Can Tho University Journal of Science
website: ctujs.ctu.edu.vn

DOI: 10.22144/ctu.jen.2022.010

An object detection method for aerial hazy images

Tran Tuan Minh*, Tran Van Bao, Vo Duy Nguyen and Nguyen Tan Tran Minh Khang
University of Information Technology, VNU-HCM, Viet Nam
Vietnam National University, Ho Chi Minh City, Viet Nam
**Correspondence: Tran Tuan Minh (email: 18520314@gm.uit.edu.vn)*

Article info.

Received 01 Sep 2021
Revised 11 Sep 2021
Accepted 13 Dec 2021

Keywords

Object detection, dehazing, aerial images, deep learning

ABSTRACT

Image processing and object detection in aerial images have to deal with a lot of trouble due to the existence of haze, smoke, dust in the atmosphere. These factors can blur objects and severely decline image quality which might lead to incorrect or missing object detection. To solve this problem, this study shows a method that can reduce the bad effect of haze on object detection in aerial images. A combination of a dehazing method called Feature Fusion Attention Network (FFA-Net) and an object detection method named Probabilistic Anchor Assignment (PAA) was conducted to evaluate two hypotheses: (1) haze was a noisy factor and (2) haze was treated as part of objects. Through extensive experiments, the selective dehazing hypothesis, which was used for truck objects, improved the detection result of car and bus from 19.6% to 21.9% and 0.7% to 4.4%, respectively, on the UAVDT-Benchmark-M dataset. This result showed that our approach was effective.

1. INTRODUCTION

Object detection is one of the core steps of analyzing videos collected from unmanned aerial vehicles (UAVs) that have practical implementation such as security cameras and rescue detecting systems. Detecting objects in aerial photos has to deal with numerous challenges. Firstly, the context is complicated by the appearance of other objects such as buildings, traffic signs, and trees. Secondly, aerial images are usually captured with different viewpoints low resolution and are easily affected by environmental factors such as weather, brightness, speed, rotation, density, or position of the object (Du et al., 2018; Chung et al., 2020; Nguyen et al., 2020). This would lead to some undesirable effects on the object detection result.

In fact, the existence of smoke, dust, and fog in the atmosphere makes the aerial images prone to blur, low contrast, and chromatic aberration, which could create many challenges in object detection, tracking, human identification, etc. especially object detection. This paper focused on detecting objects in aerial images that were captured in limited visibility conditions caused by haze. The input is hazy aerial images, and the output is the position of vehicles in that image (such as in Figure 1).

GCA-Net (Chen et al., 2019a), DCP (He et al., 2010), AOD-Net (Li et al., 2017), Dehaze-Net (Yang & Sun, 2018) are some effective dehazing methods. However, in fact, there are still several deviations in color and contrast in haze-free images when using these methods to dehaze images in real cases. After FFA-Net (Qin et al., 2020) was published, these problems were significantly solved thanks to its effectiveness and novelty. To verify the efficiency of dehazing, this paper referenced some common object detection methods, e.g., PAA (Kim & Lee, 2020), YOLOv4 (Bochkovskiy et al., 2020), RepPoint (Yang et al., 2019), SCNet (Vu et al., 2021)

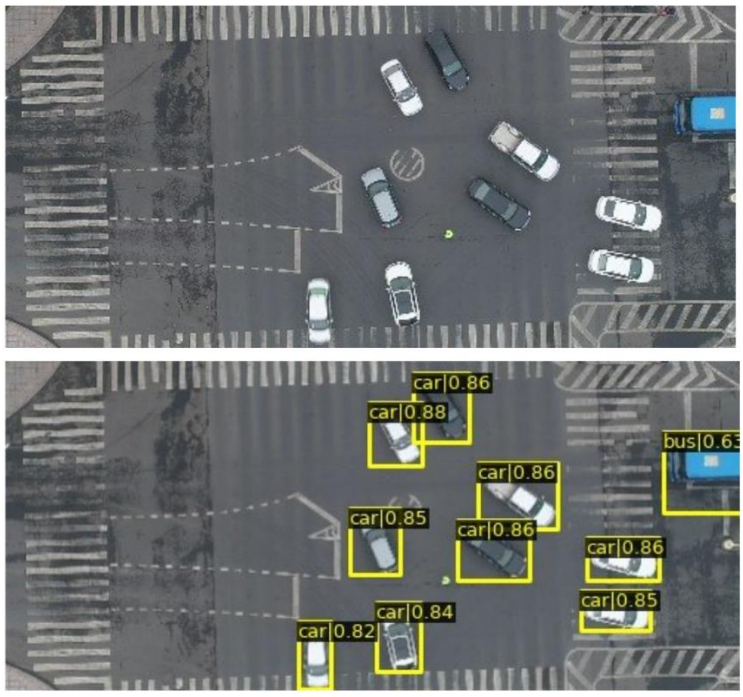

Figure 1. Object detection in hazy images
Input: Hazy aerial image - Output: Label, position of objects

