especially in some extreme weather [34], [35], such as hazy, rainy, and snowy. A Domain Adaptive Faster R-CNN is proposed in [34] to enhance the cross-domain robustness of object detection. Based on H-divergence theory, two domain adaptation components on image level and instance level are integrated into Faster R-CNN architecture, aiming to reduce the domain discrepancy at both levels. Training data with images and full supervision is used as the source domain, and only unlabeled images in test data are available for the target domain. For both components, it adapts the classifier trained on a source domain and implements the adversarial training strategy to learn domain-invariant features. The method further incorporates a consistency regularization into the Faster R-CNN model to obtain a domain-invariant region proposal network (RPN).

Inspired by Domain Adaptive Faster R-CNN in [34], the authors in [36] adopt a similar idea and designs a Domain-Adaptive Mask-RCNN (DMask-RCNN). The source domain takes the clean images in the MS COCO dataset, and the target domain takes unannotated real-world hazy images in RESIDE dataset [9], and their dehazed output images by MSCNN [7] respectively. And DMask-RCNN adds a domain-adaptive branch after the base feature extraction layers in Mask R-CNN architecture, aiming to mask the generated features to be domain-invariant between the source domain and target domain. The experimental results in [34], [36] demonstrate that the domain adaptation method can enhance the performance of both Faster R-CNN and Mask R-CNN models when tackling the object detection task in the hazy environment. Moreover, this enhancement can be more effective when feeding the target domain with images restored by a robust dehazing algorithm.

2.4 Object Detection Datasets

Object recognition is a core task to understand a visual scene in computer vision, which involves several sub-tasks, including image classification, object detection, and semantic segmentation. All three tasks have high demands for image datasets. Object classification requires each image to be labeled with a sequence of binary numbers, which indicate if object instances exist in the image or not. Object detection is more challenging, which combines object classification and localization. It not only identifies which specific class an object belongs to but also locates it in an image. The object localization requires collecting bounding boxes that locate one or more objects in an image, which is a considerable workload in a large-scale image dataset. The PASCAL VOC challenge [37] is a widely recognized benchmark in visual object recognition and detection. The VOC2007 dataset contains 20 object categories spread over 11000 images. The annotation procedure is designed to be consistent, accurate, and exhaustive, and annotations are made available for training, validation, and test data. The ImageNet dataset [38] involves millions of cleanly labeled and full-resolution images in the hierarchical structure of WordNet [39]. The dataset provides an average of 500-1000 images to illustrate each synset in a majority of 80000 synsets of WordNet. Many object recognition algorithms have made a significant breakthrough by using the training resource of the ImageNet dataset. This benchmark dataset also plays an essential role in advancing object recognition tasks.

Semantic segmentation requires labeling each pixel in an image to a category, and it is an extremely time-consuming task to build a large-scale dataset with detailed semantic scene labeling. The Microsoft COCO dataset [10] consists of about 3.3 million images, and over 2 million of them are labeled. It collects 91 common object categories, which are fewer than the ImageNet dataset. However, in contrast to both the VOC2007 dataset and ImageNet dataset, MS COCO dataset contains significantly more instances per category and considerably more object instances per image. And every instance of each object category is fully labeled and segmented by a novel instance-level segmentation mask. MS COCO dataset has been widely utilized for training some more complex CNN architectures that aim to make further progress on object recognition as well as semantic segmentation tasks.

3 Proposed Methods

3.1 Multiple Linear Regression DCP Model

DCP estimates transmission map $t(x)$ from (6) and selects the pixels in the dark channel with the highest intensity among the top 0.1 percent brightest pixels as the atmospheric light A. Then hazy images can be recovered from (3) motivated by the estimations of $t(x)$ and A. However, the rough estimations in DCP introduced in Section 2.1.3 can generate some unpredicted deviations, which are theoretically impossible to be eliminated during the estimation process. Now the authors implement a multiple linear regression model to optimize the haze-removal algorithm in DCP.

