RDC-YOLOv5: Improved Safety Helmet Detection in Adverse Weather

Dexu Yao[1], Aimin Li[1(✉)], Deqi Liu[2], and Mengfan Cheng[2]

[1] Key Laboratory of Computing Power Network and Information Security,
Ministry of Education, Shandong Computer Science Center (National Supercomputer
Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences),
Jinan, China
10431210601@stu.qlu.edu.cn, lam@qlu.edu.cn
[2] Shandong Provincial Key Laboratory of Computer Networks,
Shandong Fundamental Research Center for Computer Science, Jinan, China
{10431210382,10431210366}@stu.qlu.edu.cn

Abstract. Outdoor construction sites are frequently affected by fog and various adverse weather conditions, resulting in a decline in the quality of the captured images. This deterioration ultimately leads to a significant drop in the performance of helmet-wear detection systems at construction sites. To address this challenge, we proposed an improved YOLOv5 model. Firstly, we set up a restoration network to effectively restore hazy image quality and enhance intricate details. Secondly, we introduced a micro-scale detection layer, enabling more efficient capture of smaller objects and effectively mitigating the adverse weather-induced variations in object scale. Finally, we added a cross-layer connection to meticulously amplify the fine-grained features of objects within the network's shallower layers. Our comprehensive evaluation of the enhanced YOLOv5 model, conducted on two distinct datasets, yielded a final mean average precision (mAP) value of 95.1%. This substantial improvement effectively reduces both false detections and missed detections, enhances overall robustness, and significantly improves detection performance in challenging adverse weather conditions.

Keywords: Safety helmet detection · Adverse weather · YOLOv5

1 Introduction

Object detection plays a vital role in computer vision, and recent years have witnessed significant advancements driven by the rapid progress in deep learning techniques [1]. This technology has applications in many domains, including autonomous driving, where it helps vehicles perceive and intelligently navigate through traffic signs, pedestrians, and other vehicles [2]. Additionally, object detection is vital in the medical field, enabling automated analysis and diagnosis of lesions in medical images, thereby improving disease early detection rates [3]. Furthermore, its utility extends to image segmentation, drones [4], agriculture [5], and many other domains.

In recent years, Convolutional Neural Networks (CNN) have revolutionized object detection [6]. CNN-based detectors have become dominant and have single-stage and two-stage detection models. Single-stage detection, exemplified by the R-CNN (Region-based Convolutional Neural Networks) family, which includes R-CNN, Fast R-CNN [7], and Faster R-CNN [8], generates candidate object regions through selective search or the Region Proposal Network (RPN) [9], then employs CNNs for region classification and localization.

In addition to the traditional two-stage methods, single-stage detection methods have gained prominence for their superior ability to perform object detection. YOLO series is a widely recognized single-stage object detection model. YOLO transforms object detection into a regression problem by predicting an object's class, location, and confidence within a grid. Several iterations of the YOLO model have been introduced, including YOLOv1 [10], YOLOv2 [11], YOLOv3 [12], YOLOv4 [13], and YOLOv5 [14], among others. These iterations have realized substantial improvements through enhancements in network architectures, utilization of more powerful feature extractors, and implementation of more efficient training strategies. The YOLOv5 has garnered attention for its high accuracy and faster processing speed, achieving exceptional results across various object detection datasets and real-world applications.

Detecting wearing safety helmets at construction sites is essential for mitigating accident risks and fostering secure working environments. Waranusast et al. [15] employed object detection algorithms to accurately identify helmet usage among construction workers. Li et al. [16] used deep learning techniques to achieve real-time helmet detection at construction sites, providing immediate safety alerts. However, helmet detection in real-world outdoor construction scenarios differentiates it from other detection tasks. Outdoor construction sites are often affected by weather-related interference, such as foggy conditions or dusty and sandy weather, which can decrease image quality and detection performance.

This paper aims to improve safety helmet detection performance at construction sites for adverse weather. Our contributions can be summarized as follows:

- Set up A Restoration Network: This network enhances original image quality, improves visibility, and reduces the pressure for the detection network.
- Add A Micro-Scale Detection Layer(152×152): The micro-scale detection layer improves the ability of models to detect smaller objects.
- Add the Cross-Layer Connection: The Cross-Layer Connection to capture intricate object features, especially within the shallower network layers, facilitates finer and more context-rich information extraction.