2. MATERIAL AND METHODS

2.1. Single Image Dehazing

Single image dehazing acts like an image pre-processing step before performing other image processing tasks. In fact, the existence of haze, smoke, dust, etc. causes a lot of difficulties for classification, object detection, etc.

Previous studies (McCartney, 1976; Narasimhan & Nayar, 2000, 2002) have provided a simple formula for estimating the effects of haze on images as follows:

$$I(z) = J(z)(t) + A\big(1 - t(z)\big) \quad (1)$$

Where $I(z)$ is the observed hazy image, A is the global atmosphere light, and $t(z)$ is the medium transmission map, J is the haze-free image. According to formular (1), dehazing an image is just the problem of calculating the value of two variables A and $t(z)$. This means, a haze-free image can be calculated with the following formula:

$$J(z) = ((I(z) - A)/t(z))) - A$$

Dark Prior Channel (He et al., 2010) has been one of the outstanding prior-based dehazing methods. The authors of this method assumed that image patches of outdoor haze-free images usually have at least one-color channel that has a low-intensity value. However, these methods often overestimate the value of the transmission map because the priority values are easily influenced in practice. Therefore, in real cases, they often achieve bad results.

Along with the development of Deep Learning, a more effective way to deal with the effects of haze was born with the introduction of DehazeNet (Cai et al., 2016), multi-scale CNN (MSCNN) (Ren et al., 2016), the residual learning technique (He et al., 2016), the Densely connected pyramid dehazing network (Zhang & Patel, 2018), etc. Deep Learning tries to regress the transmission map directly. With a large amount of training data, these methods have achieved unexpected results.

2.2. Feature Fusion Attention Network (FFA-Net)

The authors of FFA-Net (Qin et al., 2020) introduced an effective single image dehazing. The experimental result proved that FFA-Net surpassed many previous state-of-the-art models by a very large margin with the score from 30.23dB to 36.39dB on the PSNR evaluation score at the time the authors published their works. This superior result was based on three-folds as below:

Feature Attention (FA) module combined Channel Attention and Pixel Attention mechanism. This mechanism helps FFA-Net flexibly deal with different features and haze pixels. This was because the author considers that the distribution of fog on different pixel regions was different.

Basic block structure including Local Residual Learning (LRL) and Feature Attention (FA) module made the training process more stable and increased the effectiveness of dehazing images. This was possible because LRL allowed the network to pay attention to important information and bypassed less important information such as the thin haze region by using multiple skip connections.

Attention-based different levels Feature Fusion (FAA) allowed the network to retain shallow layers' information and passed it into deep layers. In addition, it could combine all features and adaptively learned the different weights of different levels feature information.

2.3. Probabilistic Anchor Assignment with IoU Prediction for Object Detection (PAA)

PAA (Kim & Lee, 2020) was an object detection method based on RetinaNet architecture. This method proposed a new anchor box assigned during the training process and modified the loss function. Experiments showed that the proposed method has increased performance in object detection on the MS COCO test-dev dataset with various backbones.

For most CNN-based object detection methods, one of the most common ways to represent objects of various sizes and shapes was to slide anchor boxes with various scales and sizes on images. In this method, the process of assigning anchors determined which object that the anchors represent. The most common way to determine a positive sample was to use IoU between the anchor and ground truth. With each ground truth, one or more anchors would be assigned positive if the IoU score exceeds the threshold. However, this method did not really identify the actual content of the overlapping area, which could contain background, nearby objects. And then, the calculated IoU values did not reflect the similarity between the anchor and ground truth. Therefore, the authors investigated the gap between the testing and training objectives and predict the Intersection-over-Unions of detected boxes as a measure of localization quality to reduce the discrepancy. The core of PAA is the determination of positive and negative samples in favor of the model so that it can infer the separation in a probabilistically suitable way, which leads to easier training in comparison with the heuristic IoU hard assignment or non-probabilistic assignment strategies.