Multiple linear regression is a statistical technique using several explanatory variables to predict the output response variable. As a predictive analysis, multiple linear regression aims to model the linear relationship between two or more independent variables and a continuous dependent variable. The authors still adopt the rough estimations on transmission map $t(x)$ and atmospheric light A in DCP. Three components $\frac{I(x)}{t(x)}$, $\frac{A}{t(x)}$ and A in (3) can be regarded as explanatory variables, since $I(x)$ refers to the pixel of input hazy image while $t(x)$ and A can be estimated from DCP. The scene radiance $J(x)$, which refers to the pixel of output recovered image, can be regarded as the response variable. The authors add regression coefficient weights to each explanatory variable and a constant term (bias) to the atmospheric scattering model. Then (3) can be reformulated as a multiple linear regression model (8), which describes how the mean response $J(x)$ changes with explanatory variables:

$$J(x) = \omega_0 \frac{I(x)}{t(x)} + \omega_1 \frac{A}{t(x)} + \omega_2 A + b \tag{8}$$

As the authors introduced, DCP is already an effective dehazing algorithm even with the rough estimations on $t(x)$ and A. The authors implement a multiple linear regression model to optimize DCP by learning the relationships between hazy images with haze-free images, which is extremely challenging to find out in traditional pixel-based dehazing algorithms. Since both $J(x)$ and A are defined and estimated on RGB color channels, the dimensions of three weights and a bias should be (3,1), intending to refine the parameters at all three color channels. The Outdoor Training Set (OTS) in RESIDE dataset [9] provides 8970 outdoor haze-free images, each of which also contains 30 synthetic hazy images with haze intensity ranging from low to high. During the training process, the network learns the relationships between output recovered images and their relevant haze-free ground truth images. However, it is theoretically impossible to obtain the ground truth image of a real-world hazy image. The most efficient approach is referring to a real-world haze-free image as ground truth image and its synthetic hazy image as input image respectively. The various intensity of synthetic haze guarantees that the MLDCP model can be applied to more than a fixed haze intensity in real-world scenes.

When MLDCP model is trained on OTS dataset, the authors refer to haze-free images as target ground truth images J, refer to synthetic hazy images as input images I, and refer to the recovered images as output images J_ω. In order to simplify the formula (8) during the training process, the authors assign $x_0 = \frac{I(x)}{t(x)}$, $x_1 = \frac{A}{t(x)}$ and $x_1 = A$. Then (8) can be reformulated as:

$$J_\omega(x) = \omega_0 x_0 + \omega_1 x_1 + \omega_2 x_2 + b \tag{9}$$

The deviations between the output images J_ω and the target ground truth images J are estimated by mean-squared error (MSE):

$$MSE = (J - J_\omega)^2 \tag{10}$$

MSE measures the average squared difference between an observation's actual and predicted values, which is commonly adopted in linear regression models. The loss function of MLDCP is evaluated based upon MSE:

$$L(\omega) = \frac{1}{2n} \sum_{i=1}^{n} (J^{(i)} - J_\omega^{(i)})^2 \tag{11}$$

During the training process, the authors calculate the gradient of loss function by Stochastic Gradient Descent (SGD), a widespread

and effective technique to address optimization tasks. Then the optimal weights and bias can be obtained by minimizing MSE through the following steps:

- Implement SGD to compute the derivative of loss function to find the gradient of the error generated from current weights and bias, and adjusts the weights by moving in the direction opposite of the gradient to decrease MSE. Then authors iteratively update three weights via:

$$\omega_k = \omega_k - \alpha \frac{\partial}{\partial \omega} L(\omega) \tag{12}$$

where α is the learning rate, and $k \in (0, 1, 2)$.

- Expand the derivative of loss function (11), then (12) can be reformulated as:

$$\omega_k = \omega_k - \alpha \frac{1}{n} \sum_{i=1}^{n} (J^{(i)} - J_\omega^{(i)}) x_k \tag{13}$$

- Update the bias with the same process repeatedly by training each image in OTS dataset via:

$$b = b - \alpha \frac{1}{n} \sum_{i=1}^{n} (J^{(i)} - J_\omega^{(i)}) \tag{14}$$

When the optimal weights and bias are obtained from the training process, they can be added to the haze-removal model (8) directly, and real-world hazy images can be dehazed in the same way as that in conventional DCP. The improved performance is determined by the quality of synthetic hazy images. If the synthetic haze is more close to the real-world haze, the MLDCP model can learn more realistic haze information during the training process, and the dehazing performance can be further enhanced.

3.2 Synthetic Haze Generated Model

There have been some large-scale image datasets released for the object detection task. However, few image datasets are built just for object detection in extreme weather like hazy and rainy. The Real-world Task-driven Testing Set (RTTS) in RESIDE [9] contains 4322 real-world hazy images annotated with both object categories and bounding boxes, which is an essential contribution taken as a test dataset for object detection task in the hazy environment. However, even with data augmentation techniques, its amount of training images is far less efficient in training an object detection model in comparison with the larger scale of the MS COCO training dataset [10]. It means a considerable workload to build a large-scale image dataset with detailed annotations and segmentation instances like MS COCO or PASCAL VOC. And the rarity of real-world hazy images, especially in urban cities, makes it more challenging to collect the same amount of high-quality hazy images as those in the MS COCO training dataset.

