UPDN: Pedestrian Detection Network for Unmanned Aerial Vehicle Perspective

Minghua Jiang[1,2], Yulin Wang[1], Mengsi Guo[1], Li Liu[1], and Feng Yu[1,2](✉)

[1] School of Computer Science and Artificial Intelligence,
Wuhan Textile University, Wuhan 430200, China
[2] Engineering Research Center of Hubei Province for Clothing Information,
Wuhan 430200, China
yufeng@wtu.edu.cn

Abstract. Pedestrian detection for Unmanned Aerial Vehicle (UAV) perspective has significant potential in the fields of computer vision and intelligent systems. However, current methods have some limitations in terms of accuracy and real-time detection of small targets, which severely affects their practical application. To address these challenges, we propose UPDN, a novel network designed to improve detection comprehensive performance while maintaining high speed as much as possible. To achieve this objective, UPDN incorporates two key modules: the Spatial Pyramid Convolution and Pooling Module (SPCPM) and the Efficient Attention Module (EAM). The SPCPM effectively captures multi-scale features from pedestrian regions, enabling better detection of small targets. The EAM optimizes network operations by selectively focusing on informative regions, enhancing the overall efficiency of the detection process. Experimental results on the constructed dataset demonstrate that UPDN outperforms other classic detection methods. It achieves state-of-the-art results in terms of both Average Precision (AP) and F1 score, achieving a detection speed of 107.37 frames per second (FPS). In summary, UPDN provides an efficient and reliable solution for pedestrian detection from a UAV perspective, offering a feasible approach for real-world applications.

Keywords: Pedestrian detection · UAV · Small target

1 Introduction

Unmanned aerial vehicle (UAV) target detection technology has been widely applied in various practical scenarios, such as plant protection [14], wildlife conservation [7], personnel search and rescue [17], and urban monitoring [1]. Compared to common fixed cameras on the streets, UAVs have stronger mobility and can perform detection tasks in a wider range of scenarios.

Among them, pedestrian detection technology is a key part used in personnel search and rescue, and it also plays a crucial role in fields such as video monitoring [23], and autonomous driving [2]. Although there are some related

B. Sheng et al. (Eds.): CGI 2023, LNCS 14497, pp. 27–39, 2024.
https://doi.org/10.1007/978-3-031-50075-6_3

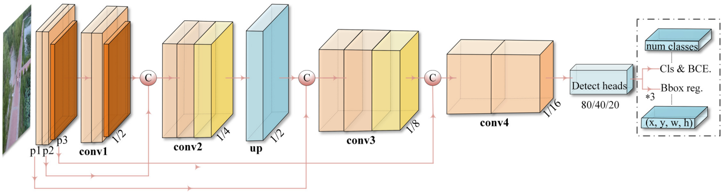

Fig. 1. The overall framework of UPDN is a pedestrian detection algorithm that utilizes a unique perspective from an UAV. The algorithm consists of several key components, including the SPCPM, the EAM, and the SIOU loss function. This diagram is intended to describe the main flow, where up represents the upsample operation and conv is not limited to conv2d operations, ∗3 because there are three detection heads.

studies on pedestrian detection [5,6,9,16], existing methods still have limitations and room for improvement in terms of terminal deployment, robustness in complex scenarios, and real-time performance. From the high-altitude perspective of UAVs, there may be relatively less occlusions in crowds, but this brings problems such as complex backgrounds, small targets, lost key information, or excessively large algorithm models, which hinder the deployment and application of pedestrian detection systems in real-world scenarios. To address these limitations, this paper proposes UPDN (Pedestrian Detection for Unmanned Aerial Vehicle Perspective), see Fig. 1, a novel network designed specifically for efficient and accurate pedestrian detection for the UAV perspective. The goal of UPDN is to improve detection speed while maintaining high accuracy, achieving real-time pedestrian detection in UAV images.

UPDN's crucial component, the Spatial Pyramid Convolution and Pooling Module (SPCPM), is designed to improve the representation of small targets by capturing multi-scale information and effectively integrating it into feature maps. Another important component of UPDN is the Enhanced Attention Module (EAM), which enhances the model's representation and discriminative capabilities by introducing spatial attention. The SIOU can help the model to carry out target detection more accurately through four loss constraints.

