Project Final Report

Ruijia Chen
Work related to Dataset

Zhiqi Gao
Work related to YOLO Model

Boyuan Zou
Work related to Defogging

1. Introduction

This project targets precise object tracking/detection in foggy weather. Key challenges include reduced visibility impact tracking systems and a lack of specific research in this area. The approach involves combining fog removal algorithms and tracking methods to create an effective solution that improves tracking accuracy.

The motivation behind this project arises from the challenges that foggy weather conditions pose to object tracking systems. Systems such as autonomous vehicles, surveillance, and navigation need to operate robustly regardless of the weather conditions. While fog can impair the performance. Crashes in adverse weather conditions have varying fatality rates: clear(6.1) rain (4.3), snow (2.8), sleet (3.9), and fog (17.3) fatalities per 1,000 crashes. This underscores the critical need for enhanced fog-resistant systems [13].

While there are established methods for rapid fog removal [7] and object tracking [9], object tracking under fog remains an unexplored area. A combination of them could benefit applications such as autonomous driving and surveillance systems. Research has shown that in foggy weather, the accuracy of object detection will have a significant drop [8]. Depending on the concentration of the fog, the accuracy will drop from 91.55% to 57.75%. Which demonstrates the importance of fog removal.

In conclusion, the primary objective of this proposed project revolves around amalgamating the prevailing fog removal algorithms and object-tracking methodologies to accomplish accurate object tracking in foggy conditions.

2. Related work

2.1. Defogging

Currently, single image defogging methods include (1) filter-based, (2) color correction-based methods, and (3) learning-based methods [1]. Those three categories each have their own advantages and disadvantages. For example, early works presented a filter-based method by Kaiming He [3], which aimed at the real-time scene due to its high speed, and has some application with object segmentation [7]. The learning-based method is relatively state-of-art and uses some CNN inside to help defog the image.

2.2. Object Detection & Object Tracking

In the current landscape of object detection algorithms, prominent choices include R-CNN [2], YOLO [12], and SSD [6]. Given our objective to achieve real-time object tracking under foggy conditions, we plan to opt for the widely acclaimed and high-speed YOLO algorithm with its latest model Yolo-V8.

3. Method

3.1. Dataset selection

A dataset that satisfies this project needs sufficient annotation for multi-object tracking and haze across the dataset. Since the shortage of annotated natural foggy videos for object detection. We decided to use the UA-DETRAC[14] part of the HazeWorld [15]. HazeWorld add haze with 4 different parameters, which are 0.005, 0.01, 0.02, and 0.03, to the original UA-DETRAC dataset, which has annotations for multi-object detection and multi-object tracking.

3.2. Defogging Method

For the defogging part, we select 1 representative method from each category of defogging methods and perform them on images with 4 levels of haze factors as mentioned above. For filtered-based, we choose Dark Channel Prior from [3]. For color correction-based, we choose the method from [5]. For the learning-based method, we decided to use Enhanced Pix2pix Dehazing Network (EPDN) since it presents a better result [11].

3.3. Model Selection

There are 5 different pretrained YOLOv8 model [4] for object detection task, YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. Which have increasing accuracy, size and decreasing speed. By balancing accuracy and speed, we selected YOLOv8m for this project.

3.4. Training and Evaluating Models

Since the UA-DETRAC set only has train and test sets, the training function of the YOLO Model needs both train and validation sets. We choose to randomly select 20% images from the UA-DETRAC training set to create the validation set for training purposes and use the original test set to evaluate the result.

To compare the performance of different training sets, I created 6 different training sets. Which are

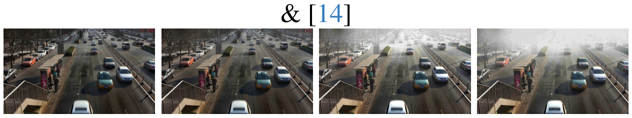
- set of original images
- set of hazy images with haze factor 0.005
- set of hazy images with haze factor 0.01
- set of hazy images with haze factor 0.02
- set of hazy images with haze factor 0.03
- a combination of 5 sets above where each category occupies 20%

for tuning YOLOv8m object-detection models [4] to see their performance and difference among original images, hazy images, and dehazed images by 3 different defogging algorithms.