The authors propose an algorithm that generates synthetic haze to any existing large-scale haze-free image datasets without much computation. Inspired by the multiple linear regression haze-removal model, the authors propose a new algorithm that implements a similar inverse MLDCP model to add synthetic haze to an image. If an object detection model is trained on the synthetic hazy dataset generated by this algorithm, its performance of detecting an object in hazy weather can be significantly enhanced.

Since DCP is an effective dehazing algorithm, and the multiple linear regression model can further exploit its potential, the authors suppose that similar ideas can also be applied to synthesize higher quality hazy images based on the regular dark channel pattern. Coefficient weights and bias are added to (1), which is expected to maximally restore the deviations between haze-free images and their relevant synthetic hazy images. Then the reformulated atmospheric scattering model is obtained as follows:

$$I(x) = \beta_0 J(x) t(x) + \beta_1 A t(x) + \beta_2 A + b \tag{15}$$

The synthetic haze generated model is still trained on OTS in RESIDE dataset [9]. In contrast to MLDCP model, the authors refer to haze-free images as input images J and refer to synthetic hazy images in OTS as target images I. And the output synthetic hazy images are taken as I_β. The cost function is defined based on MSE that evaluates the deviations between the output images I_β and target images I:

$$L(\beta) = \frac{1}{2n} \sum_{i=1}^{n} (I^{(i)} - I_\beta^{(i)})^2 \tag{16}$$

The authors implement SGD during the training process to travel down the slope of cost function until it reaches the bottom lowest value, and three weights and a bias are updated iteratively by calculating the derivative of cost function as follows:

$$\beta_k = \beta_k - \alpha \frac{1}{n} \sum_{i=1}^{n} (I^{(i)} - I_\beta^{(i)}) x_k \tag{17}$$

$$b = b - \alpha \frac{1}{n} \sum_{i=1}^{n} (I^{(i)} - I_\beta^{(i)}) \tag{18}$$

where $k \in (0, 1, 2)$ and x_k equals to $J(\mathbf{x}) t(x)$, $A t(x)$, and A in (15) respectively.

In fact, traditional DCP does not support the inverse functionality of generating synthetic haze to a haze-free image. Self-learning the haze information from input synthetic hazy images, the MLDCP model fully utilizes the superiority of DCP to generate a much higher quality synthetic haze with a more similar visual and inherent effect to real-world haze. Additionally, the inverse MLDCP model can add various intensities of haze to an image dataset, making it possible to train an object detection model on training datasets with different synthetic haze intensities corresponding to the density of real-world haze in test images. Since the method focuses on the enhancement by preprocessing the training dataset, it can be applied simultaneously with any improvements on the object detection models or preprocessing images in test datasets with more effective dehazing algorithms.

4 Experiment Results and Analysis

4.1 MLDCP Dehazing Performance

4.1.1 Experiment Setup

In the experiment, both training and test datasets are obtained from RESIDE dataset [9]. In the training process, 8970 haze-free images

and their corresponding synthetic hazy images in the OTS dataset are utilized for training the MLDCP model. In the test phase, the Synthesis Object Testing Set (SOTS) is used to test the dehazing performance. Since MLDCP focuses on recovering outdoor hazy scenes, the authors do not evaluate its dehazing performance over the 500 synthetic indoor images in SOTS. The dehazing performance of MLDCP is compared with conventional DCP on both SSIM and PSNR metrics over 500 outdoor synthetic hazy images and their haze-free ground truth images.

4.1.2 Dehazing Performance on SOTS Dataset

In Figure 2, the authors present the comparison of recovered images by conventional DCP and our MLDCP model. It is obviously observed that the recovered image by DCP (c) suffers from the color distortion in the sky region. In contrast, the recovered image by MLDCP (d) performs competently even in the sky region with strong sunlight. And the non-sky region in (c) is much darker than that in (d), and it looks unreal. In comparison with the input synthetic hazy image (a), MLDCP almost removes the haze completely. Besides, its recovered image (d) looks natural and shares a high similarity with the ground truth haze-free image (b).

