RESEARCH ARTICLE

A combined super-resolution and semantic segmentation approach of urban landscape image in rain and fog environment

He Jiang[1], Xiang Wu[2, *], Yuanhao Ma[2], Yinduo Bai[2], Yaohuan Tan[2], Jiajian Zhang[2], Yuxi Zhang[2]

[1]Interdisciplinary Creative Application Research (ICAR), College of Landscape Architecture and Art, Henan Agricultural University, Zhengzhou, Henan, China. [2]College of Electrical Engineering, Henan University of Technology, Zhengzhou, Henan, China.

Received: December 18, 2024; **accepted:** March 17, 2025.

Recent studies have encountered a significant challenge in understanding landscape images using artificial approaches. To address the issues of decreased segmentation accuracy and insufficient detail of fog areas in foggy scenes, a fusion network integrating de-fogging, hyper-division, and segmentation was proposed. The All-in-One Network (AOT-Net) for Dehazing algorithm that took an atmospheric scattering model was first utilized to effectively eliminate noise and fog from the image. Subsequently, Real Enhanced Super-Resolution Generative Adversarial Networks (Real-ESRGAN) super-resolution technology was introduced to enhance image quality post-fog removal by filling in missing pixels, restoring edge details, and addressing issues such as excessive contrast and dark tones. The results indicated that this improved approach achieved outstanding performance on the Uavid datasets with added fog noise, showing a 4.4% increase in Mean Intersection Over Union (mIoU) compared to the original U-shaped Network Transformer (UNetFormer) model. This method significantly improved segmentation accuracy in foggy or rainy conditions and demonstrated its potential for large-scale data processing in smart city development and related landscape applications.

Keywords: urban landscape; deep learning; foggy weather; image segmentation.

[*]**Corresponding author:** Xiang Wu, College of Electrical Engineering, Henan University of Technology, Zhengzhou, Henan, China. Email: xiangw@haut.edu.cn or baoshesuit8900@163.com.

Introduction

Urban landscapes are shaped by streets, buildings, and vegetation, which significantly affect the health, lifestyles, and well-being of residents. Accurate quantification of these features is vital for analyzing urban growth patterns [1]. To achieve such precise quantification, semantic segmentation has become an essential technique as it allows for the automatic identification and classification of urban elements at the pixel level. This approach enables detailed analysis of urban environments, providing valuable insights into the distribution and structure of various features within a city. In this context, Unmanned Aerial Vehicles (UAVs) have become an invaluable tool for urban landscape analysis [2]. UAVs enable the efficient collection of high-resolution aerial imagery over large areas, providing a flexible and cost-effective means of obtaining detailed data for semantic segmentation tasks. Their ability to capture real-time, high-quality images from various angles greatly enhances the accuracy of urban feature extraction, making them ideal for large-scale environmental monitoring and urban planning.

However, adverse weather conditions such as fog can severely diminish image clarity, reduce visibility, and obscure critical details, which presents significant challenges for semantic segmentation, affecting the quality of the captured data and complicating further processing tasks [3, 4].

Traditional image segmentation methods such as thresholding, region growing, and edge detection struggle with the complexity of urban environments, where features are variable and often overlap. These methods are further hindered by challenges like varying illumination and shadows, making precise segmentation difficult. Consequently, deep learning-based models have become the go-to solution for semantic segmentation. Unlike traditional approaches, deep convolutional neural networks (CNNs) can learn complex features from raw data, enabling better capture of urban scene details. Early deep convolution-based models such as Fully Convolutional Networks (FCN) marked a significant improvement in segmentation accuracy by enabling pixel-level classification [5]. Subsequent advancements introduced more sophisticated architecture to address the limitations of early models. In 2015, Olaf Ronneberger introduced the UNet model, featuring a symmetric encoder-decoder structure that allowed for the restoration of fine details in the image, improving segmentation precision [6]. The Segmentation Network (SegNet) model optimized the pooling process by preserving information in the encoder's pooling layers and refining the decoder process to recover spatial details [7]. The DeeplabV3+ model incorporated depthwise separable convolutions and an Atrous Spatial Pyramid Pooling (ASPP) module, which enhanced the model's ability to segment multi-scale objects [8]. However, despite significant advancements in deep learning-based segmentation models, accurate semantic segmentation in foggy conditions remains a challenge as fog severely impairs image quality. Image de-fogging techniques are therefore critical for overcoming these challenges and can be categorized into three primary methods including physical models, non-physical models, and deep learning-based models. Physical models typically use atmospheric scattering theory to replicate the scattering and absorption of light caused by fog, thereby restoring clear images. A widely used defogging algorithm based on this theory is the dark channel prior model, which has proven effective due to its high stability and efficiency in handling fog-related distortion [9]. Non-physical models, on the other hand, do not rely on the physical imaging process but instead apply image processing techniques to enhance the visual quality by adjusting the intensity distribution of the image. One such method is histogram equalization, which enhances the contrast by modifying the distribution of gray levels in the image [10]. Deep learning-based de-fogging methods have also made significant strides. The Dehaze Networks (DehazeNet) developed by Cai *et al*. is a deep learning-based system that uses convolutional neural networks (CNNs) to assess atmospheric haze factors and enhance image quality [11]. The key concept behind DehazeNet is its ability to map foggy images directly to their clear counterparts using CNNs. This approach has been improved by subsequent models like NIN-DehazeNet and Light-DehazeNet, which have optimized the alignment and adapted the model to different application scenarios [12, 13].