In the experimental session, we evaluated the multidimensional performance of UPDN on a constructed dataset using 9 metrics and compared it with seven classical algorithms. The results demonstrate that UPDN outperforms several classical algorithms, achieving state-of-the-art performance in terms of average accuracy (AP), F1 score, and recall. Furthermore, UPDN achieves a remarkable detection speed of 107.37 frames per second (FPS), making it suitable for real-time applications. UPDN addresses the unique challenges of detecting pedestrians from the UAV perspective, and the potential for various applications in surveillance, security, and autonomous systems. In summary, the four contributions of this study are:

- UPDN algorithm shows excellent comprehensive performance in pedestrian detection, especially in the UAV scenario.

- The SPCPM boost the representation and detection of small targets. The EAM enhance the model's representation and discriminative capabilities.
- The combination of SPCPM, EAM, and SIOU helps to the aggregate performance and efficiency of the UPDN algorithm, enhancing the capability of feature representation, small target detection, complex background detection, and target location.
- The dataset we built provides a broad benchmark for pedestrian detection and provides strong support for the evaluation of algorithms, while also providing a useful resource for future research.

The rest of this paper is organized as follows. Section 2 introduces related work on pedestrian detection for UAV images. Section 3 outlines the proposed UPDN framework. Section 4 discusses the experimental settings, dataset description, evaluation metrics, and effect analysis. Finally, Sect. 5 summarizes the entire paper and proposes future research directions.

2 Related Works

2.1 Pedestrian Detection Algorithms

Pedestrian detection is a crucial task in the field of computer vision. Traditional pedestrian detection methods relied on handcrafted features and machine learning classifiers [20,24], such as Haar features and Adaboost algorithm. However, these methods have limited performance when dealing with complex scenes.

Nevertheless, pedestrian detection algorithms based on Convolutional Neural Networks (CNNs) have made significant progress. Classic CNN-based pedestrian detection algorithms, such as Faster R-CNN, SSD, and YOLO [3,8,13,25], have achieved more accurate and efficient detection. However, these algorithms still face many challenges.

2.2 UAV Perspective Pedestrian Detection

In UAV perspective pedestrian detection, traditional methods struggle to meet the practical requirements due to the varying camera viewpoints and reduced target sizes. Consequently, researchers have proposed a range of pedestrian detection methods [11,18] specifically tailored for the UAV perspective.

One common approach is to enhance pedestrian detection performance through multi-scale feature fusion. These methods leverage feature pyramid networks or multi-scale convolution and pooling operations to capture target information at different scales [19,22], thereby improving the model's perception capabilities. Another approach is the incorporation of attention mechanisms to enhance the focus on pedestrians [10,21]. By introducing attention modules within the network, models can automatically learn which features are crucial for pedestrian detection, thereby boosting detection performance.

Additionally, researchers have explored techniques such as data augmentation [15], and model optimization [12] to enhance the performance and generalization capabilities. However, despite the progress made, UAV vision pedestrian detection still faces challenges in accurately detecting small-scale pedestrians, handling complex scenarios and lost target key information with cluttered backgrounds and occlusions.

3 Approach Overview

3.1 SPCPM

The SPCPM enhances the representation of small targets in UAV-based pedestrian detection scenarios, as shown in Fig. 2. It leverages spatial pyramid convolution and pooling operations to effectively capture multi-scale information. SPCPM improves the network's ability to handle objects of different sizes and provides comprehensive feature information. Experimental results demonstrate its effectiveness, with improvements observed in evaluation metrics such as AP, F1 score, precision, and recall.