(a) Synthetic hazy image   (b) Haze-free image

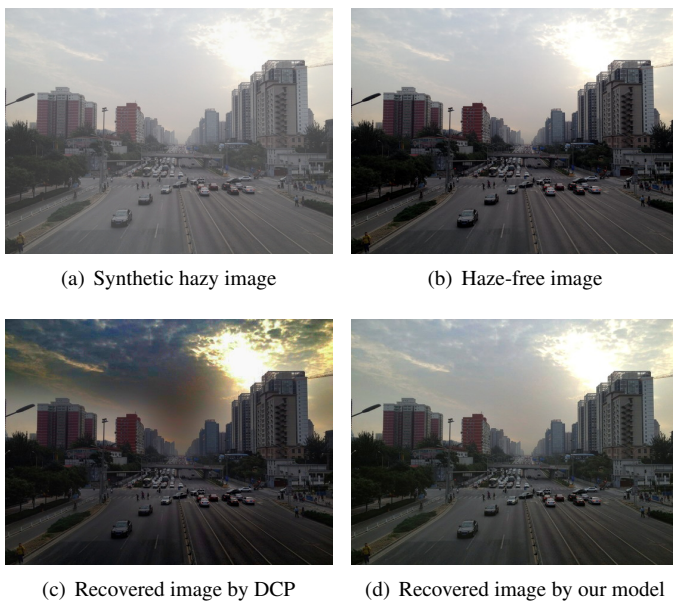(c) Recovered image by DCP   (d) Recovered image by our model

Figure 2: Comparison on synthetic hazy image. (a) is a synthetic hazy image in SOTS, (b) is the haze-free ground truth image, (c) is recovered by conventional DCP algorithm, (d) is recovered by our MLDCP model.

The authors compare the dehazing performance of MLDCP with several conventional dehazing algorithms as well as some CNN-based haze removal models over two major image benchmark metrics PSNR and SSIM. PSNR evaluates the quality of recovered images, while SSIM measures the similarity between recovered images and haze-free ground truth images. In Table 1, MLDCP model increases PSNR and SSIM by 5.3 and 0.23 respectively in comparison with the conventional DCP. Compared with the performances of other widely adopted dehazing algorithms obtained from [9], the proposed MLDCP model achieves a reasonably acceptable PSNR value and the highest SSIM value. Considering that SSIM

is essentially more dominant than PSNR when indicating the pixel-wise effect of haze removal, MLDCP outperforms all other dehazing algorithms in Table 1.

Table 1: Average SSIM and PSNR comparison among different dehazing algorithms over 500 outdoor synthetic hazy images in SOTS.

500 outdoor images		
Dehazing method name	PSNR	SSIM
Improved DCP model	23.84	**0.9411**
DCP	18.54	0.7100
FVR	16.61	0.7236
CAP	23.95	0.8692
NLD	19.52	0.7328
BCCR	17.71	0.7409
GRM	20.77	0.7617
DehazeNet	**26.84**	0.8264
MSCNN	21.73	0.8313
AOD-Net	24.08	0.8726

4.1.3 Real-world Hazy Images Dehazing

Even though synthetic benchmark datasets and metrics have achieved great success in comparing dehazing performances, the visual effect of synthetic haze is still different from the real-world haze. In Figure 3, the authors also present the recovered images by conventional DPC and MLDCP from real-world nature hazy scenes (first three rows) and hazy scenes in urban cities (last three rows). It is obviously observed that the MLDCP model performs better than conventional DCP in both sky and non-sky regions. The output images restored by MLDCP prevent the color distortion in sky regions, and the brightness over a global view seems more natural and closer to realistic haze-free scenes.

4.2 Object Detection with Pre-dehazed Test Dataset

4.2.1 Experiment Setup

Some object detection tasks on RESIDE RTTS dataset have been tested in [6], [9], [36]. The Domain-Adaptive Mask R-CNN model proposed in [36] achieves the highest detection accuracy on RTTS test dataset among several well-known object detection models including Faster R-CNN [31], Mask R-CNN [11], SSD [40] and RetinaNet [41]. A concatenation of dehazing algorithm and object detection modules is proposed in [6] to detect objects in the hazy environment. In [36], the authors further try more combinations of effective dehazing algorithms and object detectors in the cascade and evaluate their performances on the mean average precision (mAP) values. MLDCP model is utilized to dehaze the RTTS test dataset, and Mask R-CNN and DMask-RCNN modules are fed with the recovered images. Then the authors compare the detection accuracy with the results of dehazing-detection cascades in [36].

