ISBN: 978-981-94-1156
2024 the 14th International Workshop on Computer Science and Engineering (WCSE 2024)
doi: 10.18178/wcse.2024.06.009

An Improved End-to-end Method for Improving Smoked Image Recognition in Outdoor Rescue Scene

Lei Zhang[1], Zhen Ruan[1], Zhen Li[1], Xi Liu[1], Hongliang Chen[1] and Ziyou Wang[1]
Shanghai Fire Research Institute of MEM,
Shanghai, China

Abstract. This paper discusses a method for image clarity during rescue vehicle operations. This method first deals with the uncertainty caused by image jitter, and then uses defogging methods to improve the image, bringing more convenience and safety to unmanned rescue processes. The dehazing process is processed using an improved AOD Net method. Firstly, replace ReLU with LeakyReLU to minimize peripheral interference, and preprocess the collected images for brightness before sending them to the detector. Secondly, using bilateral mesh filtering to make the dark scene clearer after defogging. Through experiments, it can be found that the method can improve atomized images and maintain the brightness of the images, which can help to enhance the visual experience.

Keywords: DOA-Net, Smoke detect, Dehaze Image, image enhancement, color recovery.

1. Introduction

At the rescue site, carbon oxides formed by the combustion of combustibles form mist in the air [1]. If the on-site photos can be effectively clarified, it can play a very important role in rescue work [2]. The traditional multi-scale Retinex algorithm (MSR) based on image enhancement can play a certain role [3]. In 2021, ABDUL et al. assumed low contrast of smoke components and designed a wavelet variational method to solve the problem of smoke removal [4].The AODNet (All in One Dehazing Network) algorithm [5] is an end-to-end dehazing model.In 2022, Pan et al. [6] proposed DeSmoke which is based on the unpaired image-to-image cycle-consistent generative adversarial network for removing smoke from real robotic laparoscopic hysterectomy video. The LAP algorithm [7] trains unpaired images using GAN. The outdoor rescue site is located in an environment where the atmospheric light intensity is difficult to estimate, which can cause contrast distortion, color distortion, and halo effects. There are also other methods that use Decom RNet to decompose images and then use Enhance RNet to enhance them Strong lighting and the introduction of attention modules have improved the accuracy of learning[8].In terms of scene recognition, the YOLO method was used in the recognition process for scene recognition[9]. There are also studies on the use of deep neural networks for human and human behavior judgment in public applications[10,11].However, the above algorithms Used in relatively open visual scenes cannot simultaneously take into account the low brightness and low contrast of outdoor rescue images which would lead to a decrease in recognition rate.

This paper proposes an improved method for image dehazing using AOD Net, taking into account the visual effect after dehazing. By combining high frequency and contrast, a method for improving recognition for rescue images is proposed.

The rest of the paper is structured as follow. Firstly, we introduce the proposed architecture and method in Section II. In Section III, we analyze the parameters.Experiments and results are presented in Section IV and conclusions are drawn in Section V.

2. Architecture and Algorithm

[1] Lei Zhang. Tel.: +8613621966275; fax: +86-021-54961900.
E-mail address: zhanglei2014@shfri.cn.

This paper uses a camera in motion of rescue-vehicle to capture the corresponding image sequence through an analyzer, and then uses a model for dehazing processing to finally classify and recognize the images in the scene. The specific architecture is shown in the following figure.

During the dehazing process, AOD improvement method is adopted, in which LeakyReLU is used to replace ReLU during the training phase. Retix[10] method is used to pre-preprocess the image scene in an end-to-end direction, and then placed in the model for dehazing processing.

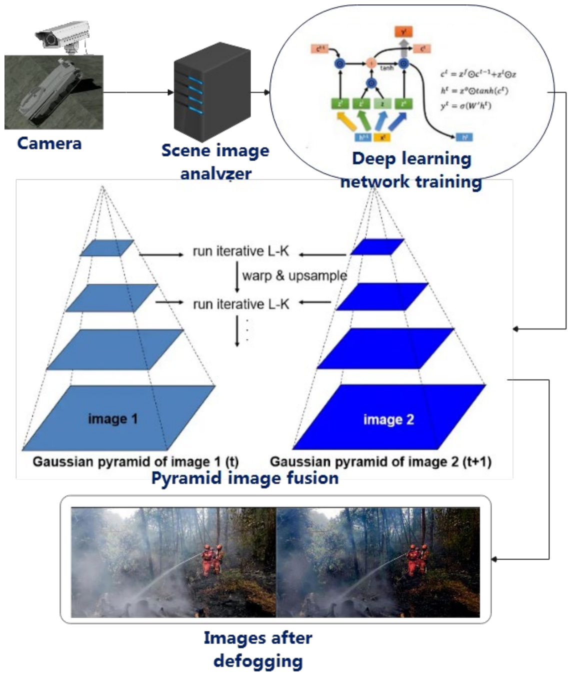

Fig.1: System architecture

In Figure 1, the overall structure is captured by the camera equipment on the vehicle platform. The rescue vehicle has outdoor and motion jitter characteristics. The front end uses scene image comparison to capture stable images, and then uses neural network methods for image denoising. Through different image fusion, the final image is generated by dehazing.