Given the importance of both image de-fogging and semantic segmentation, achieving accurate urban streetscape segmentation in foggy conditions requires an integrated approach. Combining de-fogging, super-resolution techniques and deep learning-based semantic segmentation has shown promise in overcoming the challenges caused by fog. This integrated solution improves image resolution, restores fine details, and enhances contrast, all of which are crucial for boosting segmentation accuracy [14]. This research proposed a method that effectively combined de-fogging, super-resolution, and semantic segmentation networks to process UAV imagery in adverse weather, ultimately enabling accurate urban streetscape segmentation even under foggy conditions. The proposed method not only addressed significant challenges in current segmentation tasks but also laid the groundwork for future applications in urban planning, environmental monitoring, and other areas requiring precise image analysis under complex weather conditions.

Materials and methods

Urban landscape segmentation based on super-resolution image enhancement in fog environment

When segmenting urban landscapes in foggy conditions, image quality often deteriorates, complicating the accurate extraction of relevant information. To solve the above problems, this study applied image dehazing techniques to reduce blurriness and contrast loss caused by fog and enhance clarity and visibility. However, details might be lost during dehazing, prompting the use of super-resolution enhancement techniques to recover these details and improve resolution, ensuring image quality meeting segmentation needs. The processed images underwent semantic segmentation using the UnetFormer network, accurately identifying key elements like buildings, roads, and vegetation, thereby providing essential semantic information to support urban management and landscape analysis.

(1) Image defogging algorithm

Under the influence of fog, urban streetscape images captured from a UAV perspective were prone to interference, leading to a decline in image quality that subsequently affected downstream analysis and processing tasks. In this context, the adoption of an efficient dehazing algorithm became particularly crucial. AOD-Net with its lightweight architecture and real-time processing capabilities had emerged as an ideal choice for dehazing aerial images captured by drones [15]. AOD-Net relied on an atmospheric model, indicating that image degradation was mainly due to atmospheric scattering effects. Traditional algorithms estimated the transmittance and atmospheric light separately, which not only incurred high computational costs but also introduced errors. In contrast, AOD-Net merged these two elements into a single transition matrix, simplifying its structure, reducing computational complexity, enhancing efficiency, and minimizing errors. Its network architecture employed a multi-scale feature fusion network, effectively capturing information across various scales to accurately estimate transmittance. Ultimately, AOD-Net utilized the simplified model to generate dehazed images through the transition matrix. This lightweight design made it highly suitable for real-time dehazing tasks in UAV applications. The AOD-Net model consisted of the following five steps. Step 1 was the atmospheric scattering model that represented a simplified relationship between a clear image and its foggy version expressed as follows.

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

where *t(x)* was the transmission map. *A(*)* represented the atmospheric light value. To recover a clear image, accurate estimation of both values was essential. In AOD-Net, the two unknowns in Equation (1) were combined into a single variable K through mathematical transformations. The solution for K was shown in equation (2), while the simplified atmospheric scattering model was presented in equation (3).

$$K(x) = \frac{A - 1 + \frac{(I(x) - A)}{t(x)}}{I(x) - 1} \quad (2)$$

$$J(x) - 1 = K(x)I(x) - K(x) \quad (3)$$

Step 2 was the multi-scale feature fusion network leveraged a multi-scale fusion strategy to extract fog features, enhancing its overall performance (Figure 1). The architecture comprised five convolution layers and three merging layers. These convolution layers used kernels of varying sizes to capture features at multiple scales, enabling the network to gather diverse image details. The merging layers integrated feature maps from these scales, enhancing feature extraction and minimizing information loss during the convolution process [16]. Each convolution layer in the network employed only three kernels, creating a simple, shallow structure without complex branches. This streamlined design significantly reduced processing time, enabling the algorithm to achieve real-time performance.

