Multiple Linear Regression Haze-removal Model Based on Dark Channel Prior

Binghan Li

226002451

1 Introduction

Images captured in outdoor scenes are usually degraded by haze, fog and smoke. Suffering from poor visibility, reduced contrasts, fainted surfaces and color shift, hazy images will miss many details. Haze removal is highly desired in computational photography and computer vision applications, for that many computer vision algorithms can only work well with haze-free images. With the development of self-driving vehicles, object detection in traffic has developed a lot in recent years. However, in some cities, the traffic environment suffering from heavy haze will reduce the accuracy of object detection. Because of the distribution difference between the dehazing images and the clean images, only focusing on dehazing algorithm or object detection algorithm is not the best way to improve the detection precision. In this project, our purpose is to tackle this cross-domain object detection problem. The first task is to boost single image dehazing performance as an image restoration problem. The second task is to improve object detection accuracy in the presence of haze. Both two tasks will use the REalistic Single Image DEhazing (RESIDE), which is a new large-scale benchmark consisting of both synthetic and real-world hazy images.

2 Overview of existing algorithms

2.1 Dark Channel Prior

In computer vision, the classical model to describe the generation of a hazy image is the atmospheric scattering model:

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

where $I(x)$ is the observed intensity, $J(x)$ is the scene radiance, A is the atmospheric light, and $t(x)$ is the medium transmission matrix. The first term $J(x)t(x)$ on the right-hand side is called direct attenuation, and the second term $A(1 - t(x))$ is called airlight. When the atmosphere is homogeneous, the transmission matrix $t(x)$ can be defined as:

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

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

Given the atmospheric scattering model (1), most state-of-the-art single image dehazing algorithms estimate the transmission matrix $t(x)$ and the global atmospheric light A. Then they recover the clean images $J(x)$ via computing the transformation of (1):

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

To formally describe the observation in DCP [4], the dark channel of an image J is defined as:

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

Then by minimizing both sides of equation (1), and putting (2) into (1), we can eliminate the multiplicative term and estimate the transmission $\widetilde{t}$ by:

$$\widetilde{t} = 1 - \min_{y\in\Omega(x)}(\min_{c} \frac{\boldsymbol{J}^{c}(y)}{\boldsymbol{A}^{c}}) \tag{5}$$

However, DCP has many limitations when estimating transmission map $t(x)$ and atmospheric light A, which make DCP fail in some specific cases that are very common in real world. When the scene object is inherently similar to the air light (e.g., snowy ground or a white wall) over a large local region and no shadow is cast on it, it will underestimate the transmission of these objects and overestimate the haze layer. So the brightness of the restored image will be darker than the real haze-free image, which often impacts object detection negatively.

2.2 Object detection method

Object detection model has developed a lot in recent years. The development from R-CNN to Fast R-CNN, from Fast R-CNN to Faster R-CNN, from Faster R-CNN to Mask R-CNN, shows that the deep model will become faster with higher accuracy. In [10], the authors proposed a cascade of AOD-Net dehazing and Faster-RCNN detection modules to detect object in hazy images. And trying different combinations of more powerful dehazing and object detection modules in the cascade. What's more, such a cascade could be subject to further joint optimization.

2.2.1 Fast R-CNN

Fast R-CNN is Fast Region-based Convolutional Network. A Fast R-CNN network takes as input as entire image and a set of object proposals. The network first processes the whole image with several convolutional and max pooling layers to produce a conv feature map. Each feature vector is fed into a sequence of fully connected layers that finally branch into two sibling output layers: one that produces softmax probability estimates over K object classes. Each set of 4 values encodes refined bounding-box positions for one of the K classes. [2]

2.2.2 Faster R-CNN

Faster R-CNN consists of two stages. The first stage, called a Region Proposal Network (RPN), proposes candidate object bounding boxes. The second stage, which is in essence Fast R-CNN [8], extracts features using RoIPool from each candidate box and performs classification and bounding-box regression. The features used by both stages can be shared for faster inference.

2.2.3 Mask R-CNN

Mask R-CNN extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existng branch for bounding box recognition and classification. It efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance. [3]

2.2.4 Domain-Adaptive Mask-RCNN

Domain Adaptive Faster R-CNN in [1] is based on the recent state-of-the-art Faster R-CNN model. It designs two domain adaptation components, on image level and instance level, to reduce the domain discrepancy. And inspired by the Domain Adaptive Faster R-CNN, [7] applied a similar approach to design a domain-adaptive mask-RCNN (DMask-RCNN). The primary goal of DMask-RCNN is to mask the features generated by feature extraction network to be as domain invariant as possible, between the source domain (clean input images) and the target domain (hazy images). Specifically, DMask-RCNN places a domain-adaptive component branch after the base feature extraction convolution layers of Mask-RCNN.