These contributions address the challenges of adverse weather in helmet detection at construction sites. Our experimental results also validate the effectiveness of our improvements.

2   Related Work

2.1   Object Detection in Adverse Weather

Detecting helmets in adverse weather presents substantial challenges, primarily due to low image quality, lighting changes, and background interference. Adverse weather conditions, such as haze, reduce the contrast between objects and the background, leading to blurred edges and details of small objects. Unstable lighting conditions introduce further complexity, with factors like shadows and glare altering the appearance of small objects. Additionally, background interference, often characterized by blurriness and interfering objects, can confound the features of small objects and result in a higher likelihood of false and miss detection.

Researchers have proposed many methods, including multimodal sensor fusion [17], where multiple sensors like visible cameras, infrared thermal cameras [18], and LIDAR [19] work together to improve helmet detection accuracy in adverse weather. However, these methods may have detection accuracy and hardware requirements limitations.

Deep learning algorithms [20] and image enhancement techniques have gained prominence in recent research. These techniques involve preprocessing images to enhance quality by removing elements like raindrops, snowflakes, and fog. Training models on these improved images significantly enhance detection accuracy.

2.2   YOLOv5 Model

YOLOv5 stands out as a prominent model in object detection, offering four versions: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x.

In this paper, we use the YOLOv5s version mainly because of its lightweight network architecture, which better meets the real-time detection requirements of the device. YOLOv5 employs CSPNet (Cross-Stage Partial Network) as its infrastructure, building a streamlined design that makes it ideal for real-time object detection. CSPNet offers excellent feature extraction efficiency and few parameters, improving inference speed while ensuring accuracy. It is worth emphasizing that YOLOv5 excels in multi-scale object detection, effectively detecting objects with different feature scales. Furthermore, we implement an adaptive training strategy that can be dynamically adjusted based on object confidence levels. This method effectively tackles the class imbalance challenge in object detection, thereby improving the model's ability to detect smaller, more challenging samples.

3   Our Proposed Method

3.1   Set up a Restoration Network

In adverse weather conditions, the quality of captured images reduced pixel quality, diminished sharpness, blurred edges, loss of detail, and a significant presence of noise and interference, all of which adversely affect detection performance.

To address these challenges, we set up a restoration network. The primary objective of this network is to mitigate the adverse effects of noise and blurring present in these foggy images, revealing the underlying features and structural information akin to clear images. This transformation process effectively converts a blurred or distorted image into a more visually coherent and interpretable form. Post-restoration, the image is forwarded to the detection network for further processing. This method restores detailed feature information, providing valuable cues for subsequent detection tasks.
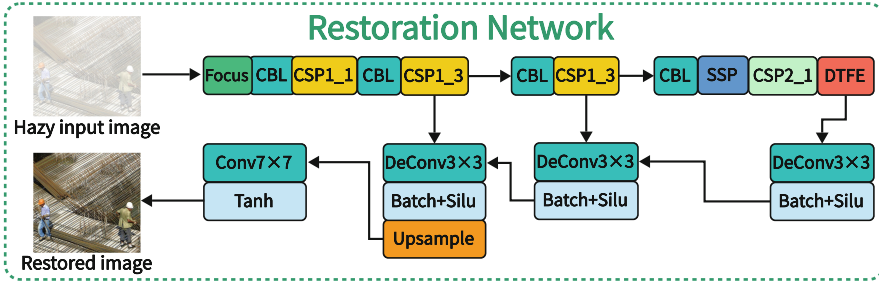

Fig. 1. The architecture of restoration network.

The architecture of the restoration network is shown in Fig. 1. The restoration network commences with feature extraction, and complex latent features inherent to the input image are captured. After extracting features, we introduce the Dynamic Transformer Feature Enhancement (DTFE) module to enhance feature extraction and representation capabilities. The DTFE utilizes three DeConv operations, one Upsample operation, and one Tanh activation function to restore clean image features. These operations generate a final clean image, subsequently input into the detection network for object detection.

Dynamic Transformer Feature Enhancement Module. The DTFE module comprises two deformable $conv3 \times 3$ layers and a vision transformer component, as shown in Fig. 2. These elements play critical roles in extending the adaptive shape-aware domain of the model.