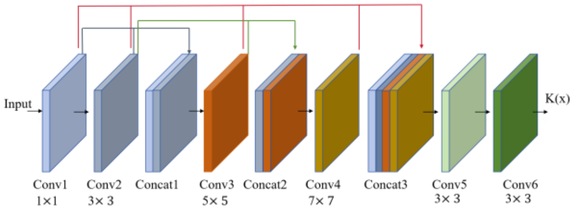

Figure 1. Network structure of AOD-Net.

Step 3 was the generation of foggy dataset. AOD-Net augmented the NYU2 indoor clear image dataset by applying various atmospheric light intensities and scattering coefficients, producing a synthetic foggy dataset with different fog density levels. Step 4 was the designing the loss function for network training by employing the mean squared error (MSE) loss as described below.

$$L_{\mathrm{MSE}} = (J_i - f(x_i))^2 \tag{4}$$

where $\mathrm{x_i}$ was the foggy image input to the network. $\mathrm{J_i}$ was the corresponding clear image of the synthetic fog data. $\mathrm{f(x_i)}$ was the dehazed image generated by the AOD-Net network. Step 5 was that, after training, the AOD-Net network could get the weight file of each layer of the trained network, load the weight file, and read the fog image with AOD-Net to directly obtain the de-fog image.

(2) Super resolution image enhancement

Although the AOD-Net dehazing algorithm demonstrated exceptional efficiency, the resulting images often suffered from issues such as incomplete dehazing, excessive contrast, blurred edge details, and dark tones. These blurred images obscured building outlines and made road signs difficult to discern, negatively impacting subsequent segmentation tasks. The introduction of super-resolution techniques could significantly enhance image quality by not only supplementing missing pixels and restoring details but also improving contrast and color accuracy, which was particularly beneficial for extracting critical features like buildings, roads, and vegetation in urban street scene recognition. In this study, Real-ESRGAN technology was used for real-world scenarios [17]. The Real-ESRGAN framework built upon Generative Adversarial Networks (GANs) applied high-order degradation modeling to replicate actual degradation effects. This technique accounted for various image acquisition conditions like noise, blurring, and color distortion, enhancing the realism and accuracy of the restoration results. The classical degradation model involved convolving the ground truth image y with a blur kernel k, then downsampling, adding noise, and applying JPEG compression as described in equation (5) [18]. The high-order degradation model reproduced the degradation present in real images by repeatedly using the classical degradation procedure with n representing the number of iterations as shown in equation (6). Additionally, Real-ESRGAN utilized a sinc filter to minimize ringing and overshoot artifacts as illustrated in equation (7), where i and j denoted the filter coordinates, and ω_c was the cutoff frequency. By implementing high-order degradation and sinc filtering, Real-ESRGAN effectively mimicked real-world image degradation, ultimately improving image quality.

$$x = D(y) = [(y \# \ k) \downarrow_r \ + n]_{\mathrm{JPEG}} \tag{5}$$

$$x = D^n(y) = (D_n \cdots D_2 D_1)(y) \tag{6}$$

$$k(i, j) = \frac{\omega_c}{2\pi\sqrt{i^2 + j^2}} J_1\left(\omega_c \sqrt{i^2 + j^2}\right) \tag{7}$$

The generator structure of the Real-ESRGAN model consisted of convolutional layers, 16 sequentially connected Residual-in-Residual