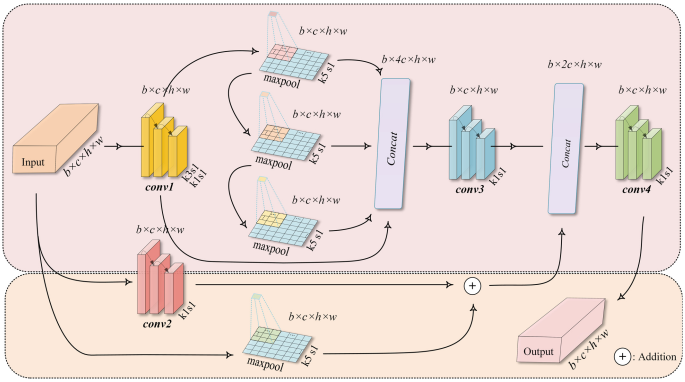

Fig. 2. The SPCPM is a crucial component of the pedestrian detection algorithm, particularly challenging UAV-based scenarios with small targets. It is designed to enhance the representation and detection of small targets by effectively capturing multi-scale information through spatial pyramid convolution and pooling operations. The SPCPM effectively captures multi-scale information and emphasizes important features in the input feature maps, enhancing the performance of the pedestrian detection algorithm. Its purpose is to enhance the extracted features from the input image and improve the model's performance in detecting pedestrians in UAV-based images.

The operations performed by SPCPM preserve the feature information while increasing the receptive field of the network. By capturing multi-scale information through spatial pyramid convolution and pooling processing, SPCPM enhances the network's ability to handle objects of different sizes. The module effectively provides more comprehensive feature information, particularly for small targets, thereby improving the network's detection capabilities in UAV-based pedestrian detection scenarios.

Experimental results demonstrate the usefulness and effectiveness of the SPCPM. The performance evaluation indicates a significant improvement in some evaluation metrics. The AP shows a notable rise of 3.81%, while the F1 score boosts from 0.19 to 0.22 and recall increases from 10.37% to 12.78%. These improvements validate the module's effectiveness in enhancing the model's detection performance. The module maintains operational efficiency, with a slight decrease in FPS and reasonable increases in parameter count and computational complexity. The benefits in detection performance justify the module's inclusion in the pedestrian detection methodology. Furthermore, the SPCPM contributes to the operational efficiency of the network. The FPS measure shows a slight decrease from 119.17 to 117.50, indicating that the module does not significantly affect the network's processing speed. Although the module introduces additional parameters and computational complexity, with the parameter count growing from 7.022M to 13.450M and the GLOPS from 15.946G to 21.093G, the resulting boost in model size and computational load is reasonable considering the notable improvements in detection performance.

3.2 EAM

In the field of object detection, the attention mechanism is often introduced into the deep learning network structure. The EAM as shown in Fig. 3 introduced in this study builds upon the foundation of channel attention mechanisms while incorporating spatial features to enhance the performance of the network.

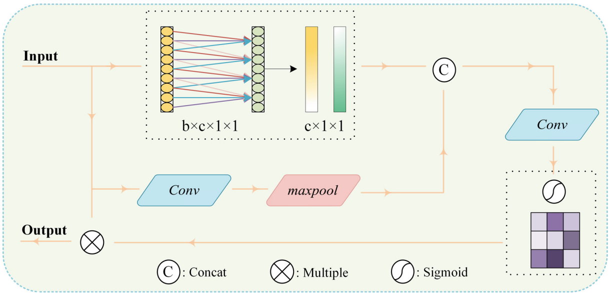

Fig. 3. The EAM is another key component of the pedestrian detection algorithm proposed in the research paper. The module introduces additional spatial attention, which is emphasizing important features and suppressing irrelevant ones, the EAM improves the comprehensive performance of the model in detecting pedestrians from UAV-based images.