3. Improvement Strategy

The AODNet (All in One Dehazing Network) calculation method is an end-to-end dehazing model that does not require air mist evaluation, but directly generates clear images through lightweight CNN. The framework model is shown in the following figure 2.

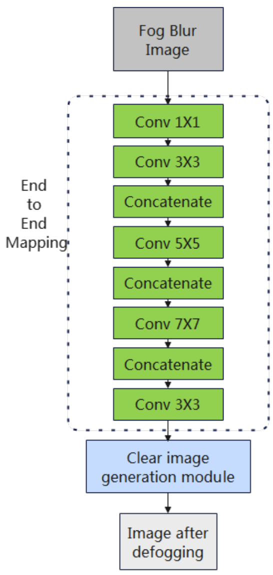

Fig. 2: Framework model

In this paper, the network used 7 convolutional layers with the aim of forming multi-scale features by fusing filters of different sizes, and the loss function used a simple mean square error loss. The calculation is mainly divided into two parts. One is the K value estimation module, which uses the input image I (x) to calculate the unknown variable K (x), calculate the concentration and depth of fog. Another is the clear image generation module, which inputs the calculated K (x) as an adaptive variable into the network to obtain a clear image J (x).

Combining with AOD-NET network, considering the problem of image and atomized surface being classified as Class 2, this paper proposes a network architecture that replaces the input LeakyReLU with ReLU. In addition to Conv in the feature layer, a maximum output structure is added to avoid oversaturation of the output when atomization is too severe. Output image can be limitted output by formula $\mathrm{F}^{\mathrm{i}}(x) = \max_{j \in (1,K)}(N^{i,j})$, this symbol i outputs the maximum output of k-neurons N at close range, indirectly using high-frequency information to improve the visual effect of the image.Then, improvements were made to YOLOV3 from the perspective of scene recognition.

To verify the recognition of clear images, first of all, target features are extracted by darknet-53 feature extraction network using a large number of 5×5, 3×3 convolution kernels, residual modules and fast links. Considering the small size of outdoor animals, the feature map tensor of different scales was obtained through 24 and 16 times down-sampling respectively. Finally the characteristics of different scale figure tensor addition, feature fusion to get 3 dimension feature maps, including the different scales of the grid number and the characteristics of each figure predicted tensor, namely 3xD (4+1+a), including four representative test box four parameters x, y, w, h, namely the center coordinates and testing frame's width and length, 1 represents the targeted evaluation, A represents the number of categories, and 3 represents the three sets of such information that each grid needs to predict. In terms of identification method, considering the size and special scene of the identified animal, and based on the principle of homology and similarity, the suspected target area is segmented to analyze 1-N pyramids, which is completed through the following process.

Pyramid image recognition processing method:

1.Collection of suspected areas <= depth frame

2.For the suspected area，and the discriminant points S_i

3. A new image size $\in$ [1/2*[Width], 1/2*[Height]]

Collection of suspected areas <=depth frame

End For

4.MaxC $_i$= {continuous suspected region set discrimination + discriminant points S={i|max（S_i）}}

At the network level, Leaky ReLu function is used as the convolution layer activation function. The output of neurons is located as the following activation function.

$$f(x)_{out} = \begin{cases} x, x > 0 \\ a * x, x <= 0 \end{cases}, \text{ here } \alpha = 0.13$$

4. Results and Analysis

4.1. Dataset

Considering the particularity of the scene, this paper adopts images with fog taken from outdoor rescue scenes. The training set uses 2000 images, and an additional 500 images are used for testing the sample set.

4.2. Experiment

In this experiment, smoke images of outdoor atomization were tested and compared using PSNR and SSIM methods[12]. To compare the comparability of processed images with human visual perception, this paper defines a block based approximate brightness comparison (BMSE, Block Mean-Square Error), and the formula for separating brightness comparison is as follows.

$$\text{BMSE} = \frac{1}{\text{H} * \text{W}} \sum_{x=0}^{(H-1)} \sum_{y=0}^{(W-1)} [\mathop{Min}_{j\in[0,B]} (I^{j}(x,y) - J^{j}(x,y))]$$

Where I represent the original image, and J represents the deburred image. The testing environment: GPU TYPE GTX4090, 10-core 20-thread E5-2690V2 96GB memory with 1 camera channel.

