A Study on Fog and Haze Weather Object Detection Method Based on DeblurGANv2 and YOLOv4

Xiaocong Qiu[1],Jacob Taylor[2]

Misericordia University[1],Misericordia University[2]

Xiaocongxiaocong@gmail.com[1],Jacobtaylor897@gmail.com[2]

Abstract:

To address the issue of low detection accuracy in object detection under foggy and hazy conditions, a novel defogging object detection method that integrates DeblurGANv2 and YOLOv4 is proposed. This approach incorporates an image enhancement algorithm from the DeblurGANv2 generative adversarial network into the preprocessing module of YOLOv4. This integration aims to preprocess foggy images, preserving high-quality texture and color information. Additionally, the method replaces the CSPDarkNet53 backbone network in YOLOv4 with the lightweight ShuffleNet V2 neural network to enhance model detection speed. Experimental results demonstrate that this proposed method effectively reduces issues of significant color discrepancy and fog residue, achieving a mean average precision (mAP) of 86.56% on the test dataset. These results indicate the method's efficacy in real-world defogging object detection scenarios.

Keywords:

Generative Adversarial Nets; DeblurGANv2; Coordinate Attention; ShuffleNet V2;Defogging Target Detection.

1. Introduction

Traditional target detection is to quickly and accurately identify the category information of the target in complex scenes and accurately locate the position of the target in the image. Target detection is the most basic and common problem in the field of computer vision. With the rapid development of artificial intelligence and deep learning, traditional target detection has also made a series of progress. Common target detection can be divided into single-stage target detection and two-stage target detection. Common single-stage target detection include SSD [1], Yolo (you only look once) series [2-5] and retina net [6]. Common two-stage target detection include FPN [7] and r-cnn [8].

In recent years, environmental pollution has led to frequent smog weather in many cities. Smog weather will not only cause harm to human body, but also cause serious inconvenience to people's daily travel. At the same time, it will also affect target detection and reduce the accuracy of target detection. In recent years, with the rapid development of artificial intelligence and machine learning, traditional target detection has made significant progress in detection accuracy and speed, However, due to the influence of image quality, the feature extraction in the process of image feature extraction by neural network is inaccurate, which seriously affects the accuracy of target detection.

At present, defogging target detection can be divided into two categories. One is the uncorrelated model based on defogging and detection, that is, defogging operation first and then target detection. For example, Guo Chunle [9] and others proposed a defogging network por net combined with fast r-cnn [10] for uncorrelated defogging target detection. The detection accuracy is improved by 2%. The other is to jointly optimize the defogging algorithmand target detection algorithm, and carry out defogging and detection at the same time. For example, Li [11] and others proposed an end-to-end defogging target detection based on aodnet and fast r-cnn in 2017, which improves the accuracy of target detection in foggy days, but the defogging effect is not ideal, and artifacts will appear in defogging pictures. Since then, the teamhas also conducted a series of studies [12-14] and proposed the fog data set rest [12]

In order to improve the accuracy of fog detection based on gaurv2 and gaurv4, this paper puts forward a method of fog detection based on gaurv4 in order to improve the fog detection accuracy of the image, and get a better fog detection result based on gaurv4, Firstly, the lightweight neural network shufflenet V2 [15] is used to replace the cspparknet53 network used for backbone feature extraction in ylov4 to improve the speed of model mark detection.

2. Improved YOLOv4

2.1. Deblurganv2 Defogging Pretreatment

In order to remove the fog function, this paper uses deblurganv2 network to remove the fog in the preprocessing module of yolov4 network. The specific process is to process the foggy picture through deblurganv2, retain the color information and texture information in the original picture to the greatest extent, and generate a defogging image close to the real scene. When the network receives the image input with the size of 416 * 416, first use deblurganv2 network for defogging preprocessing, After the defogging process is completed, the image size is reset to 416 * 416 size through the deconvolution network layer, and then the features are extracted by the improved yolov4 feature extraction network.

The image composition model proposed by Professor gaurv2 at the Catholic University of Ukraine in 2019 is shown in gaurv2 Different from deblurganv1 [16], deblurganv2 performs feature fusion based on FPN structure FPN can not only fuse multiple different scale information, but also achieve a good balance between speed and accuracy. FPN structure includes bottom-up and top-down paths. The bottom-up path is a convolution network for feature extraction. The main body of deblurganv2 includes two discriminators with different sizes and a generator for image restoration. After considering the difference between global size and local size, the discriminator combines fullgan discriminator and patchgan discriminator to deal with complex fuzzy problems. In the generator, the feature pyramid (FPN) framework is used to extract different feature mapping layers, in which the bottom-up process is convolution neural network to extract features, In the process of feature extraction, the image is down sampled. After the down sampling, the sampling results are horizontally linked with the top-down up sampling to obtain more feature information. Finally, the output feature maps of five different scales are obtained. Then, the five characteristic graphs are sampled to a quarter of the input size and connected to a tensor. This tensor contains different levels of semantic information. Finally, the upper sampling layer and convolution layer are added to deblurganv2 network to restore clear images and remove artifacts.