4. Experiments and Results

4.1. Comparison table for defogging methods

Hazy Image with factor 0.005, 0.01, 0.02, 0.03 from [15] & [14]



De-haze by Dark Channel Prior from [3]

De-haze by Correction-based Method from [5]

De-haze by Learning-based Method EPDN from [10]

Figure 1. Comparison table for dehazing result of 3 different methods

4.2. Difference by Trained/Tuned YOLOv8m

- The original YOLO model has more than 4 classes (car, van, bus, others) as the UA-DETRAC dataset, training could let the model ignore irrelevant info.
- Training on custom dataset can improve its performance on a specific task, which is vehicle detection in this case.

Official Annotation of images from [15] & [14]

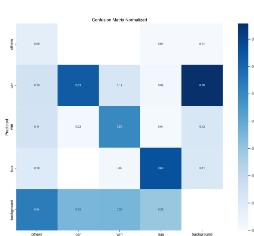
From the figure below, we can see the improvement of the Trained YOLOv8m model (right) compared to the original model (left). We can see that the trained model ignores the person and classify the van correctly.

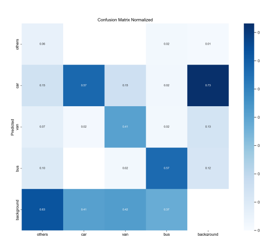
The detection result for the original image

The detection result of dehazed by Dark Channel Prior

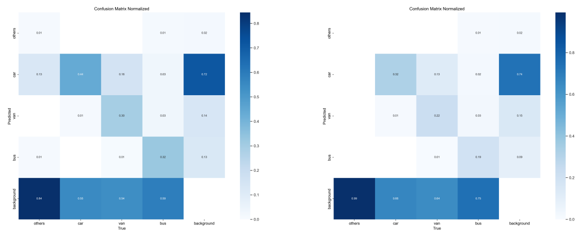
Figure 2. Comparison table for original model and trained model

Evaluation Result on Ground Truth Dataset





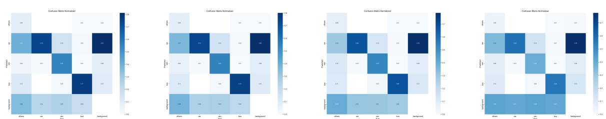
Evaluation Result on Hazy Image with factor 0.005, 0.01



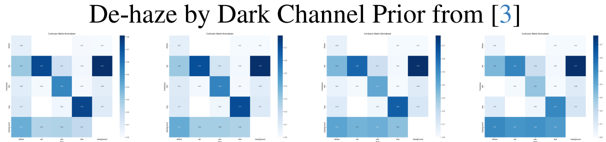
Evaluation Result on Hazy Image with factor 0.02, 0.03

4.3. Comparison for performance for YOLOv8m trained on original image

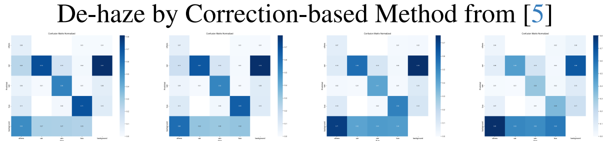
From the figure above, we can see that the detection precision drops as the hazy factor increases. For example, the precision of cars drops from 0.73 to 0.32 from the ground truth set to the hazy set with a factor of 0.03.



De-haze by Dark Channel Prior from [3]



De-haze by Correction-based Method from [5]



De-haze by Learning-based Method EPDN from [10]

Figure 3. Comparison table for Detection Precision for 3 different defogging techniques. In each row, the result comes from defogging for haze factor 0.005, 0.01, 0.02, 0.03 correspondingly

By Figure 3, we can see that the tracking accuracy of Dark Channel Prior from [3] outperforms the other 2 methods in each haze factor by a minor advantage. Hence, we will ignore the other 2 methods in comparison for the next section.