Table 1: Average PSNR and SSIM results

Type	AOD-Net	This Method
PSNR	19.76	20.13
SSIM	0.85	0.87
BMSE	0.72	0.83

From Table 1, it can be seen that the improved method has limited objective improvement, but can achieve better visual quality of images. The high recognizability of images can be more suitable for outdoor rescue scenes.

A)Original B) AOD-Net C) This Method

Fig.3: Comparison results of different methods

From Table I, it can be seen that the image quality of the scene has improved by about 10%, with a significant improvement for dark scenes. As can be seen from Image A, the improvement is not significant for images already in the scene. We have basically achieved the expected goal of solving the problem of image darkness caused by defogging.From Fig.3., it can be seen that classification recognition of the red marked areas in the scene can better avoid target loss with improved scene recognition.

5. Conclusions

This paper combines the issue of image clarity in firefighting and rescue to study the removal of outdoor smoke. A method is proposed to improve the clarity of dark images while removing smoke. This method has a significant improvement in subjective vision, especially in the contrast and brightness of images, and its effectiveness has been proven through experiments. Further research on image intelligence will focus on firefighting and rescue issues in the follow-up work.

6. Acknowledgment

This paper is supported by the Project supported by the National Key Research and Development Program of China(Grant No.2021YFC3090400) and Technology Plan project of the National Fire and Rescue Administration(Grant No.2021XFZD07).

7. References

[1] Lin, G., Zhang, Y., Xu, G. et al. Smoke Detection on Video Sequences Using 3D Convolutional Neural Networks. Fire Technol 55, 1827–1847 (2019).

[2] Talaat, F.M., ZainEldin, H. An improved fire detection approach based on YOLO-v8 for smart cities. Neural Comput & Applic 35, 20939–20954 (2023).

[3] Sanjay Agrawal, Rutuparna Panda, P.K. Mishro, Ajith Abraham,A novel joint histogram equalization based image contrast enhancement,Journal of King Saud University - Computer and Information Sciences,Volume 34, Issue 4,2022,Pages 1172-1182

[4] Abduru Suli Osman; Eleminu Abulijan. (2021). Tone-invariant color image enhancement algorithm based on wavelet transform and histogram equalization. Journal of Optoelectronics Laser, 40(1), 14–18.

[5] Ping-Juei Liu, "Enhance Dehazed Images Rapidly Without Losing Restoration Accuracy", IEEE Access, vol.12, pp.40795-40808, 2024.

[6] Pan Y, Bano S, Vasconcelos F, Park H, Jeong TT, Stoyanov D. DeSmoke-LAP: improved unpaired image-to-image translation for desmoking in laparoscopic surgery. Int J Comput Assist Radiol Surg. 2022 May;17(5):885-893.

[7] Q. Wu, M. Qin, J. Song and L. Liu, "An Improved Method of Low Light Image Enhancement Based on Retinex," 2021 6th International Conference on Image, Vision and Computing (ICIVC), Qingdao, China, 2021, pp. 233-241.

[8] H. Qiang, Y. Zhong, Y. Zhu, X. Zhong, Q. Xiao and S. Dian, "Underwater Image Enhancement Based on Multichannel Adaptive Compensation," in IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-10, 2024.

[9] Y. Ma, J. Yang, Z. Li and Z. Ma, "YOLO-Cigarette: An effective YOLO Network for outdoor smoking Real-time Object Detection," 2021 Ninth International Conference on Advanced Cloud and Big Data (CBD), Xi'an, China, 2022, pp. 121-126.

[10] Florian Spiess, Lucas Reinhart, Norbert Strobel, Dennis Kaiser, Samuel Kounev, and Tobias Kaupp, "People Detection with Depth Silhouettes and Convolutional Neural Networks on a Mobile Robot," Journal of Image and Graphics, Vol. 9, No. 4, pp. 135-139, December 2021.

[11] Edisanter Lo, "Target Detection Algorithms in Hyperspectral Imaging Based on Discriminant Analysis," Journal of Image and Graphics, Vol. 7, No. 4, pp. 140-144, December 2019.

[12] Agrawal S, Panda R, Mishro PK, Abraham A (2022) A novel joint histogram equalization based image contrast enhancement. J King Saud University-Comput Inform Sci 34(4):1172–1182.