Unsupervised Segmentation of Haze Regions as Hard Attention for Haze Classification

Jingyu Li[1], Haokai Ma[1], Xiangxian Li[1], Zhuang Qi[1], Xiangxu Meng[1], and Lei Meng[2,1] *

[1] Shandong University, Jinan, Shandong, China
[2] Shandong Research Institute of Industrial Technol, Jinan, China
{jingyu_lee,mahaokai,xiangxian_lee,z_qi}@mail.sdu.edu.cn
{mxx,lmeng}@sdu.edu.cn

Abstract. Haze classification plays a crucial role in air quality and visibility assessment. In contrast to traditional image classification, haze classification requires the classifier to capture the characteristics of different levels of haze. However, existing methods primarily focus on feature extraction while neglecting the interference of background information. To address this issue, this paper proposes a hard attention infused network (HAINet) for haze classification, consisting of an unsupervised segmentation module (USM) and a hybrid information fusion module (HIF). The USM is used to extract haze area information in an unsupervised manner, generating various forms of haze images. The HIA selects different various forms of haze images, as a hard attention mechanism, to reduce the impact of background and improve classification performance. We conduct experiments on two datasets, Hazel-level and Haze-Wild, in terms of performance comparison, ablation study, and case studies. The results show that our method effectively reduces the impact of background noise in haze images and consistently improves the classification performance.

Keywords: Haze classification · Hard attention · Unsupervised segmentation · Image classification.

1 Introduction

In recent years, deep learning has witnessed remarkable advancements across various fields, including classification [6, 16, 17, 20, 29, 38, 40], recommendation [22–25, 30], image generation [14, 15, 35, 39] and federal learning [21, 32]. Haze classification [7, 37, 43] has gained widespread employment in the field of air quality and visibility assessment [13, 28], especially autonomous driving [8, 9]. Unlike conventional image classification tasks, haze classification focuses on determining the level of haze in an image rather than identifying objects within the image. Previous research has primarily concentrated on feature extraction for haze classification [2, 19, 33, 42]. However, even the state-of-the-art deep learning

* Corresponding author

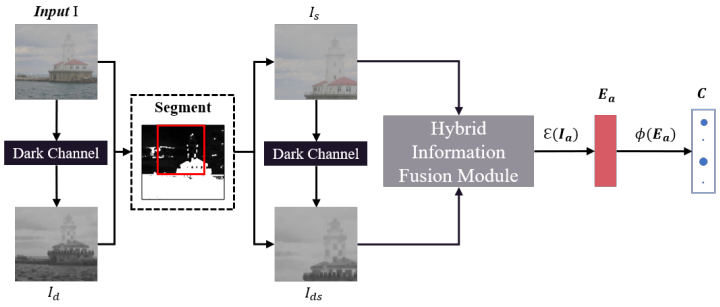

Fig. 1. Illustration to HAINet for haze classification. HAINet generates multiple haze images including the original image $\mathbf{I}$, the dark channel of original image $\mathbf{I}_d$, the segmented image $\mathbf{I}_s$ and the dark channel of segmented image $\mathbf{I}_{ds}$, which contain multiple haze information from the input haze image via the dark channel and segmentation module. The final feature $\mathbf{E}_a$ is obtained by a fusion of the images from the selected images for haze classification $\mathbf{I}_o \oplus \mathbf{I}_s = \mathbf{I}_a \mapsto \mathbf{E}_a \mapsto C$.

methods [19, 41, 44] face difficulties in accurately classifying images with heavy haze, which can be attributed to the presence of complex backgrounds. Therefore, there is a need for a method can mitigate the influence of background on haze classification results.

The haze classification methods currently available can be broadly categorized into three groups: pixel value statistics methods, engineering methods and haze feature learning approaches. The pixel value statistics methods obtain haze levels by analyzing statistics of prior knowledge haze images, such as color [3, 10, 27, 31] and dark channel prior [8, 18], which are limited by their reliance on prior knowledge. Engineering methods methods [36, 43, 45] extract visible features like edges and colors from input images and use classifiers like SVM to classify haze levels, but their effectiveness is heavily dependent on the feature selection [2, 33, 42]. Haze feature learning approaches use deep learning methods [46, 47] to classify haze images, leveraging techniques such as network ensembles [33,41], multi-branch training [26,42], and pre-training [2,7]. However, these approaches have lower performance when classifying heavy haze images, as the features of the background can interfere with the classification results.