2.3 RESIDE dataset

The REalistic Single Image DEhazing (RESIDE) dataset [6] is the first large-scale dataset for benchmarking single image dehaziang algorithms and includes both indoor and outdoor hazy images. Further, RESIDE contains both synthetic and real-world hazy images, thereby highlighting diverse data sources and image contents. It is divided into five subsets, each serving different training or evaluation purposes. And the test sets address different evaluation viewpoints including restoration quality (PSNR, SSIM and no-reference metrics), subjective quality (rated by humans), and task-driven utility.

3 Multiple linear regression haze-removal model

After the transmission map $t(x)$ and the Atmospheric Light A are estimated, hazy image can be recovered by:

$$\boldsymbol{J}(x) = \frac{\boldsymbol{I}(x)}{t(x)} - \frac{\boldsymbol{A}}{t(x)} + A \tag{6}$$

In Section II, we already introduced the estimation of transmission map $t(x)$ and the Atmospheric Light A are both based on the hazy images. Although many algorithms attempt to refine the estimations on transmission map $t(x)$, and the Atmospheric Light A, estimations can still generate some unexpected deviations. And this kind of errors are normally impossible to eliminate. Now we introduce the multiple linear regression model to optimize the atmospheric scattering model (6). Multiple linear regression is the most common form of linear regression analysis. As a predictive analysis, multiple linear regression is used to explain the relationship between one continuous dependent variable and two or more independent variables. When we train with thousands of synthetic images and haze-free images, the scene radiance $J(x)$, which is also the RGB pixel of haze-free image, can be regarded as the the continuous dependent variable. For each image, after we estimate $t(x)$ and A by original DCP, $\frac{\boldsymbol{I}(x)}{t(x)}$, $\frac{\boldsymbol{A}}{t(x)}$ and A in (6) can be regarded as three independent variables, where $I(x)$ is the pixel of hazy image. Then both two parameters $t(x)$ and A together with the pixels of input images and target images can be simplified to a prediction problem with multiple linear regression model, which describes how the mean response $J(x)$ changes with the three explanatory variables:

$$\boldsymbol{J}(x) = w_0\frac{\boldsymbol{I}(x)}{t(x)} + w_1\frac{\boldsymbol{A}}{t(x)} + w_2\boldsymbol{A} + b \tag{7}$$

Then we implement Stochastic Gradient Descent (SGD) to our multiple linear regression model, which is one of the most widely used algorithms to solve optimization problems. The Outdoor Training Set (OTS) of RESIDE dataset provides 8970 outdoor haze-free images. For each haze-free image, OTS also provides about 30 synthesis hazy images with haze intensity from low to high. Different haze intensity means a lot to our multiple linear regression haze-removal model, because we don't expect that our algorithm can only perform good on hazy images with fixed haze intensity. When we train our model on OTS, we refer haze-free images as target images J, refer synthesis images as input images I, take the images recovered from our model as output images J_ω. Then (7) can be re-formulated as:

$$\boldsymbol{J}_\omega(x) = w_0\boldsymbol{x}_0 + w_1\boldsymbol{x}_1 + w_2\boldsymbol{x}_2 + b \tag{8}$$

For that $I(x)$ and A are both defined based on RGB color channels, the dimension of three weights and bias is (3,1), aiming to refine the parameters from all three color channels. And the deviations of the output images J_ω from the target images J are estimated by mean-squared error (MSE):

$$\boldsymbol{MSE} = (\boldsymbol{J} - \boldsymbol{J}_\omega)^2 \tag{9}$$

By training our model, we try to find the optimal weights and bias that minimize the mean-squared error.

- Define the cost function of our model based on (9):

$$\boldsymbol{L}(\omega) = \frac{1}{2n}\sum_{i=1}^{n}(\boldsymbol{J}^{(i)} - \boldsymbol{J}_\omega^{(i)})^2 \tag{10}$$

- For each image in training set, we repeatedly update three weights via:

$$\omega_k = \omega_k - \alpha\frac{\partial}{\partial\omega}\boldsymbol{L}(\omega)x_k \tag{11}$$

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

- Since the derivative of cost function can be computed via (10), Equation (11) can be simplified as:

$$\omega_k = \omega_k - \alpha\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{J}^{(i)} - \boldsymbol{J}_\omega^{(i)})x_k \tag{12}$$

- For each image in training set, we repeatedly update the bias via:

$$b = b - \alpha\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{J}^{(i)} - \boldsymbol{J}_\omega^{(i)}) \tag{13}$$

And after we get the optimal weights and bias, they can be simply added to traditional DCP. What's more, the multiple linear regression model will not modify the inner theory of any state-of-the-art dehazing algorithm. Our model can be applied to further improve the performance of some state-of-the-art dehazing algorithms that optimize the estimations of $t(x)$ and A.