In Table 2, the proposed cascade of MLDCP dehazing model and Mask R-CNN or Mask R-CNN increases the detection accuracy by about 2% or 1.7% respectively, which is the highest in comparison with other dehazing approaches, including AOD-Net [6], MSCNN [7] and conventional DCP [8]. And pre-dehazing RTTS dataset with MSCNN and DCP can both enhance the performance of object

detection in the hazy environment. In [6], AOD-Net outperforms several dehazing algorithms on PSNR and SSIM values. However, it decreases the mAP result of DMask R-CNN from 61.72% to 60.47%, which means that a better dehazing performance on PSNR and SSIM values does not align with a higher accuracy on object detection in the hazy environment.
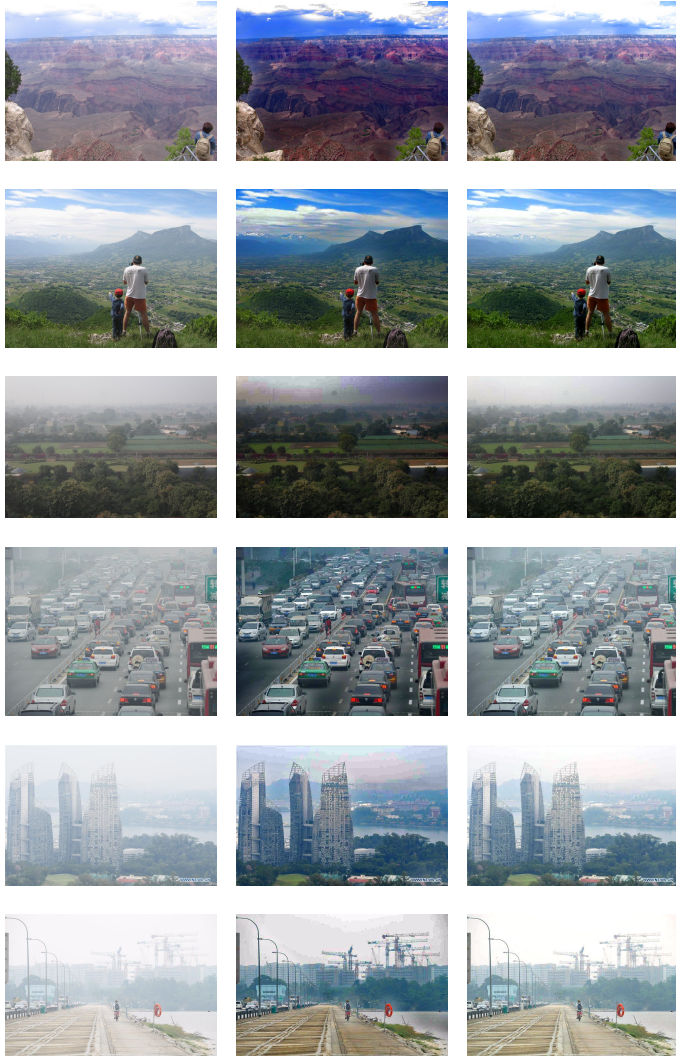

Figure 3: Visual comparison between real-world hazy images (first column), recovered images from DCP (second column) and recovered images from MLDCP (third column).

Table 2: Object detection accuracy (mAP) comparison among different dehazing-detection cascades

Framework	mAP(%)
Mask R-CNN	61.01
DMask R-CNN2	61.72
MLDCP + Mask R-CNN	**63.06**
AOD-Net + DMask R-CNN2	60.47
MSCNN + DMask R-CNN2	63.36
DCP + DMask R-CNN2	**62.78**
MLDCP + DMask R-CNN2	**63.42**

4.3 Object Detection with Hazy-COCO Dataset

4.3.1 Experiment Setup

The authors train Mask R-CNN on the new Hazy-COCO training dataset and evaluate its performance in comparison with the original Mask R-CNN trained on the MS COCO training dataset. During the training process, the backbone network is set as resnet101, and the training starts with the pre-trained COCO weights. The authors also tried various combinations of training stages to check the effect on its detection performance. It turns out that when training the head layers by ten epochs with a learning rate of 0.001 and fine-tuned layers from Resnet stage 4 and up by 60 epochs with a learning rate of 0.0001, the authors obtained the weights and bias with the highest detection accuracy on both RTTS and UG2 test datasets.
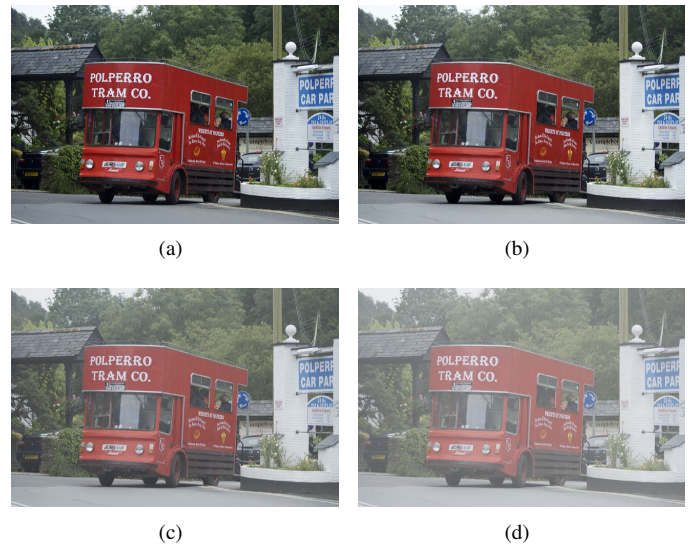