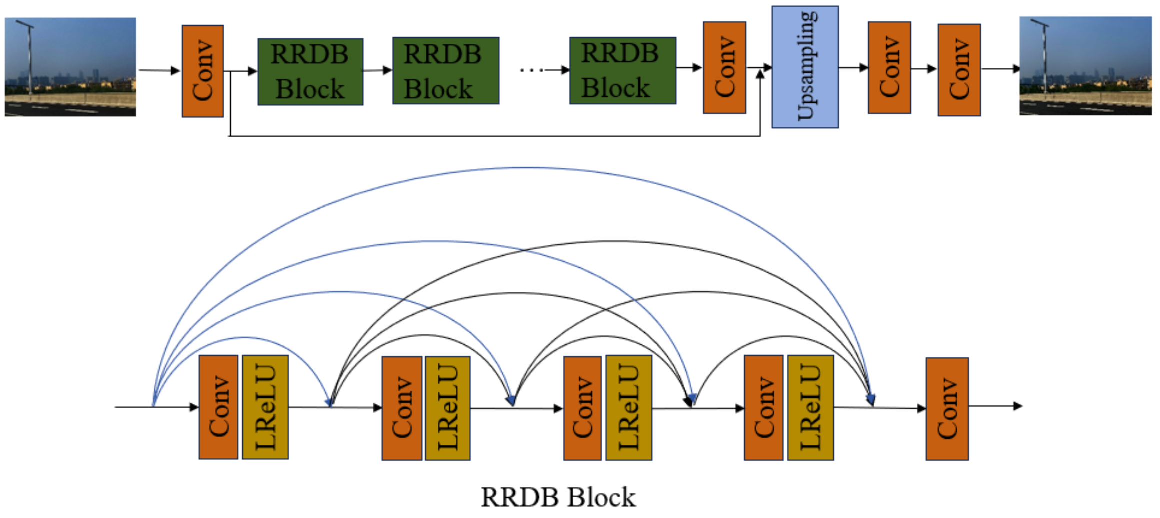

Figure 2. The generator structure of the Real-ESRGAN model.

Dense Blocks (RRDBs), up-sampling layers, and a convolutional output layer (Figure 2). In contrast to the ESRGAN model, the discriminator architecture replaced the Visual Geometry Group Network (VGG) with a U-Net model that incorporated Spectral Normalization (SN) (Figure 3). This design allowed the discriminator to assess the generated images from a pixel-level perspective, enabling it to maintain the overall realism of the images while also focusing on intricate details.

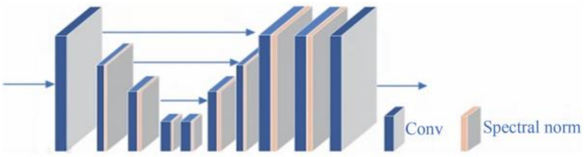

Figure 3. Structure of Real-ESRGAN model discriminator (U-Net model with SN).

(3) Semantic segmentation in urban streetscape

Semantic segmentation of urban streetscapes is widely used in land cover measurement, urban development monitoring, environmental protection, and economic planning. The Fully Convolutional Network (FCN) is an end-to-end architecture for semantic segmentation, establishing the groundwork for using Convolutional Neural Networks (CNNs) in this domain. Despite its innovation, FCN's decoder structure is relatively simple, resulting in low-resolution outputs that affect segmentation precision. To overcome this limitation, UNet implemented a balanced encoder-decoder framework, where features were extracted *via* downsampling and resolution is restored through upsampling [19]. Although CNN-based encoder-decoder techniques had advanced, they struggled in complex urban environments because they primarily captured local details and fail to account for global context. Transformers, on the other hand, transformed 2D image tasks into 1D sequences, allowing them to capture global information more effectively and achieve better results in core vision tasks [20]. UNetFormer was an innovative network architecture that seamlessly integrated the encoder-decoder framework based on CNNs with the transformer's exceptional global information modeling capabilities (Figure 4). In these environments, the images might suffer from incomplete defogging or blurring, leading to a decline in segmentation accuracy with traditional methods. UNetFormer effectively addressed these challenges by precisely capturing local details and integrating global information. By enhancing the recognition and fidelity of key details such as building outlines and road boundaries, this network achieved high accuracy and clarity in image segmentation under complex conditions, significantly advancing the capabilities of remote sensing image analysis.

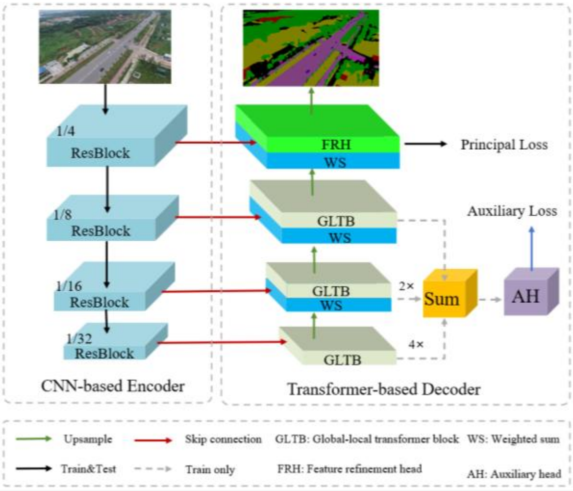

Figure 4. UnetFormer flowchart of the algorithm.