We present a novel approach named HAINet to address the aforementioned challenges in haze classification, which comprises two main modules: the unsupervised segmentation module (USM) and the hybrid information fusion module (HIF). The USM module uses the dark channel prior and an unsupervised contrastive method to identify and segment the haze regions from the background scenes. It has two sub-modules: Dark Channel and Segment. The HIF module takes the original and segmented images from the pre-processed pool as inputs to the classifier. By concatenating both images and extracting an enhanced feature, HIF module creates a hard attention mechanism that combines information from different forms of haze images to improve the performance of the classifier. The overall framework of the proposed approach is illustrated in Figure 1. The USM module pre-processes the input image and generates a pool of pre-processed im-

ages, including the original image, the dark channel of the original image, the segmented image and the dark channel of the segmented image. The HIF module fuses these images as input to the classifier and generates an enhanced image. Overall, HAINet addresses the limitations of existing haze classification methods by segmenting the haze regions and incorporating hard attention into models.

Experiments were conducted on the Hazel-level [18] and the Haze-Wild datasets, where the Hazel-level dataset contains 3024 images in 9 classes, and the Haze-Wild dataset contains 100,000 images in 10 classes. The performance of the proposed method is evaluated with performance comparison, ablation studies, and case studies. The results show that our method achieves consistent performance gains as compared with backbones.

To summarize, this paper includes two main contributions:

- A novel haze classification approach HAINet, is proposed to address the issue of background noise in haze classification tasks by focusing on the haze region in the image, which is achieved through the extraction of various forms of haze images, and effectively exclude the negative effects of background noise on the classification process.
- We explore the role of different haze image forms in the classification task and show that extracting haze region information can shift the model's attention from objects to haze. Experiments demonstrate that HAINet can integrate haze information of pre-processed images and continuously improve the classification results.

2 Related Works

2.1 Using Pixel Value Statistics as Haze features

The pixel value statistics method involves counting the input haze image information, computing the haze value of the input image based on the statistical results, and comparing the haze value and the threshold value to determine the haze level [1]. The information comprises the lowest/highest pixel value of the original RGB image [10, 27], dark channel priors [8], depth map [18], transmitted image [3], etc. These parameters undergo operations such as logarithmic, division, and pooling calculations to achieve transformation [18]. However, these methods are constrained by the construction of specific functions, which results in poor scalability.

2.2 Engineering method

Engineering methods involve extracting features from images through manual feature engineering and training the model using machine learning methods. Typically, color histograms [36, 43], color model parameters [45], regions of interest, and power spectral slopes [19] are extracted as features from raw RGB images, depth maps, and dark channel maps [44]. A regression model, such as a support vector regression, is then used to predict the level of haze and find the corresponding class from the index. Alternatively, multiple or cascaded support vector machines can be employed to directly classify images [44]. While this method enhances the classification robustness by fitting to a large amount of data, the selection of features limits the precision of model classification.

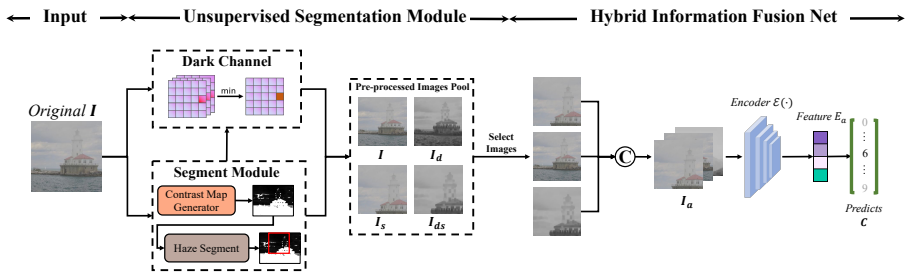

Fig. 2. Illustration of proposed methods HAINet. In forward propagation, the original image $\mathbf{I}$ goes through the USM module to get the pre-processed images pool including the original image $\mathbf{I}$, the dark channel of the original image $\mathbf{I}_d$, the segmented image $\mathbf{I}_s$ and the dark channel of segmented image $\mathbf{I}_{ds}$. The HIF module selects images from the pool and fuses them to generate the augmented image $\mathbf{I}_a$.