When deblurganv2 uses different backbone networks for training, different image processing effects can be obtained. In order to obtain more efficient fuzzy image processing methods, this paper selects inception as the backbone network of deblurganv2.

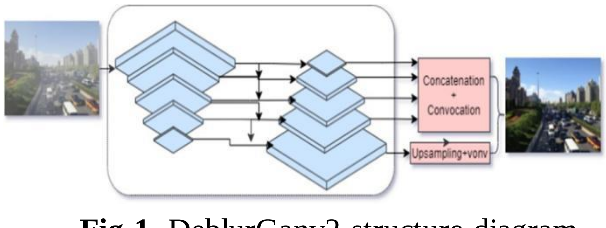


Fig 1. DeblurGanv2 structure diagram

2.2. Lightweight Shufflenet V2 Replaces Cspparknet53

In order to effectively reduce the model parameters in the algorithm, reduce the model size and improve the target detection speed of the algorithm, this paper uses shufflenet V2 as the backbone feature extraction network of yolov4.

Ma et al. Proposed four efficient and lightweight network design criteria in shufflenet V2, followed these four principles to redesign shufflenet V1 [17], and obtained shufflenet v2. The network structure of shufflenet V2 is shown in Fig 2. The basic components of shufflenet V2 network model can be roughly divided into two types. The two basic unit modules of shufflenet V2 are shown in Fig 3 below.

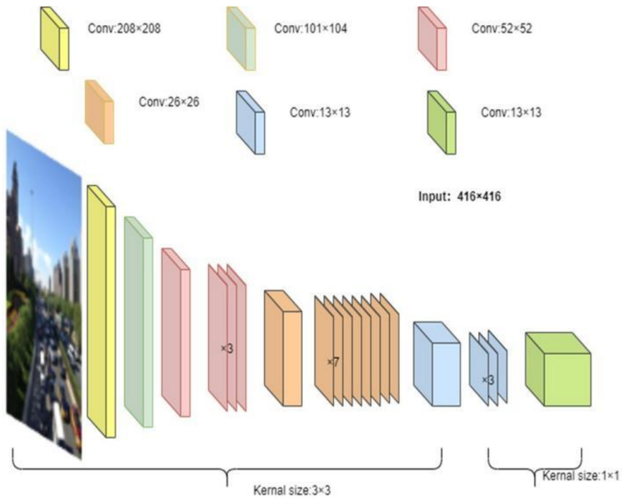
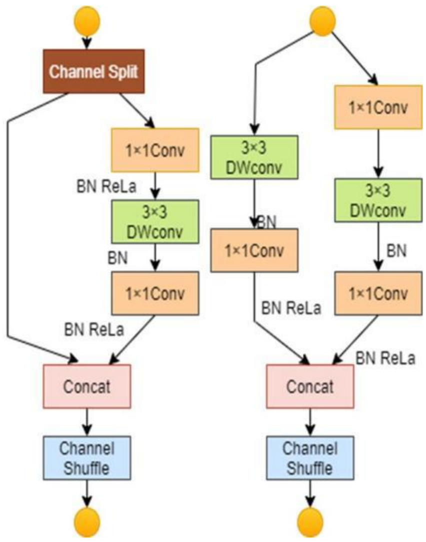


Fig 2. ShuffleNet V2 structure diagram



Unit 1 Unit 2

Fig 3. ShuffleNet V2 two modules

In shufflenet V2 unit 1, the input characteristic map is first divided into two branches with the number of channels accounting for 1 / 2. The left branch is constant. The right branch will use the same number of input and output channels and need to go through three convolutions with a step length of 1. Among them, 1 * 1 convolution represents ordinary convolution and 3 * 3 convolution represents deep convolution (dwconv). When the three convolutions are completed, the two branches will perform concat operation, the number of channels will be added, and the features will be fused at the same time. Finally, the channel shuffle operation will be carried out to fully integrate different channels. Unlike unit 1, as shown in Fig 3 (b), in shufflenet V2 unit 2, the channels will not be divided, and the feature map will be directly input to the two branches. Both branches use 3 * 3 depth convolution with step size of 2, Reduce the dimension of the length (H) and width (W) of the feature map, to reduce the amount of network calculation. Then, the concat operation is carried out after the output of the two branches, and the sum of the number of channels is twice the original input, which increases the width of the network, increases the number of channels without significantly increasing flops, and makes the network more capable of extracting features. Finally, the channel shuffling is also carried out to realize the information exchange between different channels.

3. Result Analysis

The training data set used in this paper is OTs in rest. OTs in rest contains 8971 clear pictures, and each clear picture generates ten foggy data sets. In this paper, we select one clear picture including five kinds of objects: people, cars, trucks, bicycles, and traffic lights and one corresponding fuzzy picture.