4.4. Comparison for performance for YOLOv8m trained on different dataset

In this part, we will compare the performance of different training sets by comparing several benchmarks of them on the GroundTruth test set and defogged set with different hazy factors by the Dark Channel Prior method.
Explanation for Abbreviations in tables:
GT: GroundTruth
Number X in Test Set: dehazed test set (by DCP) with original haze factor X
Training set meaning: see "section 3.4 Training and Evaluating Models" above

Table 1. mAP50 for Different Training/Testing Set

Test Set	Training Set					
	GT	0.005	0.01	0.02	0.03	Mixed
GT	0.556	0.542	0.51	0.475	0.481	0.557
0.005	0.551	0.53	0.491	0.461	0.475	0.553
0.01	0.546	0.541	0.501	0.471	0.474	0.552
0.02	0.534	0.561	0.509	0.511	0.492	0.55
0.03	0.475	0.561	0.505	0.527	0.506	0.539

Table 2. mAP50-95 for Different Training/Testing Set

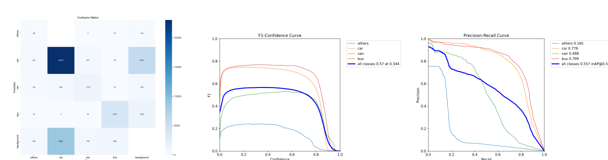
Test Set	Training Set					
	GT	0.005	0.01	0.02	0.03	Mixed
GT	0.412	0.389	0.36	0.335	0.334	0.418
0.005	0.404	0.376	0.341	0.325	0.325	0.412
0.01	0.403	0.385	0.35	0.327	0.325	0.411
0.02	0.399	0.403	0.367	0.354	0.345	0.413
0.03	0.357	0.406	0.368	0.375	0.364	0.405

By comparing table 1 and table 2, we can see that the model trained by the mixed training set has the best mAP50 and mAP50-95. While the model by Ground truth set and set with 0.005 haze factor are also competitive, but not universalizing enough.

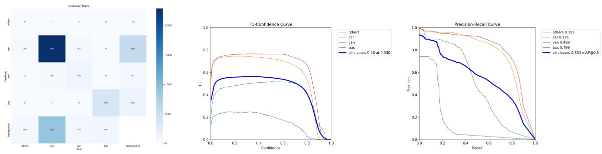
The model trained by the ground truth training set has relatively low performance on the test set with a high haze factor (0.02 & 0.03). The reason behind it might arise from the fact that the images with a high haze factor that dehazed by DCP still have a blurry part. The blurry part could hamper the detection model if it is not seen in the training set.

For the model trained by the set with a hazy factor of 0.005, we can see that it has better performance than other models trained on non-mixed hazy sets. We induce that the reason is larger hazy factor might confuse the model about the vehicle's characteristics, which is more obvious in the dehazed images.

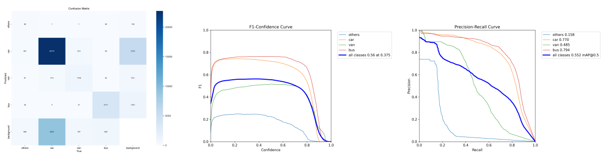
4.5. Detailed metric for the YOLOv8m model trained on mixed dataset



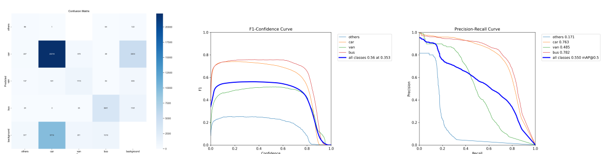
Benchmark on Groundtruth dataset



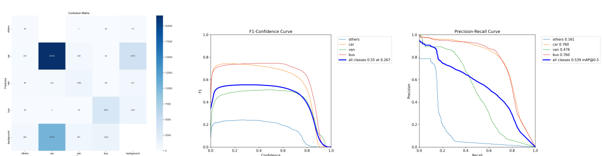
Benchmark on DCP-dehazed dataset with original haze factor 0.005



Benchmark on DCP-dehazed dataset with original haze factor 0.01



Benchmark on DCP-dehazed dataset with original haze factor 0.02



Benchmark on DCP-dehazed dataset with original haze factor 0.03

5. Conclusion

5.1. Precision for classes

It is worth noticing that the precision for cars and buses is generally higher than for the van, and ”others” are deficient under any combination of training and testing sets.