2.3 Haze Feature Learning method

The deep learning approach in deep learning allows the model to learn autonomously by extracting features and outputting classification results. However, since it is difficult to train models using complete haze images, related works have improved the model training framework. For instance, model ensemble methods [33, 41] achieve feature enhancement by training multiple basic models and adapting a meta-learner to learn to fuse the basic models' output. Similarly, the multi-branch method [26, 42] proposes multiple classifiers in different training branches to obtain better predictions. Pre-training [2, 7] and multi-task training [46, 47] are also viable options. Although the deep learning method does not require feature engineering, most current works lack constraints on the background problem of haze classification.

Overall, deep learning method has made significant progress in haze classification, but effective methods for effectively addressing the issue of background noise in haze images, which could significantly enhance the performance of haze classification models, are still lacking.

3 Technique

3.1 Framework Overview

As depicted in Figure 2, the proposed framework consists of two main modules: the USM module and the HIF module. The USM module serves as a preprocessing step for the input image and generates a processed image pool containing various versions of the original image, such as the dark channel of the original image, the segmented image, and the dark channel of the segmented image. On the other hand, the HIF module is responsible for selecting images from the processed image pool and concatenating them at the channel level for haze classification.

3.2 Unsupervised Segmentation Module

As illustrated in Figure 2, the unsupervised segmentation module is comprised of two sub-modules: the Dark Channel module and the Segment module. The

Dark Channel module is responsible for extracting the darkest pixel in the input image, while the Segment module utilizes the UCM [12] algorithm to extract haze information regions from the input image I. Unlike UCM, the Segment module focuses on the effect of different representations of haze images on the classification results. The output of this module is a pool of pre-processed images, including the original image I, the dark channel of the original image I_d, the segmented image I_s, and the dark channel of the segmented image I_{ds}.

Dark channel Module Existing work [8] has shown that the haze pixel values are stable across RGB channels and dark channels. Specifically, the dark channel image $\mathbf{I}_d$ is obtained by taking the minimum pixel value across the RGB channels at each pixel location in the original input image $\mathbf{I}$.

– **Calculate the Dark Channel Images.** This procedure aims to count the pixel information in the image $\mathbf{I}$. We use a filter $\mathbf{F}_d$ to get the channel map $\mathbf{I}_d$. The filter $\mathbf{F}_d$ uses the equation below:

$$\mathbf{F}_d(x) = \min_{y \in \Omega(x)} (\min_{c \in \{R,G,B\}} \mathbf{I}^c(y)), \tag{1}$$

where x represents a pixel, $\mathbf{I}^c(\cdot)$ denotes the c channel of image $\mathbf{I}$, $\Omega(x)$ means the neighboring pixels of x.

Segment Module Due to the excellent performance of UCM in segmenting haze, we have chosen to use it as the main program for the Segment module. However, we have made some modifications to the original method to reduce the computational burden and improve the efficiency of the program without sacrificing performance. Specifically, we have changed the method of acquiring the haze region by removing the step of calculating the denoising map. These improvements have resulted in faster running times without compromising the quality of the output. The improved steps are as follows.

– **Get the Contrastive map.**We use two filters $\mathbf{F}_b$ and $\mathbf{F}_d$ to get the contrastive image $\mathbf{I}_c$. $\mathbf{F}_d$ is defined as Eqs 1. $\mathbf{F}_b$ and $\mathbf{I}_c$ is defined as the following.

$$\mathbf{F}_b(x) = \max_{y \in \Omega(x)} (\max_{c \in \{R,G,B\}} \mathbf{I}^c(y)), \tag{2}$$

$$\mathbf{I}_{\mathrm{c}}(x) = \begin{cases} 255, & \text{if } x > v_{\text{threshold}} \text{ or } \mathrm{img}_{\text{ mean}} < \text{ gray} \\ x, & \text{otherwise} \end{cases}, \tag{3}$$

where $v_{threshold} = \frac{1}{2}\left(mean(v_c) + median(v_c)\right)$, $v_{\mathrm{c}} = I_{\mathrm{b}}(x) - I_d(x)$, $\mathrm{img}_{\mathrm{mean}} = \frac{1}{3}\sum_{c \in \{R,G,B\}}(I^c)$ and $\mathrm{gray} = 0.3I^R + 0.33I^G + 0.45I^B$.

– **Get the haze segmented map.** We use a filter $\mathbf{F}_p$ defined as Eqs 4 to get the position information of haze region. Then we can obtain the haze segmented map $\mathbf{I}_s$ according to the matrix $\mathbf{P}$ which get by the Eqs 5.

$$\mathbf{F}_p(x) = \max_{y \in \Omega(x)} \mathbf{I}_c(y), \tag{4}$$

$$\mathbf{P}(x) = \begin{cases} 0 & \text{if } \mathbf{F}_p(x) == 255 \\ 1 & \text{otherwise} \end{cases}. \tag{5}$$

.