2.4. Proposal method

To evaluate the effectiveness when applying the dehazing method to the job of detecting objects in foggy images, two training methods were conducted as follows: Training an object detection model on hazy the original dataset and an object detection model on the dataset which was dehazed by FFA-Net shown in Figure 2.

Image Dehazing: A state-of-the-art (SOTA) method for dehazing images called FFA-Net was employed. Because of difficult conditions (not having enough time and ability to build a training dataset that had a diversity of factors such as views and scenarios like RESIDE), this research used a pre-trained model trained on the RESIDE Outdoor Training Set (OTS) dataset provided by the author. The input of FFA-Net was a hazy image with uneven frequency and the output was a haze-free image that has the same size and resolution as the input.

Object Detection: In this research, the PAA method (Probability Anchor Assignment with IoU Prediction for Object Detection) was used to perform object detection. The input of the PAA model was a haze-free image and the output was prediction annotations.

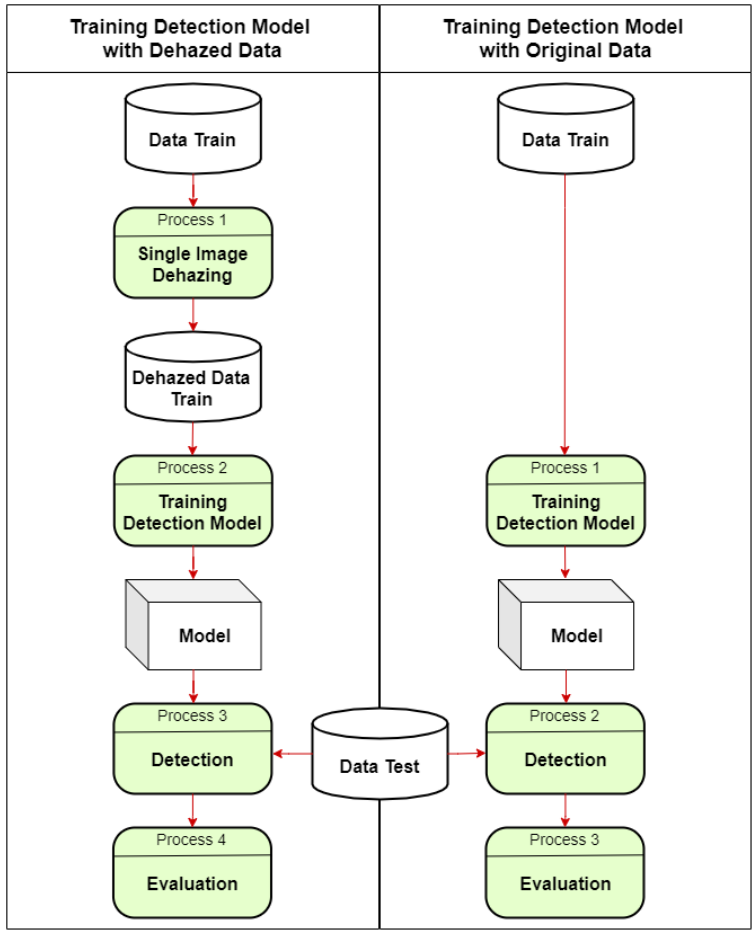

Figure 2. Experimental architecture. Hazy image after being dehazed by FFA-Net becomes the input of object detection model

2.5. Dataset

2.5.1. RESIDE Outdoor Training Set (OTS)

Published in 2017, REalistic Single Image DEhazing (RESIDE) (Li et al., 2018) has been a large-scale benchmark dataset including both synthetic and real-world hazy images. The RESIDE dataset was divided into five subsets, each serving different training or evaluation purposes.

The pre-trained FFA-Net model was trained on the RESIDE Outdoor Training Set (OTS), which had 72,135 outdoor hazy images and consisted of paired clean outdoor images and generated hazy ones. It was a subset of the RESIDE-β dataset and could be used for training. This dataset was designed to manifest a diversity of evaluation viewpoints.