The two deformable $conv3 \times 3$ layers perform dynamic feature transformations, facilitating adaptive shape-aware modelling. Deformable convolution introduces filters capable of dynamic shape and position adjustments, aligning with the feature distribution of the input data. This operation significantly enhances the network's feature transformation capabilities. By introducing this deformable $conv3 \times 3$, we effectively extend the receptive domain, accommodating adaptive shapes and reinforcing the model's feature extraction capabilities. This adaptability improves the precision in capturing object shape, size, and location information in adverse weather.
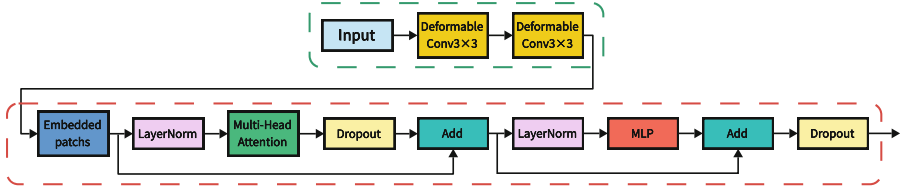
Fig. 2. The architecture of DTFE.

The vision transformer achieves a broader contextual understanding through global awareness and self-attention mechanisms. These mechanisms enhance feature representation capabilities, enabling the comprehensive capture of global features and relationships within the input image. This capability is particularly critical in adverse weather where atmospheric factors may lead to object blurriness or occlusion, necessitating a greater reliance on contextual information for accurate detection. Furthermore, the vision transformer improves the model's resilience to interference, making it better equipped to handle image disruptions caused by adverse weather conditions such as fog.

In summary, the restoration network enhances key and detailed features in the grayscale image, highlighting features that facilitate object detection. The clean image generated by the restoration network is then fed into the detection network to improve the detection accuracy further. This cooperation between the restoration and detection networks is essential for object detection in adverse weather.

3.2   Adding a 152×152 Detection Layer for Micro-Scale

In object detection networks, comprehending the capabilities of shallow and deep networks holds great significance. Shallow networks specialize in low-level feature extraction, characterized by compact receptive fields and high spatial resolutions. These attributes enable them to effectively capture fine details, including object contours and subtle colour variations, making them proficient at detecting smaller objects. However, as network depth increases, the capacity to capture detailed features related to smaller objects diminishes. Conversely, deep networks in higher layers possess broader receptive fields, rendering them more adept at detecting medium and large objects.

YOLOv5 can be divided into three detection layers: 19×19, 38×38, and 76×76. The smallest object scales these three layers can detect are large-scale with 32×32, medium-scale with 16×16, and small-scale with 8×8. Consequently, micro-scale objects with less than 8×8 pixels cannot be detected. However, safety helmets commonly found on construction sites are generally small objects, which can pose a significant detection challenge.
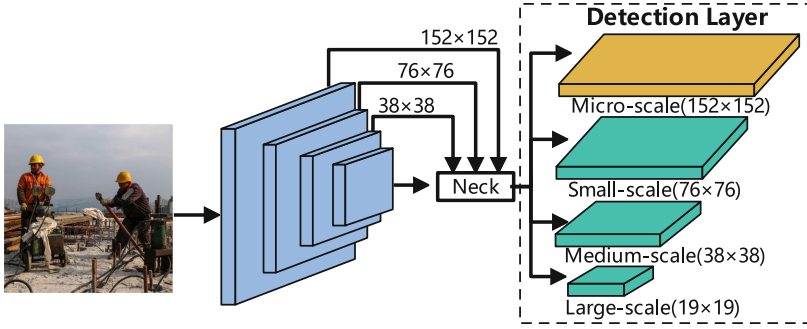
Fig. 3. Add a 152×152 detection layer for micro-scale.

To address this issue, we add a 152×152 detection layer for micro-scale objects, 608 divided by 152 is 4, as shown in Fig. 3. This adjustment effectively reduces the minimum detectable object size from 8×8 pixels to 4×4 pixels, significantly improving the performance of detecting smaller objects.

3.3   Adding a Cross-Layer Connection

In object detection within neural networks, as information travels from shallower to deeper layers, there is a gradual reduction in the scale of the feature map, leading to the blurring of feature and gradient information. This phenomenon gives rise to a critical issue known as "feature loss". The problem can result in misclassification of objects and the loss of crucial object information during detection.