The EAM is designed to selectively capture both channel-wise dependencies and spatial information to improve the overall representation and discriminative

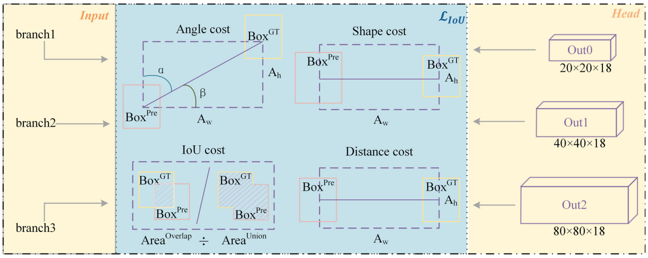

Fig. 4. The SIOU consists of four key aspects: center coordinate loss, width and height loss, confidence loss, and class loss.

power of the network. The EAM begins with a 1D convolutional layer that operates on the input feature map. This operation enables the module to capture channel-wise dependencies by modeling interactions between different channels. The resulting intermediate feature representations provide an initial understanding of the input data. To further enhance the module's understanding of spatial information, we introduce an additional pathway that involves a convolutional layer followed by max pooling. This parallel spatial pathway focuses on extracting relevant spatial features and provides valuable spatial context to the module.

To combine the channel and spatial information effectively, the outputs of the channel-wise and spatial pathways are concatenated along the channel dimension. This fusion step allows the learned channel-wise dependencies to be integrated with the captured spatial features. The concatenated feature map is then passed through a subsequent convolutional layer, enabling the module to adaptively transform its dimensionality back to the original input size. To generate attention weights, we apply a sigmoid activation function to the output of the convolutional layer. The sigmoid function scales the values to the range [0, 1], indicating the importance or relevance of each element in the feature map. Finally, the attention weights are multiplied element-wise with the original input feature map, resulting in the final output of the EAM. This operation allows the module to selectively emphasize informative regions while attenuating less important or noisy regions. SPCPM is located at the end of the backbone network, and EAM is applied in four places, after SPCPM, after conv1, after up, and after conv4.

The enhanced feature map captures both channel-wise dependencies and spatial attention, leading to improved discriminative capabilities of the network. Experimental results on the constructed data demonstrate the effectiveness of the EAM in pedestrian detection tasks. AP is increased by 1.15%, and F1 value is increased from 0.19 to 0.21. Although the accuracy index has decreased, the two core comprehensive indicators have been significantly improved, so EAM is recognized as effective.

3.3 Loss Function

The loss function used in UPDN is the standard Single Intersection over Union (SIOU) loss, as shown in Fig. 4. It is defined as the negative logarithm of the

Table 1. The outcome of the pedestrian detection comparison experiment.

Model	Shape	AP (%)	Precision (%)	Recall (%)	F1	FPS	Params (M)	GLOPS(G)	Param size (MB)	Total size (MB)
Retinanet	600 × 600	3.54	92.05	1.34	0.03	41.29	36.330	145.339	\	\
SSD	300 × 300	4.33	52.95	1.65	0.03	89.40	23.612	60.597	90.07	503.67
Faster RCNN	600 × 600	4.35	39.11	5.68	0.10	17.08	136.689	369.719	521.43	2096.31
Yolov3	416 × 416	11.23	45.99	12.05	0.19	106.02	615.238	65.597	234.59	12228.39
Centernet	512 × 512	16.40	67.78	10.96	0.19	103.84	32.665	70.217	124.61	1686.61
Yolov5s	640 × 640	21.69	90.77	10.37	0.19	119.17	7.022	15.946	26.79	830.29
Yolov7tiny	640 × 640	21.05	82.35	11.97	0.21	118.47	6.014	13.181	\	\
Ours	640 × 640	**26.85**	88.41	13.88	**0.24**	107.37	14.239	21.306	54.32	905.51

IOU between the predicted bounding box (bbox) and the ground truth bbox. Although we did not make changes to the SIOU loss function, we included it in our paper to give readers an understanding of the fundamental components of our object detection model.

4 Experiments and Results

4.1 Experimental Setup

The experimental platform is configured as follows: The system version used is Windows 10 Pro 64-bit, the processor is an Intel Core i9-12900KF 12th Gen running at 3.2 GHz, and the graphics processing unit is an NVIDIA GeForce RTX 3090 Ti 24 GB. The system utilizes Python 3.8.15 along with the following libraries: numpy 1.24.1, opencv-python 4.6.0.66, torchvision 0.13.0+cu116 and torch 1.12.0+cu116.