3.3 Hybrid Information Fusion Module

The HIF module is proposed to address the problem of the influence of background information. This module selects pre-processed images from the processed image pool and fuses them for classification. As shown in Figure 2, the module outputs the haze level classification result.

Input Augmentation The input to this module is a multi-channel hybrid map, which is created by concatenating the selected pre-processed images. An encoder ϵ_a is utilized to extract the feature. Then the feature $\mathbf{E}a$ is forwarded to the classifier for the purpose of classification. The multi-channel hybrid map $\mathbf{I}a$ is acquired by utilizing the equations outlined in Eq. 6.

$$\mathbf{I}_a = concat(\mathbf{I}, \mathbf{I}_p). \tag{6}$$

where $\mathbf{I}_p$ contains $[I_d, I_s, I_{ds}]$

Training Strategies HIF is optimized by minimizing loss $\mathcal{L}_c$.

$$\mathcal{L}_c = CE(C, C^{'}). \tag{7}$$

where C means the predicated label and $C^{'}$ means the groundtruth.

4 Experiments

4.1 Datasets

Experiments were conducted on two datasets to evaluate visual haze classification. One is Hazel-level, while the other is Haze-wild. Further details regarding the two datasets are provided below:

- **The Image Hazel-level Dataset (Hazel-level Dataset)**: The image Hazel-level dataset [18] contains 3024 synthetic images with 9 classes. These images are based on the algorithms and dataset provided by the FRIDA dataset.
- **Haze-Wild Dataset**: Source datasets contain 5000 sunny and 5000 cloudy images are used to generate 10 levels of fogging images with original images as level 0 through a monocular depth estimation model [5].

4.2 Implementation details

We implemented the proposed methods and the algorithms in comparison by python. For proposed HAINet framework, the model details are as follows:

- As for unsupervised segmentation module, We set the parameters as the setting of the original paper [12]. For the HIF module, the parameter settings vary depending on the dataset used for experimentation. Specifically, on the Haze-level dataset, batch size of 16 and a learning rate of 5e-4 are utilized. On the Haze-wild dataset, batch size of 128 and a learning rate of 1e-3 are used. In both cases, the Adam optimizer is employed. Additionally, the decay is set to 0.1 or 0.5 for every N epoches. The base models used in the paper are ResNet18 and ResNet50 [9]. The original image, along with the pre-processed images, is resized to a size of 64×64.

4.3 Performance Comparison

This section presents a performance comparison between HAINet and existing haze classification methods, including three pixel value statistics methods: Saturation & RGB-correlation Detection [1], Filter-Based Fog Detection [10], HSV-Based Fog Detection [27], the engineering method SVM [45], and haze feature learning method CNN PAPLE [41], LeNet5 [11], ResNet18 [9], ResNet50 [9] and VIT [4]. For both algorithms, we fine-tune their hyper-parameters to obtain their best performance in the experiments. We can observe the followings as shown in the table 1:

Table 1. Precision comparison of haze classification algorithms on Hazel-level and Haze-wild datasets.

Type	Model	Datasets	
		Hazel-level	Haze-Wild
Pixel Value Statistics Method	Saturation & RGB-correlation [1]	0.1529	0.1018
	Filter-Based [10]	0.2063	0.2121
	HSV-Based [27]	0.2431	0.1779
Engineering Method	SVM [45]	0.4532	0.2601
Haze Feature Learning Method	PAPLE [41]	0.7636	0.5181
	LeNet5 [11]	0.8745	0.6843
	ResNet18 [9]	0.8998	0.7650
	ResNet50 [9]	0.9031	0.7694
	ViT [4]	0.8459	0.7890
	HAINet(ResNet50)	**0.9372**	0.8102
	HAINet(ResNet18)	0.9328	**0.8320**