To tackle this challenge, we add the cross-layer connection. This connectivity mechanism enables the fusion and sharing of information across distinct network layers, fostering a seamless integration of lower-level and upper-level features. This integration provides a more comprehensive understanding of an object's structural and contextual characteristics. Furthermore, incorporating cross-layer connections enhances model responsiveness by expediting model convergence and decision-making processes, which is particularly beneficial for real-time object detection in adverse weather.

The cross-layer connection structure is realized through the density block module, as illustrated in Fig. 4. The density block module plays a crucial role in facilitating the transfer of feature information. Its image features are directly transmitted to the feature fusion network, where they are spliced and fused. The splicing layer then conveys the fused feature information to the subsequent layer, augmenting the feature representation capability. Within this structure, the splicing layer is responsible for merging two feature maps from different layers, each with dimensions $K_1 \times H_1 \times W_1$ and $K_2 \times H_1 \times W_1$, respectively. The output feature information of the splicing layer can be expressed as:

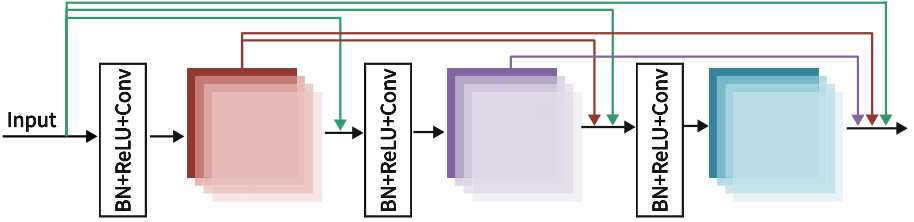$$Z_{concat} = (K_1 + K_2) \times H_1 \times W_2 \tag{1}$$

Fig. 4. The architecture of the density block module.

K_1 and K_2 represent the number of channels in the input feature maps from different layers, while H_1 and W_1 denote the height and width of the input feature maps required for the splicing layer. Equation 1 demonstrates that combining features from different layers in the splicing layer enhances subsequent network layers' feature representation. The cross-layer connection then transmits the new feature information $K_3 \times H_1 \times W_1$ to the corresponding splicing layer, resulting in the following feature information:

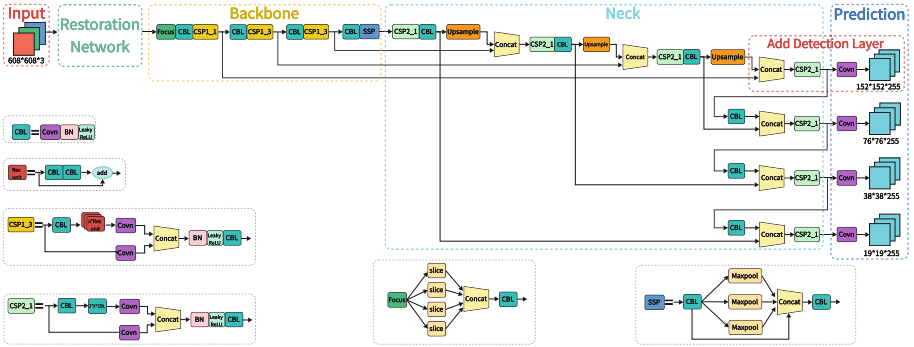$$Z_{concat} = (K_1 + K_2 + K_3) \times H_1 \times W_2 \tag{2}$$

Fig. 5. The structure of the improved YOLOv5 model.

The height and width of the output feature maps of the fifth and seventh layers in YOLOv5's feature extraction network match the input scales of the two splicing layers. The addition of cross-layer connections in these two layers combines feature information extracted from shallower layers with that from deeper layers, enriching feature information for small and medium-sized objects. This improvement allows the network to better capture features of objects at different scales in subsequent stages, thereby improving the accuracy and robustness of the model. The structure of the final improved YOLOv5 model is shown in Fig. 5.

4 Experimental Results and Analysis

4.1 Dataset and Environment Construction

No public datasets can accurately represent construction scenarios under foggy conditions, so we established a Synthetic Fog Dataset that simulates construction sites under foggy conditions, offering a resource for evaluation. This dataset comprises 6,000 images and is obtained from natural construction scenes. In addition, to test our model more realistically, we also used the real-world Foggy Driving Dataset [21] for comparison tests with other models. The synthetic fog dataset applied an atmospheric scattering model to replicate foggy conditions to generate synthetic foggy images, as shown in Fig. 6.