The class ”car” and class ”bus” have high precision. The reason behind the buses is especially large and special, and cars have the most number of instances in the training set.

For the class ”van”, we can see that it is likely to be mispredicted as a car by the model due to its similar exterior shape.

For the class ”others”, the general accuracy is very low in any combination of the training set and testing set. We believe that is due to a lack of specific characteristics in this class since every vehicle that does not belong to a car, van, or bus needs to be classified as ”others” in this dataset. Although we can have around 80% precision in the training set for this class, the total number of instances is only around 1000, which is not sufficient for classifying a ”hard” class.

5.2. Gap between different training set

By our result, we can find the mixed dataset a combination of 5 sets above where each category occupies 20% has the best performance in object detection. It may be because the combined dataset has a diverse distribution, so the model can learn more general features about vehicles since it sees variation under different levels of haze conditions.

5.3. From detection to tracking

The testing set for our dataset has 30 sets of images, which can be converted to 30 videos in the case of tracking.

Although the Yolov8m model [4] only has detection benchmarks such as mAP, precision, and recall, we manually check through the tracking result of our model and discover that the model can track an object as soon as it detects it in most cases. The only case it fails is when it classifies a vehicle as a van in a frame but as a car in another frame.

Additionally, the speed for detection is 0.1ms preprocess, 6.5ms inference, 0.0ms loss, and 0.6ms postprocess per image with an NVIDIA GeForce RTX 4070 Laptop GPU (8188MiB). The running speed for applying Dark Channel Prior on a single image is less than 10.0ms with AMD Ryzen 5 5600X 6-Core Processor (3.70 GHz), which makes the total processing time less than 20ms in a home-level computer. This should be a sufficient speed for real-world object tracking in foggy conditions.

References

[1] Bijaylaxmi Das, Joshua Peter Ebenezer, and Sudipta Mukhopadhyay. A comparative study of single image fog removal methods. *Vis Comput*, 38:179–195, 2022. 1

[2] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014. 1

[3] 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, 2010. 1, 2, 3

[4] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. YOLO by Ultralytics, 2023. 1, 2, 4

[5] Se Eun Kim, Tae Hee Park, and Il Kyu Eom. Fast single image dehazing using saturation based transmission map estimation. *IEEE Transactions on Image Processing*, 29:1985–1998, 2020. 1, 2, 3

[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016. 1

[7] Xinchao Liu, Liang Hong, and Yier Lin. Rapid fog-removal

strategies for traffic environments. *Sensors (Basel)*, 23(17): 7506, 2023. 1

[8] Zhaohui Liu, Yongjiang He, Chao Wang, and Runze Song. Analysis of the influence of foggy weather environment on the detection effect of machine vision obstacles. *Sensors*, 20 (2):349, 2020. 1

[9] Isaac Oluwadunsin Ogunrinde and Shonda Bernadin. Improved deepsort-based object tracking in foggy weather for avs using sematic labels and fused appearance feature network. 2023. 1

[10] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8152–8160, 2019. 2, 3

[11] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8160–8168, 2019. 1

[12] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 1

[13] Brian C. Tefft. Motor vehicle crashes, injuries, and deaths in relation to weather conditions, united states, 2010 – 2014. Report, 2016. January. 1

[14] Longyin Wen, Dawei Du, Zhaowei Cai, Zhen Lei, Ming-Ching Chang, Honggang Qi, Jongwoo Lim, Ming-Hsuan Yang, and Siwei Lyu. UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking. *Computer Vision and Image Understanding*, 2020. 1, 2

[15] Jiaqi Xu, Xiaowei Hu, Lei Zhu, Qi Dou, Jifeng Dai, Yu Qiao, and Pheng-Ann Heng. Video dehazing via a multi-range temporal alignment network with physical prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1, 2