- Among all the methods evaluated, the precision of the Haze-Wild dataset was found to be lower than that of the Haze-level dataset. This phenomenon suggests that images with more complex backgrounds can negatively impact the precision of haze classification.
- The pixel value statistics methods were observed to perform poorly on both datasets, which could be attributed to the fact that haze levels cannot be accurately assessed solely based on numerical values. These methods are commonly used to identify the presence of haze in an input image, and the final haze value can be easily affected by background factors when measured in numerical terms.
- In comparison to the pixel value statistics methods, the engineering method has been observed to achieve a performance improvement of 46.21% on both datasets. However, the performance of this method is constrained by the selection of features utilized.
- Among the haze feature learning methods, the CNN method has been found to outperform the handcrafted feature extraction method, achieving a precision improvement of 64.07% on the Hazel-level dataset and 99.19% on the Haze-Wild dataset, which has a simpler background. This result highlights the advantages of data fitting.
- **ViT Performance Analysis.** Although ViT achieves competitive results, its performance is still inferior to that of ResNet on the Hazel-level dataset. This could be attributed to the synthetic backgrounds used in the hazel-level dataset, which may not fully represent real-life situations. Another reason for the disparity could be the challenge of transferring pre-trained knowledge from traditional image classification tasks to the haze classification task. On the other hand, ViT performs better than ResNet on the Haze-Wild dataset, suggesting its potential for handling real-life haze scenarios.
- The proposed method has been found to achieve a significant precision improvement over existing methods that use different backbones on both datasets. This demonstrates the effectiveness of removing background factors from haze images to enhance the performance of the network.

In summary, pixel value statistics and engineering methods have limited results due to information loss and assumptions that may not match the data. In contrast, the proposed HAINet achieves significantly better precision on both the Hazel-level and Haze-

Table 2. Classification precision of different selections of pre-processed images on Hazel-level dataset and Haze-Wild dataset. O = the original image; D = the dark channel of original image; O/D(HRS) = the segmented image of O or D

Model	Hazel-level		Haze-Wild	
	ResNet18	ResNet50	ResNet18	ResNet50
O	0.8998	0.9031	0.7650	0.7694
+O(HRS)	0.9251	0.9196	0.8170	0.7878
+D(HRS)	0.9196	0.9262	0.8179	0.8137
+O(HRS)+D	0.9240	0.9207	0.8274	0.7423
+D(HRS)+D	0.9284	0.9328	0.8287	**0.8236**
+O(HRS)+D(HRS)+D	**0.9328**	**0.9372**	**0.8320**	0.8102

wild datasets compared to existing methods with different backbones. This demonstrates the effectiveness of using haze images with background factors removed to improve network performance.

4.4 Ablation Study

In this section, we investigate the effect of the pre-processed images on the classification results under different methods.

Evaluation on the Input Augmentation with segmented images: As illustrated in Table 2, the "+O(HRS)" and "+D(HRS)" rows indicate that using the original images with the augmentation of segmented images as input, both achieve better performance on two datasets and backbones. Notably, except for the case of using ResNet18 on the Hazel-level dataset, it was found that the augmentation of dark channel segmented images outperforms the augmentation of original segmented images alone. This suggests that the dark channel of segmented images contains more informative features about haze, which can be effectively captured by the model. Overall, these results demonstrate the importance of incorporating segmented images as an augmentation strategy for haze classification.

Evaluation on the Input Augmentation with segmented images and the dark channel of original images: As shown in the row "+O(HRS)+D" and "+D(HRS)+D" of the Table 2, the performance improvement achieved by these augmentations is significant compared to the previously mentioned. Specifically, the classification results obtained using the "+D(HRS)+D" augmentation outperform those obtained using the "+O(HRS)+D" augmentation, albeit only slightly. This finding further highlights the importance of removing background factors in haze classification. By reducing the impact of background information, the model is able to focus more on the informative features of the haze itself, leading to improved performance.

Evaluation on Input Augmentation with all pre-processed images: As shown in the row "+O(HRS)+D(HRS)+D" of Table 2, the model achieves the best performance on the Hazel-level dataset and the Haze-Wild dataset except using ResNet50. Even when using ResNet50 on the Haze-Wild dataset, the model still shows significant improvement compared to the base model. This finding suggests that by combining the respective advantages of segmented and dark channel images, the model can better capture informative features of haze and thus improve classification performance. Overall, the results indicate the importance of leveraging background factors and incorporating appropriate input augmentation techniques for haze classification.