(a)Clean images

(b)Synthetic foggy images

Fig. 6. Example images in the proposed fog dataset.

Specifically, atmospheric scattering models are employed to simulate the scattering and absorption of light in fog. The following equation quantifies the scattering effect:

$$I(x) = J(x) \cdot t(x) + A \cdot (1 - t(x)) \tag{3}$$

$I(x)$ represents the image under foggy conditions, $J(x)$ is the original image, A stands for the estimated atmospheric light, and $t(x)$ denotes the transmittance,

which dictates the extent of light attenuation within the fog. $t(x)$ is calculated using the following equation:

$$t(x) = \exp(-\beta \cdot d(x)) \tag{4}$$

β represents the fog concentration coefficient, indicating the fog's concentration level. $d(x)$ signifies the distance between the camera and the object, defined as:

$$d(x) = \sqrt{(x_2 - x_1,)^2 + (y_2 - y_1)^2} \tag{5}$$

(x_1, y_1) and (x_2, y_2) denote the coordinates of two pixel points. Our experiments standardized the global atmospheric light parameter A to 0.6. We introduced variability in the atmospheric scattering parameter β to manipulate the fog level, selecting random values from 0.08 to 0.12.

Our synthetic fog dataset was divided into two subsets: 5,000 images for training and 1,000 images for thorough testing. Additionally, we employed a systematic approach by categorizing the dataset into four scales based on object pixel size: micro-objects, small-objects, medium-objects, and large-objects. This segmentation enabled a comprehensive assessment of the model's detection performance across a spectrum of object sizes, ensuring a thorough evaluation of its capabilities.

Our improved YOLOv5 model was built upon the PyTorch 1.8.1 framework, and both training and testing phases were executed on NVIDIA A100 GPUs. We adhered to the hyperparameter settings defined in the YOLOv5 model architecture throughout the training process. To combat overfitting, we applied a weight decay coefficient of 0.0005, and to prevent the model from converging to local optima, we set the momentum value at 0.937. The initial learning rate was 0.01, with termination occurring at 0.2. After 200 training iterations, the model attained optimal performance with the specified weight configuration.

4.2 Evaluation Criteria

This paper employs widely accepted evaluation criteria in object detection, mean Average Precision (mAP). The mathematical expressions for mAP are detailed as follows:

$$P = Precision = \frac{TP}{TP + FP} \tag{6}$$

$$R = Recall = \frac{TP}{TP + FN} \tag{7}$$

$$AP = \int_0^1 P(R)dR \tag{8}$$

$$mAP = \frac{\sum_{i=0}^n AP(i)}{n} \tag{9}$$

P is precision, TP is true positives, FP is false positives, and FN is false negatives. AP represents the average accuracy for a specific category, where n is the total number of categories, and mAP is the overall average value of AP across all categories.

4.3   Ablation Studies

To test the impact of each component in the improved model, we conducted ablation experiments to analyze different components, including the restoration network, the detection layer(152×152), and the cross-layer connections. We used YOLOv5 as the baseline model and added the different components to the baseline model step by step as follows:

1. **F1: Restoration Network**
2. **F2: Detection Layer(152×152)**
3. **F3: Cross-Layer Connection**

Table 1. The Results of Ablation Experiments

Variants	Micro	Small	Medium	Large	mAP(%)
YOLOv5	92.6	92.9	93.2	91.7	92.6
YOLOv5+F1	93.1	93.2	93.5	91.8	92.9
YOLOv5+F2	92.8	92.9	93.5	92.2	92.9
YOLOv5+F3	93.0	93.1	93.7	92.6	93.1
YOLOv5+F1+F2	93.5	93.6	93.6	93.1	93.5
YOLOv5+F1+F3	92.8	93.6	94.0	93.1	93.4
YOLOv5+F2+F3	93.2	93.6	94.5	93.2	93.6
YOLOv5+F1+F2+F3	**93.8**	**94.6**	**96.0**	**95.2**	**95.1(↑2.5)**

The experimental results are shown in Table 1, that each component of the improved YOLOv5 model helps to improve the object detection performance. The data shows that the restoration network and 152×152 Detection Layer are more helpful in detecting small objects, and the cross-layer connection is more inclined to detect large objects' performance improvement. Through this experiment, we verified the effectiveness of these improved methods. Our improved YOLOv5 model improved the performance of helmet detection in adverse weather by 2.5% compared to the YOLOv5 model.