The CNN network's encoder used ResNet18 as its base and was divided into four stages. Down sampling was taken in each stage by a factor of 2. The feature maps of each stage were linked to those of the decoder *via* skip connections using 1×1 convolutions with a channel size of 64. This process involved calculating a weighted sum to combine the semantic features from the Residual Block (ResBlock) with those from the decoder's Global-Local Transformer Blocks (GLTB) [21]. The weighting was adjusted based on how much each feature contributed to segmentation accuracy with the specific expression as follows.

$$FF = \alpha \cdot RF + (1-\alpha) \cdot GLF \tag{8}$$

where FF was the fused feature created by combining features from multiple sources. RF was the features extracted by the ResBlock. GLF was the features generated by the GLTB. The decoder section consisted of three GLTBs and a Feature Refinement Head (FRH), forming a streamlined transformer-based decoder. The Global-Local Transformer Block (GLTB) employed dual parallel branches to extract both global and local information (Figure 5). The local branch included two convolutional layers with kernel sizes of 1 and 3 followed by batch normalization to stabilize the training process. The global branch, which was more intricate, implemented a self-attention mechanism to capture global image features, which transformed the 2D feature map into Query, Key, and Value vectors. The Query vector targeted specific location details, while the Key stored reference information, and the Value vector held the corresponding content. By computing the similarity between the Query and Key vectors, the network identified relevant areas and adjusted the Value vector to create a global context. To improve spatial understanding, the model integrated relative positional bias and a cross-shaped window context interaction, capturing long-range features horizontally and vertically, enhancing global information exchange. Softmax normalization was then applied to generate the global context representation, enabling effective processing of long-distance pixel relationships, particularly in complex urban street scenes.

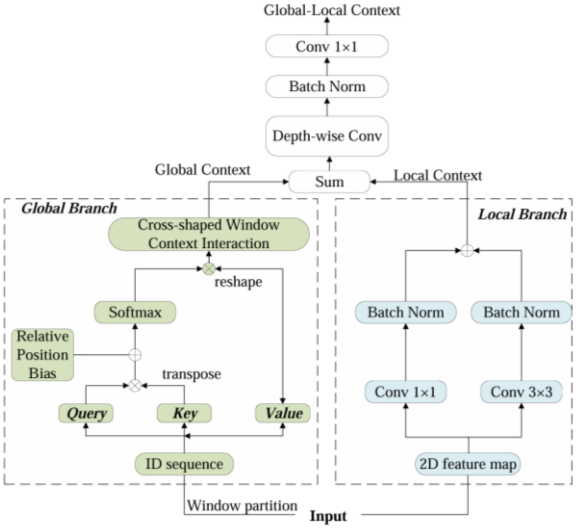

Figure 5. Diagram of global-local attention.

Experimental dataset

The UAVid dataset was employed in this study [22], which was provided by a collaboration among three universities including the University of Twente (Twente, the Netherlands), Wuhan University (Wuhan, Hubei, China), and the Ohio State University (Columbus, Ohio, USA). The dataset was specifically designed for semantic segmentation tasks in urban scenes captured by UAVs, which consisted of over 300 high-resolution remote sensing images of urban street scenes with the pixels of 4,096 × 2,160, capturing various urban environment features such as buildings, roads, vehicles, pedestrians, and vegetation in different settings, including residential, commercial, industrial areas, and parks. The dataset was annotated with eight primary semantic categories with each associated with a distinct color label to clearly differentiate the various objects in the UAV images, which included black for clutter, red for building, purple for road, magenta for static car, green for tree, olive for vegetation, brown for human, and blue for moving car. The color labels had been carefully processed to ensure high label quality. To create foggy scenarios, synthetic fog noise was applied to the UAVid dataset with scene depth serving as the main parameter. The process merged clear images and their corresponding depth maps, generating realistic urban scenes shrouded in fog. This method accurately simulated atmospheric effects by considering the distance between the camera and the scene elements. In fog removal research, optical models were commonly used to simulate the effect of fog on visual scenes. This work simulated foggy circumstances using the atmospheric scattering model described in equation (1). The exact calculation procedure was as follows.

$$t(x) = \exp(-\beta l(x)) \tag{9}$$

where $t(x)$ was the transmission quantity, which determined the scene brightness that reached the camera. In a homogeneous medium, the transmission quantity depended on the distance $l(x)$ from the scene elements to the camera. β was the attenuation coefficient, which could effectively control the concentration of fog. The larger the value, the heavier the fog generated. Meteorological Optical Range (MOR), often referred to as visibility, was a standard measure used to assess fog density. Typically, the camera-to-scene distance exceeded 0.05 meters. The visibility at this point was defined as MOR = 2.996/β as shown in equation (10). According to meteorological guidelines, visibility during foggy conditions was classified as less than 1 km, so the value range of the attenuation coefficient was determined as below.

$$\beta > 2.996 \times 10^{-3} \mathrm{m}^{-1} \tag{10}$$

The original images could be synthesized into three categories as light fog, moderate fog, and heavy fog with corresponding β values of 0.005, 0.015, and 0.030, respectively. Following this synthesis process, the dataset had expanded from the initial 300 images to a total of 900 images.