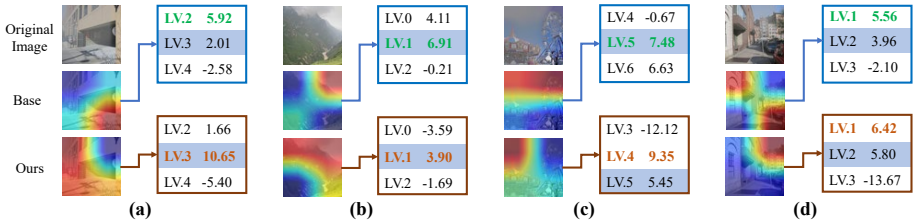

Fig. 3. Visualization of feature attentions and model predictions of Base model (the blue branch) and HAINet (the brown branch). We use GradCAM [34] to generate heatmaps and show outputs near the ground truth, the ground-truth level is emphasized by background color, while the predicted level is marked in bolded blue and bolded brown respectively.

4.5 Case Study

In this section, we utilized GradCAM [34] to investigate the variation in model focus when using haze images with removed background information as an input enhancement method versus using direct input of the original images for haze classification. The displayed images were randomly chosen from the test set of the Haze-Wild dataset, and ResNet18 was utilized as the backbone.

As shown in Figure 3 (a), when using the original input, the model predicts the wrong class, but after adding the input enhancement method, the model correctly predicts the haze class. The heat map generated by GradCAM reveals that the base model is more focused on identifying possible objects in the image rather than the haze region when the original haze image is input. In the specific example, the base model tends to focus on the lower right corner region, while the haze region is concentrated in the left rear view, leading the model to make an incorrect prediction. However, after incorporating the image with the background information removed, the model is able to focus better on the left region where the haze exists and extract more informative features about the haze, ultimately leading to a correct prediction of the haze class. This highlights the importance of input enhancement in improving model attention and performance in haze classification tasks.

As shown in Figure 3 (b), both the base model using the original image as input and the HAINet model using input augmentation achieve correct prediction results. However, in the base model using the original image as input, the heat map shows that its focus is concentrated on the upper right corner and a small part of the lower left corner area, while the haze area is mainly concentrated in the upper left corner area, and the base model only focuses on a small part of the haze area, thus predicting the correct haze level. The heat map shows the model focuses on the entire haze area and concentrates on the haze concentration area to get the correct haze level.

In Figure 3 (c) it can be seen the base model successfully predicts the correct haze level, while the model enhanced by the input augmentation incorrectly predicts the level. The haze in this image is mainly concentrated on the left side of the Ferris wheel and above the roof. The heat map indicates the base model's attention is mainly focused on the haze region, resulting in a correct classification result. However, the focus of the base model also extends to the Ferris wheel region, indicating the model may have been distracted by other elements in the image. After augmentation, the model effectively narrows its focus on the haze region, but in this particular case, it may have overlooked some contextual cues that were helpful for the correct classification. Despite the wrong prediction, the model's output is still informative. The predicted level is closely related to the true level and has much higher values than the other levels.

In Figure 3 (d), the base model and HAINet model using input augmentation both predict the wrong haze level. The heat map illustrates that the base model concentrates on identifying objects in the picture, such as the vehicles in the lower right corner and the buildings in the upper left corner. However, after applying input enhancement, the HAINet model prioritizes the haze region in the upper right corner and neglects the objects in the picture, resulting in a different but still incorrect prediction. Notably, the predicted values of HAINet are closer to the actual values compared to those of the base model, indicating better performance.

Overall, the classification of haze images is negatively affected by the interference of background information. The approach presented in this study effectively eliminates this interference by removing the background information and using it as a hard attention mechanism to direct the model's focus toward the haze region. This method proves to be effective in improving the performance of haze classification tasks.

5 Conclusion

This paper introduces an approach named HAINet that effectively tackles the challenge of separating background information in haze images and improving haze classification performance. Conventional classification methods often prioritize object detection in images and disregard haze regions, which is not ideal for the haze classification task. In our proposed method, the unsupervised segmentation module separates the background information in input images and generates multiple images containing haze information. The HAINet model implements a hard attention mechanism, which focuses the model's attention on the haze region. Experimental results show that our method successfully shifts the model's attention from objects to the haze region, leading to a significant improvement in haze classification performance.

In future work, we will investigate more suitable image fusion mechanisms that can better integrate the information from multiple images with mainly haze information, further improving the model's ability to focus on the haze region. Additionally, we plan to further enhance the robustness of the background separation technique, making the hard attention mechanism applicable to a wider range of image classification tasks beyond haze classification.

Acknowledge

This work is supported by TaiShan Scholars Program (Grant no. tsqn202211289) and Excellent Youth Scholars Program of Shandong Province (Grant no. 2022HWYQ-048).