### 4.4   Comparison Experiments

Comparison Experiments of Synthetic Fog Dataset. In order to more fully test the performance of the improved YOLOv5 model, we compared it with the state-of-the-art model. The results of the comparison experiments are shown in Table 2.

The table results show that the improved YOLOv5 model is much better than the other models in terms of mAP and achieves the best detection results. What is more noteworthy is that the results of our improved YOLOv5 model in the scale of "Micro", compared with the detection results of other scales,

Table 2. Comparative Experimental Results Of Synthetic Fog Dataset

Models	Micro	Small	Medium	Large	mAP(%)
YOLOv5	92.6	92.9	93.2	91.7	92.6
DENet [22]	92.8	93.3	93.5	93.6	93.3
DS-Net [23]	93.1	94.0	93.9	94.2	93.8
CF-YOLO [24]	93.4	93.9	94.1	94.2	93.9
R-YOLO [25]	93.2	93.9	93.5	93.8	93.7
FFA-YOLOv5 [26]	93.2	94.4	94.6	93.8	94.0
RDC-YOLOv5(Ours)	**93.8**	**94.6**	**96.0**	**95.2**	**95.1**

achieved a more obvious advantage, further indicating that our improvement in the detection of small objects is a significant effect, more conducive to the detection of small helmet objects in adverse weather. In addition, our improved model frames per second (FPS) reach 32, which can meet the real-time inspection requirements at construction sites.
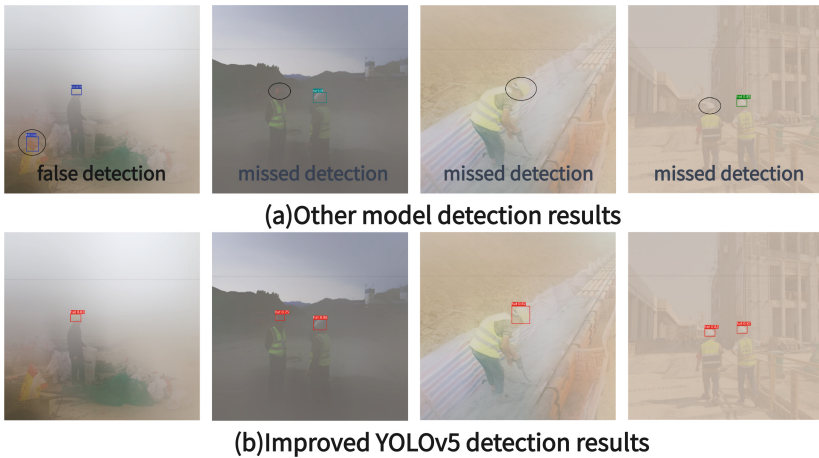
Fig. 7. Comparison of the Detection Results of the Improved YOLOv5 Model with Other Models.

We show the detection results of the improved YOLOv5 model with other models, as shown in Fig. 7. Our improved model effectively mitigates the leakage and false detection problems of other models compared to other models in the case of dense foggy, and the improved model has better robustness.

Comparison Experiments of Foggy Driving Dataset. We conducted comparative experiments with other models on a real-world foggy driving dataset.

Our goal was to demonstrate the effectiveness of our improved YOLOv5 model. The results are shown in Table 3.

Table 3. Comparative Experimental Results Of Foggy Driving Dataset

Models	Person	Motorbike	Bicycle	Car	Bus	mAP(%)
YOLOv5	89.8	89.9	90.5	91.4	91.9	90.7
DENet [22]	90.5	90.6	91.3	91.1	92.6	91.2
DS-Net [23]	90.6	91.0	91.4	92.4	92.9	91.7
CF-YOLO [24]	90.4	90.9	91.1	91.2	90.9	90.3
R-YOLO [25]	91.0	91.2	91.7	92.6	92.9	91.9
FFA-YOLOv5 [26]	91.2	91.4	91.8	92.5	93.2	92.0
RDC-YOLOv5(Ours)	**92.1**	**92.0**	**92.3**	**93.2**	**93.6**	**92.6**

According to the experimental data, our improved YOLOv5 model performs best on the foggy driving dataset, and its mAP value is significantly higher than the other models. Further analysis reveals that our improved model significantly improves the detection of smaller objects such as "Person" and "Motorbike". These findings confirm that our improved method is more favourable for detecting small objects and can improve the detection effect in adverse weather.