Figure 4: Comparison on synthetic hazy image. (a) is a haze-free image in MS COCO, (b) is the synthetic hazy image generated by inverse DCP, (c) is the synthetic hazy image with 0.1 haze density generated by our algorithm, (d) is the synthetic hazy image with 0.2 haze density generated by our algorithm.

Table 3: Object detection accuracy (mAP) comparison between MS COCO and Hazy-COCO training datasets

Test Dataset	Training Dataset	mAP(%)
RTTS	**COCO**	61.01
	Hazy COCO	**66.08**
RTTS+MLDCP	**COCO**	63.06
	Hazy COCO	**66.15**
UG2	**COCO**	32.07
	Hazy COCO	**38.53**

4.3.2 Experiment Results with Hazy-COCO Training Dataset

In Table.3, the authors evaluate the difference in detection performances when Mask R-CNN is pre-trained on MS COCO and Hazy-COCO datasets. The mAP results increase significantly by about 5% and 6% on RTTS and UG2 test datasets respectively, which presents how effective the Hazy-COCO training dataset is. Additionally, when training Mask R-CNN on the Hazy-COCO dataset, dehazing

Figure 5: Visual comparison between haze-free outdoor images from MS COCO dataset (left column) and the synthetic hazy images with 0.2 haze density generated by our algorithm (right column).

the RTTS test dataset by the MLDCP model can simultaneously increase the mAP result by 0.07%. However, this increment is much lower than the increment 2% when we preprocess the RTTS dataset individually. This is caused by the overlaps of enhancement effects between MLDCP and the advanced inverse DCP since they share some similarities over the dark channel regular pattern and multiple linear regression model.

4.3.3 Visual Effect of Synthetic Haze

Generally, traditional DCP does not support generating synthetic haze to a haze-free image by simply implementing the dark channel

regular pattern inversely. This limitation becomes possible after the authors apply the multiple linear regression techniques to the atmospheric scattering model.

In Figure 4, the effect of haze generated by the inverse DCP algorithm can be barely observed in (b), in comparison with the haze-free image (a). Synthetic hazy images created by our advanced inverse DCP algorithm with haze density 0.1 and 0.2 are presented as (c) and (d) respectively. Both of the two synthetic hazy images are covered with seemingly realistic haze. And it is obviously observed that the synthetic haze in (d) is denser than the haze in (c). The authors also present four more groups of haze-free images from the

MS COCO dataset and their corresponding synthetic hazy images in Figure 5 to demonstrate the effectiveness of our algorithm.

5 Conclusion

As autonomous driving and traffic surveillance become widespread in smart cities, object detection in extreme environments like hazy has been paid special attention to. This paper proposed the approaches that increase the detection accuracy in the hazy environment from both dehazing test dataset and preprocessing training dataset.

First, an advanced Multiple Linear Regression Haze-removal Model was proposed, aiming to overcome the deficiencies of Dark Channel Prior. The authors implemented Stochastic Gradient Descent to update and find the optimal weights and bias to refine the rough estimation of two essential parameters transmission matrix $t(x)$ and atmospheric light A. The experimental results showed that MLDCP not only achieved higher PSNR and SSIM values than other state-of-the-art dehazing algorithms and CNN-based dehazing models, but also increased the detection precision by a higher rate when concatenated with object detection models. It demonstrated that sometimes exploiting practical conventional dehazing algorithms by machine learning techniques was superior to building more complicated neural networks.

Second, the authors proposed an inverse DCP algorithm formulated on the multiple linear regression model that could generate synthetic haze to any existing image datasets. The authors expected this technique to prevent from spending excessive time and workload on building a large-scale image dataset with synthetic or real-world haze, as well as detailed annotations for object detection and segmentation. Synthetic haze was added to the MS COCO dataset and trained Mask R-CNN on this Hazy-COCO training set. The experimental results presented a significant increase in detection accuracy in the hazy environment. However, this approach could only generate the average synthetic haze with the same density over all pixels in an image. On the contrary, real-world haze density normally varies from pixel to pixel. In future work, it is expected to optimize this limitation and generate more realistic synthetic haze.