4.2 Experimental Data

In this chapter, we conduct experiments on a pedestrian detection dataset that we have constructed. The data comprises 16,953 samples gathered from various sources, including subsets of pedestrian samples from the VisDrone dataset [4], subsets of pedestrian samples from the Tinyperson dataset, and pedestrian samples collected by us. The data we collect is obtained using the DJI Magic3. These samples are incorporated to enhance the quality of the dataset and support the robustness of the subsequent algorithm model. By including a diverse range of perspectives, our data better reflects real-world scenarios and contributes to the improved performance and generalizability of the algorithm.

Table 2. The outcome of the pedestrian detection ablation experiment.

Model	Shape	AP (%)	Precision (%)	Recall (%)	F1	FPS	Params (M)	GLOPS (G)	Param size (MB)	Total size (MB)
Base	640 × 640	21.69	90.77	10.37	0.19	119.17	7.022	15.946	26.79	830.29
EAM+SIOU	640 × 640	22.84	85.53	12.07	0.21	100.86	7.451	15.795	28.42	851.52
SPCPM+SIOU	640 × 640	25.50	89.62	12.78	0.22	117.50	13.450	21.093	51.31	893.09
E+S+SIOU (Ours)	640 × 640	**26.85**	88.41	13.88	**0.24**	107.37	14.239	21.306	54.32	905.51

4.3 Evaluation Metrics

To comprehensively assess the detection performance of the proposed model, we chose AP and F1 as the primary evaluation metrics for our models. In addition, there are several other metrics commonly used for evaluating the performance and efficiency of deep learning models, including Precision, Recall, FPS, Params, GFLOPS, Params size, and Total Size.

4.4 Experimental Analysis

In this section, we provide a detailed analysis of our experimental results. We conducted experiments on our constructed dataset and report the results in terms of AP, F1, FPS, Precision, Recall, Params, GLOPS, Param size, and Total size. Instead of pursuing high accuracy through stacking a large number of modules, we aim to strike a balance between comprehensive performance (AP and F1) and speed on a concise structure, in order to prepare for possible terminal deployment, reduce reliance on high-cost hardware, and save costs while ensuring property.

In Fig. 5, it showcasing the network's ability to accurately detect pedestrians in challenging conditions. The images are annotated with the predicted pedestrian bounding boxes, corresponding class labels, and confidence scores. The visual evidence demonstrates the effectiveness and robustness of the UPDN network in handling diverse real-world scenarios, validating its suitability for pedestrian detection tasks across different conditions in UAV vision applications.

Therefore, we compare the performance of UPDN with different algorithms of their small version in our experiments. The results show that UPDN achieves excellent performance, with the best AP and F1 scores of 26.85% and 0.24, respectively. The FPS also meets real-time requirements, exceeding 25 with a value of 107.37. Other parameters, including Precision at 88.41%, Recall at 13.88%, Params at 14.239M, GFLOPS at 21.306G, Param size at 54.32 MB, and Total size at 905.51 MB, are also reported.

More specifically, we compare the performance of our proposed algorithm with seven other classic algorithms, including Retinanet, SSD, Faster RCNN, Yolov3, Centernet, Yolov5s, and Yolov7tiny in Table 1. In terms of AP and F1, our proposed algorithm has at least improved by 5.16% and 0.03 compared with other algorithms. Although our model may not reach the highest in other indicators, it demonstrated strong comprehensive performance in balancing Precision and Recall, as evidenced by the AP and F1 scores.

Next, we conduct ablation experiments on the UPDN algorithm to evaluate the impact of each component on the overall performance in Table 2. Specifically, when SPCPM is used, the AP and F1 scores of UPDN improve by 3.81% and 0.03, respectively, while when EAM is used, the AP and F1 scores contribute by 1.15% and 0.02, respectively. This effectively demonstrates the effectiveness of the SPCPM and EAM. The results show that each component contributes significantly to the overall performance, with the SPCPM contributing the most.