Experiment settings

The experimental setup used Ubuntu 22.04 and PyTorch with an NVIDIA RTX 4060 GPU and 16 GB RAM for training. NVIDIA CUDA 11.8 and cuDNN v7.6.1 were used for GPU acceleration. Python 3.8 was the software environment. During training, the Adam optimizer was utilized, specifically with a learning rate of 6×10^{-4}, a weight decay of 2×10^{-4}, a batch size of 4, and a total of 40 epochs implemented. Evaluation metric used Intersection over Union (IOU), a commonly utilized metric in semantic segmentation tasks, measuring the overlap between anticipated and ground truth masks. For each class, it was computed as the ratio of the intersection area to the union area of the predicted and target regions, defined as follows.

$$\mathrm{IOU} = \frac{I(X)}{U(X)} \tag{11}$$

where I(X) was the intersection set. U(X) was the union set and could be approximated as follows.

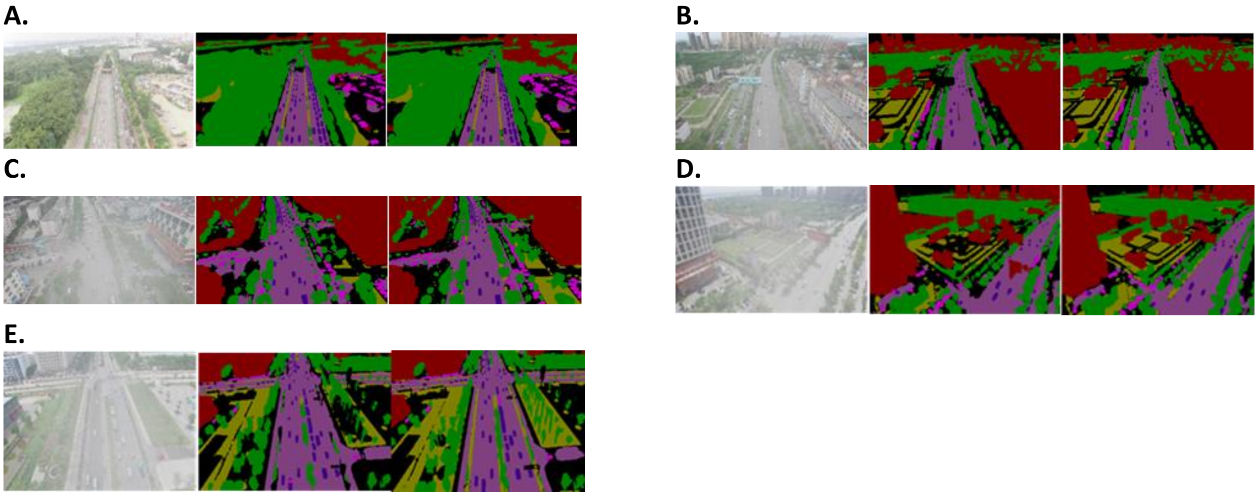

Figure 6. The comparison between mild (**A**), moderate (**B**), and heavy fog (**C-E**) effects.

Table 1. Comparison of segmentation accuracy between UNetFormer and proposed method.

Approach	Building	Tree	Clutter	Road	Plants	S-Car	M-Car	People	mIoU
UNetFormer	58.2	77.5	71.6	74.2	57.4	68.5	50.2	22.8	60.05
Proposed method	64.8	82.5	76.3	77.8	62.4	70.4	53.7	27.8	64.46

$$I(X) = \sum_{v \in V} X_v * Y_v \tag{12}$$

$$U(X) = \sum_{v \in V} (X_v + Y_v - X_v * Y_v) \tag{13}$$

When more than one class existed, the mean intersection over Union average IOU (mIOU) for all classes was calculated.