References

[1] B. Li, W. Zhang, M. Lu, "Multiple Linear Regression Haze-removal Model Based on Dark Channel Prior," in 2018 International Conference on Computational Science and Computational Intelligence (CSCI), 307–312, IEEE, 2018, doi:10.1109/CSCI46756.2018.00066.

[2] Y. Y. Schechner, S. G. Narasimhan, S. K. Nayar, "Instant dehazing of images using polarization," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, volume 1, I–I, IEEE, 2001, doi:10.1109/CVPR.2001.990493.

[3] M. Lu, et al., Arithmetic and logic in computer systems, volume 169, Wiley Online Library, 2004, doi:10.1002/0471728519.

[4] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, D. Lischinski, "Deep photo: Model-based photograph enhancement and viewing," ACM transactions on graphics (TOG), **27**(5), 1–10, 2008, doi:10.1145/1409060.1409069.

[5] S. G. Narasimhan, S. K. Nayar, "Interactive (de) weathering of an image using physical models," in IEEE Workshop on color and photometric Methods in computer Vision, volume 6, 1, France, 2003, doi:10.1109/ICIECS.2009.5365298.

[6] B. Li, X. Peng, Z. Wang, J. Xu, D. Feng, "An all-in-one network for dehazing and beyond," arXiv preprint arXiv:1707.06543, 2017.

[7] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in European conference on computer vision, 154–169, Springer, 2016, doi:10.1007/s11263-019-01235-8.

[8] K. He, J. Sun, X. Tang, "Single image haze removal using dark channel prior," IEEE transactions on pattern analysis and machine intelligence, **33**(12), 2341–2353, 2010, doi:10.1109/TPAMI.2010.168.

[9] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, Z. Wang, "Benchmarking single-image dehazing and beyond," IEEE Transactions on Image Processing, **28**(1), 492–505, 2018, doi:10.1109/TIP.2018.2867951.

[10] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, "Microsoft coco: Common objects in context," in European conference on computer vision, 740–755, Springer, 2014, doi:10.1007/978-3-319-10602-1_48.

[11] K. He, G. Gkioxari, P. Dollár, R. Girshick, "Mask r-cnn," in Proceedings of the IEEE international conference on computer vision, 2961–2969, 2017, doi:10.1109/ICCV.2017.322.

[12] R. T. Tan, "Visibility in bad weather from a single image," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, 1–8, IEEE, 2008, doi:10.1109/CVPR.2008.4587643.

[13] R. Fattal, "Single image dehazing," ACM transactions on graphics (TOG), **27**(3), 1–9, 2008, doi:10.1145/1360612.1360671.

[14] C. Chen, J. Li, S. Deng, F. Li, Q. Ling, "An adaptive image dehazing algorithm based on dark channel prior," in 2017 29th Chinese Control And Decision Conference (CCDC), 7472–7477, IEEE, 2017, doi:10.1109/CCDC.2017.7978537.

[15] K. He, J. Sun, X. Tang, "Guided image filtering," IEEE transactions on pattern analysis and machine intelligence, **35**(6), 1397–1409, 2012, doi:10.1007/978-3-642-15549-9_1.

[16] M. Zhu, B. He, Q. Wu, "Single image dehazing based on dark channel prior and energy minimization," IEEE Signal Processing Letters, **25**(2), 174–178, 2017, doi:10.1109/LSP.2017.2780886.

[17] J.-H. Kim, W.-D. Jang, J.-Y. Sim, C.-S. Kim, "Optimized contrast enhancement for real-time image and video dehazing," Journal of Visual Communication and Image Representation, **24**(3), 410–425, 2013, doi:10.1016/j.jvcir.2013.02.004.

[18] L. Shi, X. Cui, L. Yang, Z. Gai, S. Chu, J. Shi, "Image Haze Removal Using Dark Channel Prior and Inverse Image," in MATEC Web of Conferences, volume 75, 03008, EDP Sciences, 2016, doi:10.1051/matecconf/20167503008.

[19] Y. Hua, Y. Liu, B. Li, M. Lu, "Dilated Fully Convolutional Neural Network for Depth Estimation from a Single Image," in 2019 International Conference on Computational Science and Computational Intelligence (CSCI), 612–616, IEEE, 2019, doi:10.1109/CSCI49370.2019.00115.