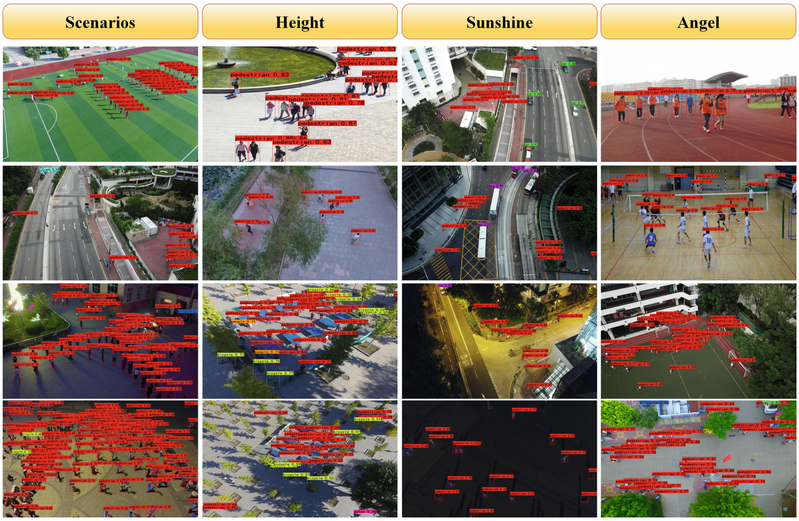

Fig. 5. It illustrates the pedestrian detection performance of the UPDN network in various scenarios, including different scenes, altitudes, lighting conditions, and camera angles. The figure is divided into four sections, each displaying four representative images.

We also test the case of applying both the SPCPM and EAM to UPDN. The results show that when these two modules work together, they achieve a synergistic effect greater than the sum of their individual effects, which is very exciting. Meanwhile, when removing the SPCPM from UPDN, the AP and F1 scores decrease by 4.01% and 0.02, respectively, while when removing the SPCPM from the model, the AP and F1 scores dropped by 1.35% and 0.03, respectively. This indicates the higher potential and synergistic effect of the SPCPM and EAM.

Combining the analysis of other evaluation indicators, the parallel design of the SPCPM enables the network to maintain a high feature processing speed even with increased processing steps and parameters. This parallelism allows the model to efficiently handle input features and enhance the overall algorithm performance. By utilizing multi-scale spatial pyramid convolution and pooling operations, the SPCPM captures multi-scale information, thereby improving the model's receptive field and feature representation capability. This further enhances the network's ability to detect small targets, which is particularly crucial in pedestrian detection scenarios based on UAVs.

Furthermore, the incorporation of multiple EAM in the model, despite increasing the model's size, leads to performance improvements for UPDN. This is evidenced by the observed increase in the GLOPS metric, indicating improved computational speed. The EAM introduce additional attention mechanisms, enabling the model to focus on important features and enhance its representation and discriminative capabilities. This attention mechanism contributes to higher precision and recall, resulting in an overall improvement in the F1 metric. Additionally, we observed that the introduction of SPCPM and EAM resulted in increased model parameters, computational complexity, and model size. Despite the two-fold increase in the largest of these parameters, such as Params increasing from 7.022M to 14.239M, the magnitude of this increase remains acceptable due to the model's initial small parameter size. Moreover, the performance improvements are evident, particularly the Recall metric.

Finally we select some representative and convincing pedestrian detection examples in real scenes, and use 4×4 image grid to visualize the prediction results of the algorithm, as shown in Fig. 6. (a) represents various sports fields, (b) depicts outdoor scenes, (c) showcases scenes with crowded environments and small targets, and (d) displays other miscellaneous scenarios. It is worth noting that in subplot (a), there are instances of motion blur due to the presence of moving drones. In subplot (b), the first column exhibits target occlusion. Subplot (c) contains a significant amount of occlusion caused by dense pedestrians. Lastly, the last column of subplot (d) includes interference from complex scenes. Despite these challenges, UPDN demonstrates robust performance in addressing these issues, underscoring its superior capabilities.

In conclusion, the introduction of the SPCPM and EAM demonstrates considerable performance gains while maintaining efficiency. The optimization of feature processing speed, enhancement of the model's receptive field and feature representation capabilities, and improved handling of small targets and complex scenes collectively contribute to the significant improvements in our algorithm's performance and efficiency. Adopting the strategy of trading minor overheads for substantial performance gains proves meaningful and impactful in the context of pedestrian detection tasks.