References

1. Alami, S., Ezzine, A., Elhassouni, F.: Local fog detection based on saturation and rgb-correlation. In: 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV). pp. 1–5. IEEE (2016)
2. Chakma, A., Vizena, B., Cao, T., Lin, J., Zhang, J.: Image-based air quality analysis using deep convolutional neural network. In: IEEE ICIP. pp. 3949–3952 (2017)
3. Chen, Y., Wang, J., Li, S., Wang, W.: Multi-feature based foggy image classification. In: IOP Conference Series: Earth and Environmental Science. vol. 234 (2019)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929
5. Godard, C., Mac Aodha, O., Firman, M., Brostow, G.J.: Digging into self-supervised monocular depth prediction (October 2019)
6. Guan, Q.L., Zheng, Y., Meng, L., Dong, L.Q., Hao, Q.: Improving the generalization of visual classification models across iot cameras via cross-modal inference and fusion. IEEE Internet of Things Journal (2023)

7. Guo, L., Song, J., Li, X.r., Huang, H., Du, J.j., He, Y.c., Wang, C.z.: Haze image classification method based on alexnet network transfer model. In: Journal of Physics: Conference Series. vol. 1176, p. 032011. IOP Publishing (2019)
8. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence pp. 2341–2353 (2010)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the CVPR. pp. 770–778 (2016)
10. Jeong, K., Choi, K., Kim, D., Song, B.C.: Fast fog detection for de-fogging of road driving images. IEICE TRANSACTIONS on Information and Systems (2018)
11. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
12. Li, J., Ma, H., Li, X., Meng, L., Meng, X.: Unsupervised contrastive masking for visual haze classification. In: Proceedings of ICMR (2022)
13. Li, R., Liu, X., Li, X.: Estimation of the pm2. 5 pollution levels in beijing based on nighttime light data from the defense meteorological satellite program-operational linescan system. Atmosphere **6**(5), 607–622 (2015)
14. Li, X., Wu, L., Chen, X., Meng, L., Meng, X.: Dse-net: Artistic font image synthesis via disentangled style encoding. In: 2022 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2022)
15. Li, X., Wu, L., Wang, C., Meng, L., Meng, X.: Compositional zero-shot artistic font synthesis. Proceedings of IJCAI (2023)
16. Li, X., Ma, H., Meng, L., Meng, X.: Comparative study of adversarial training methods for long-tailed classification. In: Proceedings of the 1st International Workshop on Adversarial Learning for Multimedia. pp. 1–7 (2021)
17. Li, X., Zheng, Y., Ma, H., Qi, Z., Meng, X., Meng, L.: Cross-modal learning using privileged information for long-tailed image classification. CVM (2023)
18. Li, Y., Huang, J., Luo, J.: Using user generated online photos to estimate and monitor air pollution in major cities. In: Proceedings of the 7th International Conference on Internet Multimedia Computing and Service. pp. 1–5 (2015)
19. Liu, C., Tsow, F., Zou, Y., Tao, N.: Particle pollution estimation based on image analysis. PloS one **11**(2), e0145955 (2016)
20. Liu, J., Xiao, J., Ma, H., Li, X., Qi, Z., Meng, X., Meng, L.: Prompt learning with cross-modal feature alignment for visual domain adaptation. In: CAAI (2022)
21. Liu, T., Qi, Z., Chen, Z., Meng, X., Meng, L.: Cross-training with prototypical distillation for improving the generalization of federated learning. ICME (2023)
22. Ma, H., Li, X., Meng, L., Meng, X.: Comparative study of adversarial training methods for cold-start recommendation. In: Proceedings of ADVM (2021)
23. Ma, H., Qi, Z., Dong, X., Li, X., Zheng, Y., Meng, X.M.L.: Cross-modal content inference and feature enrichment for cold-start recommendation. IJCNN (2023)
24. Ma, H., Xie, R., Meng, L., Chen, X., Zhang, X., Lin, L., Zhou, J.: Exploring false hard negative sample in cross-domain recommendation. In: Recsys (2023)
25. Ma, H., Xie, R., Meng, L., Chen, X., Zhang, X., Lin, L., Zhou, J.: Triple sequence learning for cross-domain recommendation. arXiv preprint arXiv:2304.05027 (2023)
26. Ma, J., Li, K., Han, Y., Yang, J.: Image-based air pollution estimation using hybrid convolutional neural network. In: ICPR. pp. 471–476. IEEE (2018)
27. Mao, J., Phommasak, U., Watanabe, S., Shioya, H.: Detecting foggy images and estimating the haze degree factor. Journal of Computer Science & Systems Biology **7**(6), 226–228 (2014)
28. Mei, S., Li, H., Fan, J., Zhu, X., Dyer, C.R.: Inferring air pollution by sniffing social media. In: 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014). pp. 534–539. IEEE (2014)