5 Conclusions

This paper proposed an improved YOLOv5 model to enhance object detection performance at outdoor construction sites in adverse weather. Our methods include the addition of a restoration network to enhance the quality of input images, the incorporation of a 152×152 detection layer for micro-scale to mitigate the challenges posed by varying object sizes, and the addition of a cross-layer connection module to enrich fine-grained object features within the network's shallow layers. We conducted experiments on two different datasets, and the result showed that our method significantly improves helmet detection in inclement weather while effectively reducing false and missed detections.

References

1. Xu, J.: Deep learning for object detection: a comprehensive review. Towards Data Science (2017)
2. Niranjan, D., VinayKarthik, B.: Deep learning based object detection model for autonomous driving research using Carla simulator. In: 2nd International Conference on Smart Electronics and Communication (ICOSEC), pp. 1251–1258. IEEE (2021)

3. Wang, J., Zhu, H., Wang, S.-H., Zhang, Y.-D.: A review of deep learning on medical image analysis. Mobile Netw. Appl. **26**, 351–380 (2021)

4. Xiao, Y., et al.: A review of object detection based on deep learning. Multimedia Tools Appl. **79**, 23-729–23-791 (2020)

5. Hasan, A.M., Sohel, F., Diepeveen, D., Laga, H., Jones, M.G.: A survey of deep learning techniques for weed detection from images. Comput. Electron. Agric. **184**, 106067 (2021)

6. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)

7. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)

8. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)

9. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, vol. 28 (2015)

10. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)

11. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271 (2017)

12. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. arXiv preprint arXiv:1804.02767 (2018)

13. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: YOLOv4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)

14. Huang, J., et al.: Speed/accuracy trade-offs for modern convolutional object detectors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7310–7311 (2017)

15. Waranusast, R., Bundon, N., Timtong, V., Tangnoi, C., Pattanathaburt, P.: Machine vision techniques for motorcycle safety helmet detection. In: 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013), vol. 2013, pp. 35–40. IEEE (2013)

16. Li, Y., Wei, H., Han, Z., Huang, J., Wang, W.: Deep learning-based safety helmet detection in engineering management based on convolutional neural networks. Adv. Civil Eng. **2020**, 1–10 (2020)

17. Zhang, C., et al.: Robust-FusionNet: deep multimodal sensor fusion for 3-D object detection under severe weather conditions. IEEE Trans. Instrum. Meas. **71**, 1–13 (2022)

18. Bhadoriya, A.S., Vegamoor, V., Rathinam, S.: Vehicle detection and tracking using thermal cameras in adverse visibility conditions. Sensors **22**(12), 4567 (2022)

19. Kim, T.-L., Arshad, S., Park, T.-H.: Adaptive feature attention module for robust visual-lidar fusion-based object detection in adverse weather conditions. Remote Sens. **15**(16), 3992 (2023)

20. Sharma, T., Debaque, B., Duclos, N., Chehri, A., Kinder, B., Fortier, P.: Deep learning-based object detection and scene perception under bad weather conditions. Electronics **11**(4), 563 (2022)

21. Sakaridis, C., Dai, D., Van Gool, L.: Semantic foggy scene understanding with synthetic data. Int. J. Comput. Vis. **126**, 973–992 (2018)

22. Qin, Q., Chang, K., Huang, M., Li, G.: DENet: detection-driven enhancement network for object detection under adverse weather conditions. In: Proceedings of the Asian Conference on Computer Vision, pp. 2813–2829 (2022)
23. Huang, S.-C., Le, T.-H., Jaw, D.-W.: DSNet: joint semantic learning for object detection in inclement weather conditions. IEEE Trans. Pattern Anal. Mach. Intell. **43**(8), 2623–2633 (2020)
24. Ding, Q., et al.: CF-YOLO: cross fusion yolo for object detection in adverse weather with a high-quality real snow dataset. IEEE Trans. Intell. Transp. Syst. (2023)
25. Wang, L., Qin, H., Zhou, X., Lu, X., Zhang, F.: R-YOLO: a robust object detector in adverse weather. IEEE Trans. Instrum. Meas. **72**, 1–11 (2022)
26. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: FFA-Net: feature fusion attention network for single image dehazing. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, pp. 11-908–11-915 (2020)