4 Results

In this section, by demonstrating our dehazing results on several groups of hazy images as well as comparing the PSNR and SSIM value, we show the better performance of our model over DCP and other dehazing algorithms. And we combine our dehazing model with DMask R-CNN object detection model, to test the mAP value on RTTS dataset. Compared with the highest mAP value so far, our result is almost the same to it.

4.1 Experiment setup

Most recently, a benchmark dataset of both synthetic and real-world hazy images provided in [6] for dehazing problems are introduced to the community. In our experiment, the Synthesis Object Testing Set (SOTS) in RESIDE dataset is used to test the dehazing performance of our method. Our algorithm focuses on recovering outdoor hazy images, therefore, we won't evaluate the performance on recovering the 500 synthetic indoor images in SOTS. And the Real-world Task-Driven Testing Set (RTTS) is used to test the precision of object detection in haze.

In the training phase, 8000 haze-free images and corresponding synthetic hazy images in Outdoor Training Set (OTS)[6] are used to train our multiple linear regression haze-removal model. In the testing phase, 500 outdoor haze-free images and 500 outdoor synthetic hazy images in Synthetic Objective Testing Set(SOTS) are tested using the well trained model. We compare the performance with original DCP from both the average SSIM and PSNR, and we also list the comparison of recovered images between our algorithm and DCP to prove that the multiple linear regression haze-removal model can overcome the weakness of DCP.

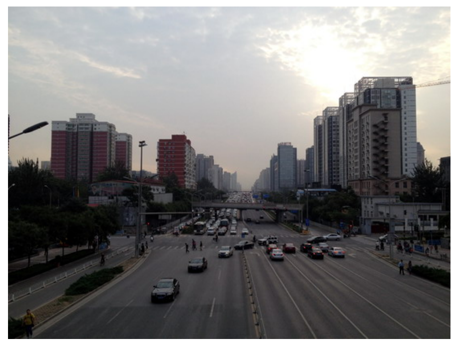
4.2 Experiment results on SOTS

In Fig. 1, We demonstrate the performance of multiple linear regression haze-removal model from the difference observed through recovered images via original DCP and our improved model. Fig. 1 (a) shows a synthetic hazy image selected from SOTS, and Fig. 1 (b) shows its haze-free image. Fig. 1 (c) shows the recovered image via original DCP. We can easily observe that the color distortion in sky region and non-sky region performs darker than its haze-free image. Fig. 1 (d) shows the recovered image via our model. Compared with the haze-free image, the haze is almost completely removed and the sky region seems more natural. The high similarity is obvious through observation.

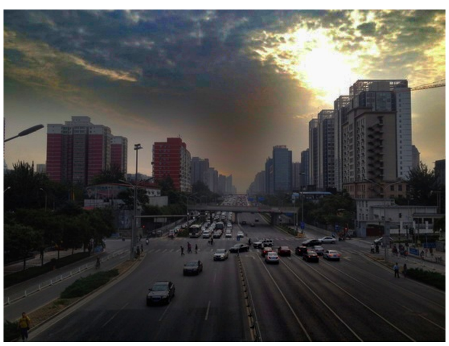
In Table 1, we compare our results with several state-of-the-art method with average SSIM and average PSNR. As we mentioned in Section III, PSNR and SSIM together can compare the image quality and similarity between recovered images and haze-free images. Since we want to apply our improved model on traffic dehazing problems, we only compare the results on 500 outdoor images in SOTS in this table. All the results of other algorithms come from [6]. When compared with well-recognized state-of-the-art dehazing algorithms and deep learning methods, our proposed model achieves reasonably good PSNR value and obtains the highest SSIM value. The results show the effectiveness of our model on dehazing outdoor images.

The comparison of dehazing real world hazy images via DCP and our model are shown in Figure 2. We can obviously observe the improvements when using the Multiple Linear Regression Model. The improved DCP model can not only perform good on PSNR and SSIM value, but also perform good on dehazing real world hazy images.