29. Meng, L., Chen, L., Yang, X., Tao, D., Zhang, H., Miao, C., Chua, T.S.: Learning using privileged information for food recognition. In: Proceedings of the 27th ACM International Conference on Multimedia. pp. 557–565 (2019)
30. Meng, L., Feng, F., He, X., Gao, X., Chua, T.S.: Heterogeneous fusion of semantic and collaborative information for visually-aware food recommendation. In: Proceedings of MM (2020)
31. Qi, Z., Chen, X.: A novel density-based outlier detection method using key attributes. Intelligent Data Analysis **26**(6), 1431–1449 (2022)
32. Qi, Z., Wang, Y., Chen, Z., Wang, R., Meng, X., Meng, L.: Clustering-based curriculum construction for sample-balanced federated learning. In: CAAI International Conference on Artificial Intelligence. pp. 155–166. Springer (2022)
33. Rijal, N., Gutta, R.T., Cao, T., Lin, J., Bo, Q., Zhang, J.: Ensemble of deep neural networks for estimating particulate matter from images. In: ICIVC (2018)
34. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of ICCV. pp. 618–626 (2017)
35. Sun, W., Li, X., Li, M., Wang, Y., Zheng, Y., Meng, X., Meng, L.: Sequential fusion of multi-view video frames for 3d scene generation. In: CAAI International Conference on Artificial Intelligence. pp. 597–608. Springer (2022)
36. Wang, H., Yuan, X., Wang, X., Zhang, Y., Dai, Q.: Real-time air quality estimation based on color image processing. In: 2014 IEEE Visual Communications and Image Processing Conference. pp. 326–329. IEEE (2014)
37. Wang, X., Zhang, L., Bo, Q., Feng, J., Hu, J., Kang, Y., Zhang, J.: Feature enhancement and fusion for image-based particle matter estimation with f-mse loss. In: IEEE ICIP. pp. 768–772. IEEE (2020)
38. Wang, Y., Li, X., Ma, H., Qi, Z., Meng, X., Meng, L.: Causal inference with sample balancing for out-of-distribution detection in visual classification. In: CAAI International Conference on Artificial Intelligence. pp. 572–583. Springer (2022)
39. Wang, Y., Li, X., Qi, Z., Li, J., Li, X., Meng, X., Meng, L.: Meta-causal feature learning for out-of-distribution generalization. In: European Conference on Computer Vision. pp. 530–545. Springer (2022)
40. Wang, Y., Qi, Z., Li, X., Liu, J., Meng, X., Meng, L.: Multi-channel attentive weighting of visual frames for multimodal video classification. IJCNN (2023)
41. Zhang, C., Yan, J., Li, C., Rui, X., Liu, L., Bie, R.: On estimating air pollution from photos using convolutional neural network. In: Proceedings of MM (2016)
42. Zhang, C., Yan, J., Li, C., Wu, H., Bie, R.: End-to-end learning for image-based air quality level estimation. Machine Vision and Applications **29**(4), 601–615 (2018)
43. Zhang, Y., Sun, G., Ren, Q., Zhao, D.: Foggy images classification based on features extraction and svm. In: Proceeding of 2013 International Conference on Software Engineering and Computer Science. pp. 142–14 (2013)
44. Zhang, Z., Ma, H., Fu, H., Liu, L., Zhang, C.: Outdoor air quality level inference via surveillance cameras. Mobile Information Systems **2016** (2016)
45. Zhang, Z., Ma, H., Fu, H., Wang, X.: Outdoor air quality inference from single image. In: International Conference on Multimedia Modeling (2015)
46. Zhao, X., Jiang, J., Feng, K., Wu, B., Luan, J., Ji, M.: The method of classifying fog level of outdoor video images based on convolutional neural networks. Journal of the Indian Society of Remote Sensing **49**(9), 2261–2271 (2021)
47. Zhao, X., Zhang, T., Chen, W., Wu, W.: Image dehazing based on haze degree classification. In: CAC. pp. 4186–4191. IEEE (2020)