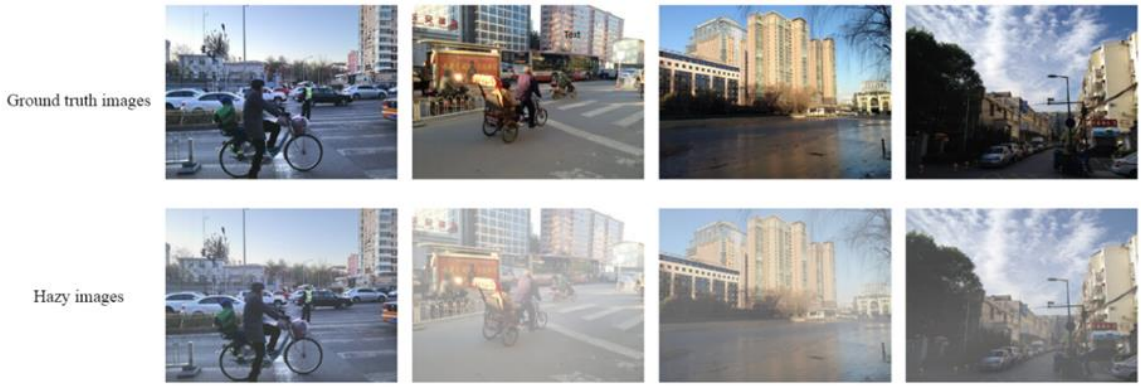

Figure 3. Ground truth and hazy images of RESIDE Outdoor Training Set (OTS)

2.5.2. UAVDT-Benchmark-M

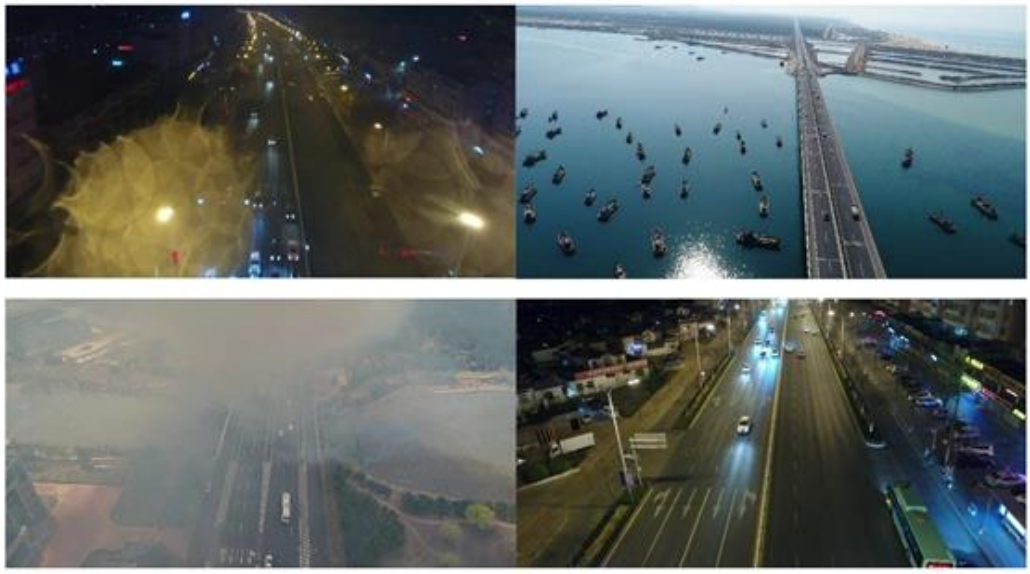

Figure 4. Some images of UATDT-Benchmark-M dataset

The aerial dataset used in this research was a subset of UAVDT-Benchmark (Du et al., 2018) called UAVDT-Benchmark-M. This dataset was specially used for Object detection and Multiple Objects Tracking. This dataset included training data: 23384 images (2998 foggy images - 12.58%) and testing data: 2181 images (2181 foggy images - 100%). The number of three classes in the training dataset and the test dataset are shown in Table 1.

Table 1. Class distribution in the object detection training dataset

	Car	Truck	Bus
Train	394,633	17,491	10,787
Test	104,705	3,703	1,080

2.6. Experiment setting

All the experimental process was executed on a GeForce RTX 2080 Ti GPU with 11018 MiB memory. In terms of the dehazing process, a pre-trained OTS model provided by the author was used. MMDetection framework V2.9.0 (Chen et al., 2019b) was employed to train the PAA detection models with the R-101-FPN backbone in 36 epochs. According to the multiple configuration training result table, which was provided on the MMDetection GitHub website, the AP score of the model trained in 36 epochs and using the R-101-FPN backbone was highest. The default configuration was used for both models trained in two different ways that were introduced in section

2: (1) training a model using the original dataset; (2) another model using the dataset which was dehazed by the FFA-Net pre-trained model.