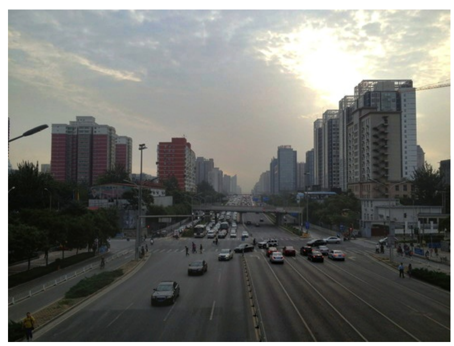
(a) Synthetic hazy image



(b) Haze-free image



(c) Recovered image by DCP



(d) Recovered image by our model

Figure 1: Comparison on synthetic hazy image

(a) Hazy image (1)

(b) Hazy image (2)

(c) Recovered image (1) via DCP

(d) Recovered image (2) of via DCP

(e) Recovered image (1) via our model

(f) Recovered image (2) via our model

Figure 2: Comparison on real-world nature hazy images

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

Table 1: Average SSIM and PSNR comparison between different dehazing methods on 500 outdoor synthetic image in SOTS.

4.3 Experiment results on RTTS

Some experiments on object detection in haze based on RTTS dataset have been made in [5] [6] [7]. From the comparison of the mAP value among Faster R-CNN, Mask R-CNN and Domain-Adaptive Mask-RCNN combined withe several dehazing algorithms, we pick the Domain-Adaptive Mask-RCNN in [7] with the best performance. Firstly, we use the original DCP with Guided Filter and our Multiple Linear Regression Model with the best performance to dehaze each hazy image in RTTS dataset separately and save the dehazed images into two folders named "result" and "9411result". We use the pretrained Domain-Adaptive Mask-RCNN to run the two dataset we added.

From the results in Table 2, after using the best performance Multiple Linear Regression Model to dehaze images in RTTS, we get mAP value that is only 0.1% lower than that of "MSCNN + DMask R-CNN2". And the mAP value of "MSCNN + DMask R-CNN2" is the highest among all object detection in haze models that tested on RTTS dataset.

However, compared with the PSNR and SSIM values of MSCNN in the last experiment, Multiple Linear Regression Model and AOD-Net outperform the MSCNN a lot. We can simply conclude that the higher PSNR value and SSIM value don't equal to higher mAP value on object detection in haze. From the previous discussion about the limitaion of original DCP, we know the dehazed

mAP values	
Framework	**mAP(%)**
Mask R-CNN	61.01
DMask R-CNN1	61.21
DMask R-CNN2	61.72
AOD-Net + DMask R-CNN1	60.21
AOD-Net + DMask R-CNN2	60.47
MSCNN + DMask R-CNN1	62.71
MSCNN + DMask R-CNN2	63.36
DCP + DMask R-CNN2	**62.78**
Multiple Linear Regression Model + DMask R-CNN2	**63.25**

Table 2: mAP values comparison among different models

image will be darker than normal scenes due to the rough estimations of $t(x)$. I used to suspect that will reduce the precision of object detection a lot. From the results in Table 2, orignal DCP with Guided Filter even performs much better than AOD-Net. One advantage of DCP is that it can remove heavy haze. We have a assumption that the amount of haze removed is more important than the effect of restored image becoming darker when we apply dehazing algorithm with object detection model. That can explain why original DCP with worse performance can outperform AOD-Net with better performance on dehazing.

5 Conclusion and Discussion

From the three experiments, we conclude that The Multiple Linear Regression DCP Model gets the highest SSIM value and reasonable PSNR value, performs very good on dehazing real world hazy images and gets almost the highest mAP value when combined with object detection model. Overall, it's much better than other dehazing model so far. Since the MSCNN dehazing model gets the highest mAP value on object detection in haze experiment with worse performance on PSNR value and SSIM value compared with AOD-Net and our improved model, we can not only focusing on improving the PSNR value and SSIM value of an dehazing algorithm. We also need to study what kind of quality of dehazed images can boost the performance on object detection in haze.

References

[1] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc 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, pages 3339–3348, 2018.

[2] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[3] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[4] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2011.

[5] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 1, page 7, 2017.

[6] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Reside: A benchmark for single image dehazing. *arXiv preprint arXiv:1712.04143*, 2017.

[7] Yu Liu, Guanlong Zhao, Boyuan Gong, Yang Li, Ritu Raj, Niraj Goel, Satya Kesav, Sandeep Gottimukkala, Zhangyang Wang, Wenqi Ren, et al. Improved techniques for learning to dehaze and beyond: A collective study. *arXiv preprint arXiv:1807.00202*, 2018.

[8] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.