Results and discussion

This study evaluated the segmentation results of the original UnetFormer model and the proposed integrated model under mild, moderated, and dense fog environments (Figure 6). Each set of images included the original model, UnetFormer, and proposed method. Due to the influence of fog on image color and texture, the original UnetFormer performed poorly, exhibiting unclear boundaries and inaccurate class differentiation. In contrast, the proposed dehazing and super-resolution techniques significantly reduced the fog effects, resulting in more accurate scene classification and boundary segmentation.

The results demonstrated that the proposed method outperformed UNetFormer in segmentation accuracy across various categories. Under mild fog, both methods achieved high precision with minimal differences. However, under heavy fog, UNetFormer's accuracy declined, especially for small targets like vehicles and trees. In contrast, the proposed approach effectively reduced fog interference, achieving higher IoU values across categories such as buildings, trees, roads, vegetation, and vehicles (Table 1). Significant improvements were observed in trees and buildings, highlighting the proposed method's advantage in detail and boundary accuracy. In terms of mean IoU (mIoU), UNetFormer scored 60.05%, while proposed method reached 64.46%, marking a 4.41% improvement, demonstrating its effectiveness in foggy conditions. To further evaluate the efficacy of the proposed strategy, the mean segmentation accuracy of three additional models, Attentive bilateral contextual network (ABCNet) [23], Multi-Scale Discriminative (MSD) [24], and Bilateral Segmentation Network (BiSeNet) [25], were compared on the Foggy UAVid dataset. The MSD model achieved an average mIoU of 53.2%, which was mainly attributed to its limitations in multi-scale feature extraction and global context integration. While it employed dilated convolutions to mitigate edge contrast, the model struggled to fully capture long-range dependencies and complex contextual information in foggy conditions. ABCNet performed slightly better with an mIoU of 54.3%, indicating some improvement in its adaptability to foggy images. However, this enhancement remained limited as the model still failed to effectively capture global and long-range dependencies in challenging weather conditions. Moreover, ABCNet lacked specific processing or enhancement methods for handling blurred regions, further hindering its performance. The BiSeNet model with an mIoU of only 52.1% performed the worst among all the models. Despite being effective for fast segmentation tasks, its relatively simple network architecture proved insufficient for extracting deep semantic features in complex environments like foggy urban scenes, resulting in lower segmentation accuracy. In contrast, the UNetFormer model achieved an mIoU of 60.05%, significantly outperforming the other models. This superior performance could be attributed to the incorporation of transformer modules, which enhanced its ability to capture long-range dependencies and global context, making it more effective for semantic segmentation under challenging weather conditions. The proposed algorithm achieved an mIoU of 64.46%, clearly demonstrating its effectiveness in enhancing image segmentation performance. By employing dehazing techniques, it significantly improved image clarity, contrast, and detail restoration. Additionally, the integration of super-resolution technology addressed potential detail loss during dehazing, refining image resolution, and enriching textures, which resulted in smoother edges and more intricate structures, leading to better segmentation outcomes.

The algorithm was built on a semantic segmentation network that combined dehazing, super-resolution enhancement, and the UNetFormer model, which together addressed the challenges of reduced segmentation accuracy and insufficient detail in dense fog areas of urban street scenes. The AOT-Net removed fog noise from the image followed by Real-ESRGAN super-resolution enhancement that corrected excessive contrast, blurred edges, and dark tones, restoring details. Finally, the UNetFormer network segmented the processed images, achieving precise results. This method showed a 4.4% improvement in mIoU accuracy on the foggy UAVid dataset compared to the original UNetFormer network. While the proposed method performed well on the foggy UAVid dataset, its effectiveness in real foggy urban scenes had not been fully tested due to the limited availability of real-world foggy scene datasets. As large-scale training on such datasets remained challenging, future work would focus on collecting labeled images from actual foggy urban environments and incorporating them into the training process to enhance the network's robustness. Additionally, more lightweight network architectures were planned to be explored to develop an efficient segmentation network tailored for foggy scenarios, ensuring its practicality in real-world applications.

Acknowledgements

This research was supported by the Funding Plan for Philosophy and Social Sciences Research Innovation Projects at Henan Agricultural University (Grant No. SKJJ2022B10).

References

1. Liu L, Zhang Z, Xiang Z, Guo J. 2024. Prediction and effectiveness evaluation of urban economic development on fine scale of streetscape. Acta Geogr Sin. 79(8):1978-1993.
2. Fu E, Mo X. 2024. Application of UAV realistic 3D modeling in intelligent surveying, mapping and planning. Geomatics & Spatial Information Technology. 47(S1):294-298.