2.7. Evaluation metric

To evaluate the effectiveness when using the dehazing method before detecting objects in aerial images, a manually selected testing dataset of UAVDT-Benchmark-M including 2191 hazy images which were contained in M0701, M1004, and M1009 was used. The trained model was evaluated on the mAP score. The AP score was calculated for 10 IoU varied in range from 50% to 95% with steps of 5%. Besides, two specified values of 50% and 75% were also reported.

3. RESULTS AND DISCUSSION

3.1. Result Analysis

In the first case, the original data was used for both train and test processes to get the default result. In the second case, the model was trained with dehazed images and then tested with original data. As shown in Table 2, a higher score was witnessed in terms of bus (from 0.7% of original to 1.4% of dehazed model). However, the figures for both car and truck were decreased from 19.6% to 18.7% and from 17.1% to 11.2% when applying the dehazing method, respectively. Consequently, there was an assumption that the dehazing process had caused some bad effects on the features of these two classes (car and truck) and the images might lose some essential information after being dehazed, which lead to the considerably lower result.

Table 2. PAA detection experimental result in mAP (%)

Training type	AP			mAP	mAP_{50}	mAP_{50}
	Car	Truck	Bus			
Original	19.6	17.1	0.7	12.5	26.7	9.7
Dehazed	18.7	11.2	1.4	10.4	21.4	9.1

On the other hand, the AP score of the truck declined by 5.9%, which was much higher than the decrease of cars with only a 0.9% drop. In addition, when the result was visualized, some truck objects existing within images were wrongly detected as car objects. Therefore, a hypothesis called the selective dehazing hypothesis was put forward. This hypothesis kept the areas inside the bounding boxes of trucks as original images, which meant that every area would be dehazed except the areas within the trucks' bounding boxes in an image and the fog was treated as part of truck objects. This would lead to some parts of a truck object (such as cabin and trunk) might be heavily covered by the fog, but original information could remain unchanged. This approach might reduce information loss and wrong detection. The experimental result of this hypothesis was shown in the following section.

3.2. Selective dehazing hypothesis

The visualization result showed that the selective dehazing hypothesis had helped the PAA detection model to detect more correctly, the mentioned wrong detection occurred with truck was reduced and the most significant example was displayed in Table 3.

Table 3. Selective dehazing hypothesis experimental result in mAP (%)

Training type	AP			mAP	mAP_{50}	mAP_{50}
	Car	Truck	Bus			
Original	19.6	17.1	0.7	12.5	26.7	9.7
Dehazed	18.7	11.2	1.4	10.4	21.4	9.1
Dehazed (truck_ex)	21.9	11.3	4.4	12.5	26.5	9.8

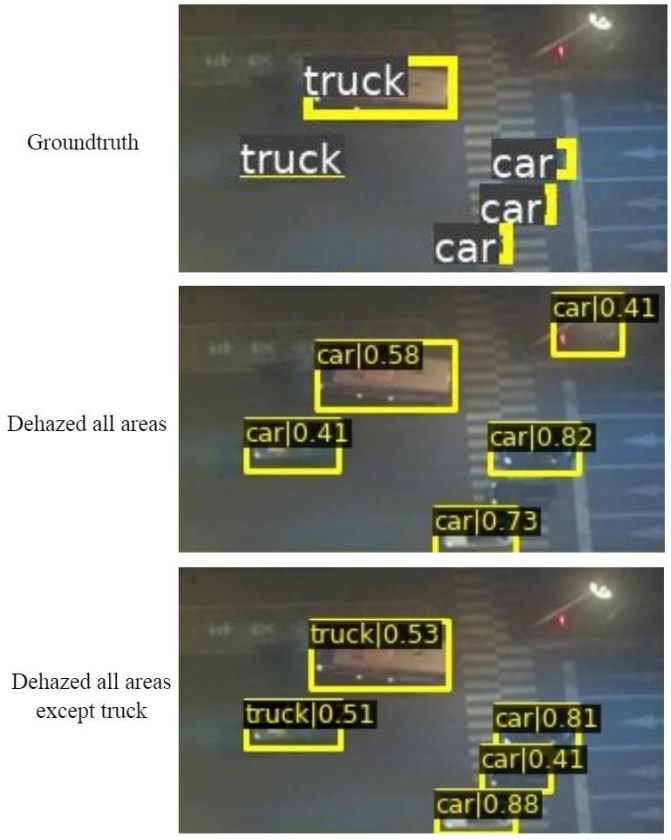

Figure 5. Visualization of the selective dehazing hypothesis experimental result

As shown in Table 3, when applying the hypothesis to our method, there were considerable increases in the score of car and bus. Meanwhile, the figure for truck just inclined 0.1%, which was not our expectation. Finally, after visualizing the result and observing the result table, two observations ware proposed:

Firstly, in terms of truck, the dehazing process made the detection result of truck in aerial images decrease significantly, even when the selective dehazing hypothesis was applied and the fog was treated as part of truck objects.