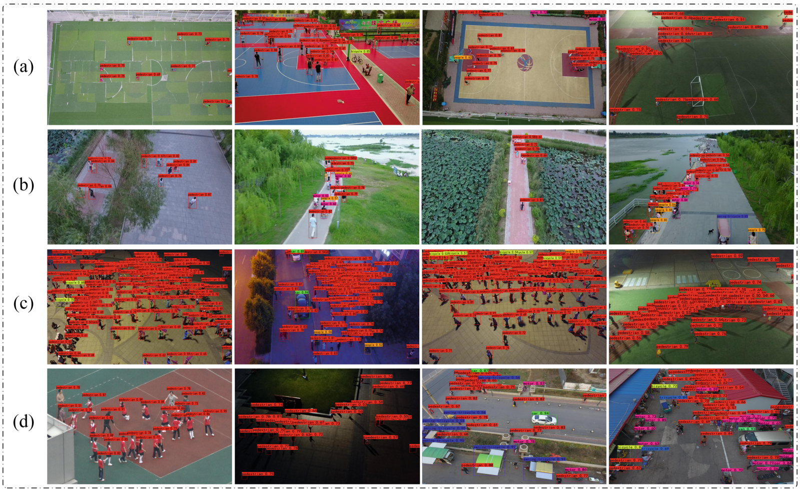

Fig. 6. (a) various sports fields, (b) outdoor scenes, (c) crowded scenes with small targets, and (d) other scenes.

5 Conclusion

In this study, we propose UPDN, a novel pedestrian detection algorithm from UAV perspective, incorporating the SPCPM, EAM, and SIOU. Through extensive experiments and evaluations, we demonstrate the superior performance of our algorithm in terms of AP, F1, Recall, and other evaluation metrics. The SPCPM effectively handles input features and enhances the algorithm's overall performance by capturing multi-scale information. The EAM introduce additional spatial attention, improving the model's characterization and discriminative capabilities. The SIOU loss function improves the algorithm's ability to handle small object detection and complex scenes. Moreover, extensive experiments on the constructed dataset, illustrated that the UPDN outperformed seven classical algorithms, reached the state-of-the-art performance in AP, F1 and Recall, achieving detection speeds of 107.37 FPS. Our research contributes to advancing the field of pedestrian detection and holds great potential for various applications in surveillance, security, and autonomous systems.

Acknowledgement. This work was supported by national natural science foundation of China (No. 62202346), Hubei key research and development program (No. 2021BAA042), open project of engineering research center of Hubei province for clothing information (No. 2022HBCI01), Wuhan applied basic frontier research project (No. 2022 013988065212), MIIT's AI Industry Innovation Task unveils flagship projects (Key technologies, equipment, and systems for flexible customized and intelligent manufacturing in the clothing industry), and Hubei science and technology project of safe production special fund (Scene control platform based on proprioception information computing of artificial intelligence).

References

1. Audebert, N., Le Saux, B., Lefèvre, S.: Beyond RGB: very high resolution urban remote sensing with multimodal deep networks. ISPRS J. Photogramm. Remote. Sens. **140**, 20–32 (2018)
2. Chen, J., Du, C., Zhang, Y., Han, P., Wei, W.: A clustering-based coverage path planning method for autonomous heterogeneous UAVs. IEEE Trans. Intell. Transp. Syst. **23**(12), 25546–25556 (2021)
3. Chen, Z., Qiu, J., Sheng, B., Li, P., Wu, E.: GPSD: generative parking spot detection using multi-clue recovery model. Vis. Comput. **37**(9–11), 2657–2669 (2021)
4. Du, D., et al.: VisDrone-DET2019: the vision meets drone object detection in image challenge results. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (2019)
5. Guo, G., Chen, P., Yu, X., Han, Z., Ye, Q., Gao, S.: Save the tiny, save the all: hierarchical activation network for tiny object detection. IEEE Trans. Circuits Syst. Video Technol. (2023)
6. Hong, M., Li, S., Yang, Y., Zhu, F., Zhao, Q., Lu, L.: SSPNet: scale selection pyramid network for tiny person detection from UAV images. IEEE Geosci. Remote Sens. Lett. **19**, 1–5 (2021)