3. Zhong K, Feng Z. 2022. Overview of visual image information enhancement techniques for adverse weather of intelligent vehicles. Agric Equip Veh Eng. 60(11):40-43.
4. Lyu D, Yang Y. 2022. Summary of vehicle foggy environment perception research based on machine vision. Automation & Instrumentation. 2022(11):1-6.
5. Long J, Shelhamer E, Darrell T. 2015. Fully convolutional networks for semantic segmentation. Proc Conf Comput Vis Pattern Recognit. 2015:3431-3440.
6. AnbuDevi KA, Suganthi K. 2022. Review of semantic segmentation of medical images using modified architectures of UNET. Diagnostics. 12(12):3064.
7. Badrinarayanan V, Kendall A, Cipolla R. 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell. 39(12):2481-2495.
8. Pan H, Chen X, Ren J, Zhou C. 2024. DeepLabv3++: Fabric defect detection model based on semantic segmentation. J Optoelectron Laser. 4:1.
9. He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. Proc Conf Comput Vis Pattern Recognit. 2016:770-778.
10. Dhal KG, Das A, Ray S, Gálvez J, Das S. 2021. Histogram equalization variants as optimization problems: A review. Arch Comput Methods Eng. 28:1471-1496.
11. Cai B, Xu X, Jia K, Qing C, Tao D. 2016. Dehazenet: An end-to-end system for single image haze removal. IEEE Trans Image Process. 25(11):5187-5198.
12. Zhang J, He F, Duan Y, Yang S. 2023. AIDEDNet: Anti-interference and detail enhancement dehazing network for real-world scenes. Front Comput Sci. 17(2):172703.,
13. Ullah H, Muhammad K, Irfan M, Anwar S, Sajjad M, Imran AS, *et al.* 2021. Light-DehazeNet: A novel lightweight CNN architecture for single image dehazing. IEEE Trans Image Process. 30:8968-8982.
14. Wang L, Li R, Zhang C, Fang S, Duan C, Meng X, *et al*. 2022. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. J Photogramm Remote Sens. 190:196-214.
15. Li B, Peng X, Wang Z, Xu J, Feng D. 2017. Aod-net: All-in-one dehazing network. Proc IEEE Int Conf Comput Vis. 2017:4770-4778.
16. Zhang J, He F, Duan Y, Yang S. 2023. AIDEDNet: Anti-interference and detail enhancement dehazing network for real-world scenes. Front Comput Sci. 17(2):172703.
17. Wang X, Xie L, Dong C, Shan Y. 2021. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. Proc IEEE Int Conf Comput Vis. 2021:1905-1914.
18. Ali M, Zakaria A, Ahmed A, Mohamed O, Ali S, Attia M, *et al*. 2024. Real-ESRGAN and SWIN-U-NET applications in enhancing NASA imagery for VR solar system models. In 2024 6th International Conference on Computing and Informatics (ICCI). 2024:229-236.
19. Zhang W, Feng J, Li Z, Sun Z, Jia K. 2021. Reconstruction for Cherenkov-excited luminescence scanned tomography based on Unet network. J Optoelectron Laser. 48(17):129-140.
20. Ghazouani F, Vera P, Ruan S. 2024. Efficient brain tumor segmentation using Swin transformer and enhanced local self-attention. Int J Comput Assist Radiol Surg. 19(2):273-281.
21. Chen Y, Li J, Hu X, Liu Y, Ma J, Xing C, *et al*. 2024. Instance segmentation from small dataset by a dual-layer semantics-based deep learning framework. Sci China Technol Sci. 67(9):2817-2833.
22. Lyu Y, Vosselman G, Xia G, Yilmaz A, Yang M. 2020. UAVid: A semantic segmentation dataset for UAV imagery. ISPRS J Photogramm Remote Sens. 165:108-119.
23. Li R, Zheng S, Zhang C, Duan C, Wang L, Atkinson PM. 2021. ABCNet: Attentive bilateral contextual network for efficient semantic segmentation of Fine-Resolution remotely sensed imagery. ISPRS J Photogramm Remote Sens. 181:84-98.
24. Zheng B, Liu Y, Zhu Y, Yu F, Jiang T, Yang D, *et al*. 2020. MSD-Net: Multi-scale discriminative network for COVID-19 lung infection segmentation on CT. IEEE Access. 8:185786-185795.
25. Yu C, Wang J, Peng C, Gao C, Yu G, Sang N. 2018. Bisenet: Bilateral segmentation network for real-time semantic segmentation. Proc Euro Conf Comput Vis. 2018:325-341.