Secondly, regarding car and bus, the new hypothesis reduced the proportion of wrong detection. Therefore, the result of these classes increased significantly, especially, the object detection result of bus had increased by 3.7% compared to the original result.

4. CONCLUSIONS

This paper focused on combining FFA-Net and PAA methods for the problem of detecting vehicles from foggy aerial images. The experiment analysis and evaluation indicated the negative effect of haze on object detection in aerial images, fog reduces the detection result of *car* and *bus* by 2.3% and 3.7%, respectively. However, when the fog was treated as part of the *truck* objects, the detection result of other classes increased noticeably.

In the future, foggier aerial images will be collected to create a more suitable distribution for the dataset that can be used to train the FFA-Net model. Furthermore, more research work will be done to improve the detection result.

ACKNOWLEDGMENT

This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number DS2021-26-01. The research team would like to express our sincere thanks to Multimedia Communications Laboratory (MMLab), University of Information Technology, Vietnam National University Ho Chi Minh City, for their support in this research.

REFERENCES

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

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

Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., ... & Hua, G. (2019a). Gated context aggregation network for image dehazing and deraining. In *2019 IEEE winter conference on applications of computer vision (WACV)* (pp. 1375-1383). IEEE. https://doi.org/10.1109/WACV.2019.00151

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., ... & Lin, D. (2019b). MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*.

Chung, Q. M., Le, T. D., Dang, T. V., Vo, N. D., Nguyen, T. V., & Nguyen, K. (2020). Data augmentation analysis in vehicle detection from aerial videos. *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)* (pp. 1-3). IEEE. https://doi.org/10.1109/RIVF48685.2020.9140740

Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., ... & Tian, Q. (2018). The unmanned aerial vehicle benchmark: Object detection and tracking. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 370-386). https://doi.org/10.1007/978-3-030-01249-6_23

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

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

Kim, K., & Lee, H. S. (2020). Probabilistic anchor assignment with iou prediction for object detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16* (pp. 355-371). Springer International Publishing. https://doi.org/10.1007/978-3-030-58595-2_22

Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision* (pp. 4770-4778). https://doi.org/10.1109/ICCV.2017.511

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

McCartney, E. J. (1976). *Optics of the atmosphere: scattering by molecules and particles* (1st ed.). Wiley, New York.

Narasimhan, S. G., & Nayar, S. K. (2000, June). Chromatic framework for vision in bad weather. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662) (Vol. 1, pp. 598-605). IEEE. https://doi.org/10.1109/CVPR.2000.855874

Narasimhan, S. G., & Nayar, S. K. (2002). Vision and the atmosphere. *International Journal of Computer Vision*, *18*(3) 233-254. https://doi.org/10.1023/A:1016328200723

Nguyen, K., Huynh, N. T., Nguyen, P. C., Nguyen, K. D., Vo, N. D., & Nguyen, T. V. (2020). Detecting objects from space: An evaluation of deep-learning modern approaches. *Electronics*, *9*(4), 583. https://doi.org/10.3390/electronics9040583

Qin, X., Wang, Z., Bai, Y., Xie, X., & Jia, H. (2020, April). FFA-Net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(7), 11908-11915. https://doi.org/10.1609/aaai.v34i07.6865

Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., & Yang, M. H. (2016, October). Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision* (pp. 154-169). Springer, Cham. https://doi.org/10.1007/978-3-319-46475-6_10

Vu, T., Kang, H., & Yoo, C. D. (2021, May). SCNet: Training Inference Sample Consistency for Instance Segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(3), 2701-2709. https://www.aaai.org/AAAI21Papers/AAAI-3154.VuT.pdf

Yang, D., & Sun, J. (2018). Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 702-717). https://doi.org/10.1007/978-3-030-01234-2_43

Yang, Z., Liu, S., Hu, H., Wang, L., & Lin, S. (2019). Reppoints: Point set representation for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9657-9666). https://doi.org/10.1109/ICCV.2019.00975

Zhang, H., & Patel, V. M. (2018). Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3194-3203). https://doi.org/10.1109/CVPR.2018.00337