7. Kellenberger, B., Marcos, D., Tuia, D.: Detecting mammals in UAV images: best practices to address a substantially imbalanced dataset with deep learning. Remote Sens. Environ. **216**, 139–153 (2018)
8. Li, J., et al.: Automatic detection and classification system of domestic waste via multimodel cascaded convolutional neural network. IEEE Trans. Industr. Inf. **18**(1), 163–173 (2021)
9. Liu, W., Liao, S., Ren, W., Hu, W., Yu, Y.: High-level semantic feature detection: a new perspective for pedestrian detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5187–5196 (2019)
10. Luo, Q., Shao, J., Dang, W., Geng, L., Zheng, H., Liu, C.: An efficient multi-scale channel attention network for person re-identification. Vis. Comput. 1–13 (2023)
11. Ma, X., Zhang, Y., Zhang, W., Zhou, H., Yu, H.: SDWBF algorithm: a novel pedestrian detection algorithm in the aerial scene. Drones **6**(3), 76 (2022)
12. Murthy, C.B., Hashmi, M.F., Keskar, A.G.: Optimized MobileNet+ SSD: a real-time pedestrian detection on a low-end edge device. Int. J. Multimed. Inf. Retrieval **10**, 171–184 (2021)
13. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271 (2017)
14. Shao, Z., Li, C., Li, D., Altan, O., Zhang, L., Ding, L.: An accurate matching method for projecting vector data into surveillance video to monitor and protect cultivated land. ISPRS Int. J. Geo Inf. **9**(7), 448 (2020)
15. Tang, Y., Li, B., Liu, M., Chen, B., Wang, Y., Ouyang, W.: AutoPedestrian: an automatic data augmentation and loss function search scheme for pedestrian detection. IEEE Trans. Image Process. **30**, 8483–8496 (2021)
16. Wang, X., He, N., Hong, C., Wang, Q., Chen, M.: Improved YOLOX-X based UAV aerial photography object detection algorithm. Image Vis. Comput. 104697 (2023)
17. Wang, Y., Liu, W., Liu, J., Sun, C.: Cooperative USV–UAV marine search and rescue with visual navigation and reinforcement learning-based control. ISA Trans. (2023)
18. Xie, H., Shin, H.: Two-stream small-scale pedestrian detection network with feature aggregation for drone-view videos. Multidimension. Syst. Signal Process. **32**, 897–913 (2021)
19. Yang, P., Zhang, G., Wang, L., Xu, L., Deng, Q., Yang, M.H.: A part-aware multi-scale fully convolutional network for pedestrian detection. IEEE Trans. Intell. Transp. Syst. **22**(2), 1125–1137 (2020)
20. Zhang, S., Bauckhage, C., Cremers, A.B.: Informed Haar-like features improve pedestrian detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 947–954 (2014)
21. Zhang, S., Yang, J., Schiele, B.: Occluded pedestrian detection through guided attention in CNNs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6995–7003 (2018)
22. Zhang, T., Cao, Y., Zhang, L., Li, X.: Efficient feature fusion network based on center and scale prediction for pedestrian detection. Vis. Comput. **39**(9), 3865–3872 (2023)
23. Zhang, Y., Xu, C., Hemadeh, I.A., El-Hajjar, M., Hanzo, L.: Near-instantaneously adaptive multi-set space-time shift keying for UAV-aided video surveillance. IEEE Trans. Veh. Technol. **69**(11), 12843–12856 (2020)

24. Zhou, H., Yu, G.: Research on fast pedestrian detection algorithm based on autoencoding neural network and AdaBoost. Complexity **2021**, 1–17 (2021)
25. Zhu, X., Lyu, S., Wang, X., Zhao, Q.: TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2778–2788 (2021)