[20] X. Wei, Y. Zhang, Z. Li, Y. Fu, X. Xue, "DeepSFM: Structure from motion via deep bundle adjustment," in European conference on computer vision, 230–247, Springer, 2020, doi:10.1007/978-3-030-58452-8_14.

[21] Y. Wang, X. Zhang, M. Lu, H. Wang, Y. Choe, "Attention augmentation with multi-residual in bidirectional LSTM," Neurocomputing, **385**, 340–347, 2020, doi:10.1016/j.neucom.2019.10.068.

[22] Y. Wang, Y. Huang, W. Zheng, Z. Zhou, D. Liu, M. Lu, "Combining convolutional neural network and self-adaptive algorithm to defeat synthetic multi-digit text-based CAPTCHA," in 2017 IEEE International Conference on Industrial Technology (ICIT), 980–985, IEEE, 2017, doi:10.1109/ICIT.2017.7915494.

[23] B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, "Dehazenet: An end-to-end system for single image haze removal," IEEE Transactions on Image Processing, **25**(11), 5187–5198, 2016, doi:10.1109/TIP.2016.2598681.

[24] A. Golts, D. Freedman, M. Elad, "Unsupervised single image dehazing using dark channel prior loss," IEEE Transactions on Image Processing, **29**, 2692–2701, 2019, doi:10.1109/TIP.2019.2952032.

[25] A. Hore, D. Ziou, "Image quality metrics: PSNR vs. SSIM," in 2010 20th international conference on pattern recognition, 2366–2369, IEEE, 2010, doi: 10.1109/ICPR.2010.579.

[26] J.-P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," IEEE Intelligent Transportation Systems Magazine, **4**(2), 6–20, 2012, doi:10.1109/MITS.2012.2189969.

[27] C. Sakaridis, D. Dai, L. Van Gool, "Semantic foggy scene understanding with synthetic data," International Journal of Computer Vision, **126**(9), 973–992, 2018, doi:10.1007/s11263-018-1072-8.

[28] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in Proceedings of the IEEE conference on computer vision and pattern recognition, 3213–3223, 2016, doi:10.1109/CVPR.2016.350.

[29] R. Girshick, "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, 1440–1448, 2015, doi:10.1109/ICCV.2015.169.

[30] K. He, X. Zhang, S. Ren, J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," IEEE transactions on pattern analysis and machine intelligence, **37**(9), 1904–1916, 2015, doi:10.1109/TPAMI.2015.2389824.

[31] S. Ren, K. He, R. Girshick, J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 91–99, 2015, doi:10.1109/TPAMI.2016.2577031.

[32] V. M. Patel, R. Gopalan, R. Li, R. Chellappa, "Visual domain adaptation: A survey of recent advances," IEEE signal processing magazine, **32**(3), 53–69, 2015, doi:10.1109/MSP.2014.2347059.

[33] R. Gopalan, R. Li, R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in 2011 international conference on computer vision, 999–1006, IEEE, 2011, doi:10.1109/ICCV.2011.6126344.

[34] Y. Chen, W. Li, C. Sakaridis, D. Dai, L. Van Gool, "Domain adaptive faster r-cnn for object detection in the wild," in Proceedings of the IEEE conference on computer vision and pattern recognition, 3339–3348, 2018, doi: 10.1109/CVPR.2018.00352.

[35] Z. Chen, X. Li, H. Zheng, H. Gao, H. Wang, "Domain adaptation and adaptive information fusion for object detection on foggy days," Sensors, **18**(10), 3286, 2018, doi:10.3390/s18103286.

[36] Y. Liu, G. Zhao, B. Gong, Y. Li, R. Raj, N. Goel, S. Kesav, S. Gottimukkala, Z. Wang, W. Ren, et al., "Improved techniques for learning to dehaze and beyond: A collective study," arXiv preprint arXiv:1807.00202, 2018, doi: arXiv:1807.00202v2.

[37] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman, "The pascal visual object classes (voc) challenge," International journal of computer vision, **88**(2), 303–338, 2010, doi:10.1007/s11263-009-0275-4.

[38] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 248–255, Ieee, 2009, doi:10.1109/CVPR.2009.5206848.

[39] G. A. Miller, WordNet: An electronic lexical database, MIT press, 1998, doi:10.7551/mitpress/7287.001.0001.

[40] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, "Ssd: Single shot multibox detector," in European conference on computer vision, 21–37, Springer, 2016, doi:10.1007/978-3-319-46448-0_2.

[41] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, "Focal loss for dense object detection," in Proceedings of the IEEE international conference on computer vision, 2980–2988, 2017, doi:10.1109/ICCV.2017.324.