In order to further prove the effectiveness and progressiveness of the algorithm proposed in this paper, this paper simply combines fast RCNN and yolov4 with dark channel prior, Retinex and deblurganv2 algorithms respectively, the six fog detection models of yolov3 dark channel, yolov3 Retinex, yolov3 deblurganv2, yolov4 dark channel, yolov4 Retinex and yolov4 deblurganv2 are tested separately under the same data set as the model proposed in this paper. The map values of various algorithms are shown in Table 1:

Table 1. MAP value of different models

Algorithm name	mAP%	AP%					FPS
		Traffic light	bus	person	car	motorbike	
YOLOv3- Dark channel	76.54	67.21	81.56	82.98	82.50	68.45	35
YOLOv3-Retinex	79.41	69.44	84.79	86.36	86.74	69.72	33
YOLOv3-DeblurGanv2	81.69	71.23	85.27	86.54	84.78	81.63	34
YOLOv4- Dark channel	84.45	73.12	90.36	91.71	90.7	76.36	41
YOLOv4-Retinex	85.13	74.25	91.95	90.14	91.79	77.52	39
Algorithm in this paper	**86.56**	**75.26**	**90.17**	**92.08**	**91.88**	**83.41**	**46**

It can be seen from table 4 that the map of the defogging target detection model proposed in this paper is as high as 86.56%, which is 10.02%, 7.15%, 4.87%, 2.11% and 1.43% higher than that of the five fog detection models of yorov3 dark channel, yorov3 Retinex, yorov3 deblurganv2, yorov4 dark channel and yorov4 Retinex. The detection speed is also as high as 46 frames / s, which is higher than that of other models. In addition, the detection speed of this paper is also better than all other models. Based on the above experimental results, it can be concluded that the yolov4 network model based on image restoration preprocessing adopted in this paper can improve the performance index of the original model to a certain extent, and has advantages over other target detection algorithm models.

4. Summary

In this paper, a defogging target detection method based on deblurganv2 and yolov4 is proposed. In the preprocessing stage of yolov4, the image enhancement algorithm generating countermeasure network (deblurganv2) is used to defog the image and retain the ultra-high- quality texture information in the image. At the same time, the network structure of yolov4 is redesigned. Firstly, the lightweight neural network shufflenet V2 is used as the backbone feature extraction network of yolov4. The experimental results show that the algorithm proposed in this paper has good fog detection effect.

At present, the algorithm proposed in this paper still needs to be improved, such as false detection and missed detection in the situation of serious reflection and haze, which is the placeto be improved in the next step. In the future work, this research will continue to improve the network structure, enhance the generalization ability of the algorithm, improve the robustness of defogging target detection, and apply the algorithm to practical engineering.

References

[1] Liu W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//European Conference on Computer Vision, Springer, 2016: 21-37.

[2] REDMON J, FARHADI A. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:779-788.

[3] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//CVPR 2017:Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition.Piscataway: IEEE, 2017: 7263 -7271.

[4] Redmon J, Farhadi A. Yolov3: An incremental improvement[C]//IEEE conference on Computer Vision and Pattern Recognition, 2018, arXiv: 1804.0276.

[5] Bochkovskiy A, Wang C-Y, Liao H-Y M. Yolov4: Optimal speed and accuracy of object detection [J/OL]. arXiv:2004.10934[2020-4-23].

[6] Lin T Y, Goyal P, Girshick R, et al. Focal loss for denseobject detection[C]/ /2017 IEEE International Conferenceon Computer Vision(ICCV) ,2017: 2999 - 3007.

[7] Lin T Y, Dollár P, Girshick R,et al. Feature pyramid net-works for object detection[C]/ /2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR) ,2017: 936- 944.

[8] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014:580-587.

[9] LI C, GUO C, GUO J, et al. PDR-Net: Perception -inspired single image dehazing network with refinement [J]. IEEE Transactions on Multimedia(S1520-9210), 2019, 22(3): 704-716.

[10] Ren S, He K , Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence(S0162-8828), 2017, 39(6):1137-1149.

[11] Li B Y, Peng X L, Wang Z Y, et al. AOD-Net: all-in-one de-hazing network[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Com-puter Society Press, 2017: 4780-4788.

[12] Li B Y, Ren W Q, Fu D P, et al. Benchmarking single image dehazing and beyond[J]. IEEE Transactions on Image Processing(S1057-7149), 2019, 28(1): 492-505.

[13] Liu Y, Zhao G , Gong B , et al. Improved Techniques for Learning to Dehaze and Beyond: A Collective Study [J]. arXiv:1807.00202 2018.

[14] Yang W H, Yuan Y, Ren W Q, et al. Advancing image under-standing in poor visibility environments:a collective bench-mark study[J]. IEEE Transactions on Image Processing(S1057-7149), 2020, 29: 5737-5752.

[15] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design [C]//Pro- ceedings of the European conference on computer vision (ECCV). 2018: 116-131.

[16] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, et al. Deblurgan: Blind motion deblurring using conditional adversarial networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8183-8192.

[17] Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6848-6856.