Research on Automatic Detection Algorithm for Pedestrians on the Road Based on Image Processing Method

Qing Zhang
Zhangjiajie College, Jishou University, Zhangjiajie, Hunan 427000, China

***Abstract*—Accurate detection of pedestrian targets can effectively improve the performance level of intelligent transportation and surveillance projects. In order to effectively enhance the accuracy of detecting pedestrian targets on the road, this paper first introduced the traditional pedestrian target detection algorithm, proposed the faster recurrent convolutional neural network (RCNN) algorithm to detect pedestrian targets, and improved it to make good use of the convolutional features at different scales. Finally, support vector machine (SVM), traditional Faster RCNN, and optimized Faster RCNN algorithms were compared by simulation experiments. The results showed that the optimized Faster RCNN algorithm had higher detection accuracy and recall rate, obtained a more accurate target localization frame, and detected faster than SVM and traditional Faster RCNN algorithms; the traditional Faster RCNN algorithm had higher detection accuracy and target frame localization accuracy than the SVM algorithm.**

***Keywords*—Pedestrian detection; recurrent convolutional neural network; scale-invariant feature transform; support vector machine; characteristic scale; Difference of Gaussians operator**

NOMENCLATURE

$D(x, y, \sigma)$: The DoG operator.

DR : The target frame predicted by the algorithm.

$G(x, y, \sigma)$: The Gaussian filter function.

$G(x, y, k\sigma)$: The Gaussian filter function.

GT : The actual target frame in the image.

IoU : The degree of target frame overlap.

$I(x, y)$: The original image.

P : The precision.

R : The recall rate.

I. INTRODUCTION

Economic development continues to improve people's living standards, and the pressure on traffic management increases as more and more vehicles are used in travel [1]. The progress of computer technology has promoted the emergence of intelligent transportation, and the detection of pedestrians is an important component of intelligent transportation [2]. Accurate detection of pedestrians can effectively improve the level of intelligent driving, intelligent monitoring, and other technologies. For example, in intelligent driving, more accurate pedestrian detection can assist drivers to make safe avoidance of pedestrians and reduce the occurrence of traffic accidents; in intelligent monitoring, computers replace humans to make recognition of pedestrians in monitoring videos, track pedestrians, and judge the behavior of pedestrians, thus improving the security level. Manual identification is relatively accurate and is also more intuitive when identifying pedestrians in video images, but human energy is limited and cannot maintain focused attention for a long time, so replacing humans with machines to automatically detect pedestrians is the current trend. Although the detection of pedestrians in images by image processing techniques is not intuitive, it is relatively more comprehensive in measuring targets with smaller scales in images. The traditional pedestrian target detection algorithm uses a feature extraction algorithm to extract image features before classification by a classification algorithm. In the traditional pedestrian target detection algorithm, feature extraction and recognition and detection of images can be considered relatively independent, and the features extracted by the feature extraction algorithm are often statistical local features, which are difficult to reflect image features comprehensively. As deep learning algorithms and computer performance improve, convolutional neural networks (CNNs) have been applied to pedestrian detection. Compared with the traditional detection method, CNNs combine image feature extraction and recognition together and integrated the local features extracted using convolutional kernels into global features, thus making the detection of pedestrian targets more accurately.

Some relevant literature is reviewed below. Xu et al. [3] reconstructed a target detection model called YOLOv3, proposed YOLOv3-promote, and introduced an attention mechanism. They found that the inference speed of the method was faster than the original model and the parameter volume was reduced to one-tenth. Xia et al. [4] put forward a pedestrian detection algorithm based on multi-scale feature extraction and attention feature fusion and found that the algorithm had good detection performance. Liu [5] proposed a deep residual network-based adaptive scale pedestrian detection algorithm and found that the algorithm was applicable to pedestrians of different scales. Yang et al. [6] designed a pedestrian target detection algorithm based on a single shot multibox detector. Subsequent simulation experimental results on VOC2007 and data_sub showed that the maximum value of mAP was 77% and the maximum

accuracy was 96.31%. Zhang et al. [7] designed a pedestrian target detection algorithm based on the histogram of oriented gradient and support vector machine (SVM). They found that the algorithm greatly reduced the computational effort when feature extraction was performed only on candidate regions, thus improving the detection efficiency. Pei et al. [8] designed a multispectral pedestrian target detection algorithm combining visual optical images and infrared images based on deep CNNs and performed simulation tests on the public multispectral benchmark dataset. They found that the log-average miss rate of the algorithm reached 27.6%. Shojaei et al. [9] used transfer component analysis and maximum independent domain in pedestrian target detection. The experimental results on the dataset of INRIA showed that the pedestrian target detection algorithm with domain adaptation had less classification error. Wang et al. [10] designed an algorithm using image fusion and deep learning to improve the performance of unmanned aerial vehicles for detecting pedestrians on the ground in low-illumination environments and verified the excellent performance of the algorithm through experiments.

The previous text is a review of some studies related to pedestrian target detection, and different researchers have used different approaches to identify and detect pedestrian targets. In general, the basic principle of these pedestrian target recognition and detection methods is to extract pedestrian features from images and recognize them based on the extracted features. However, the extracted image features under different scales were not fully considered in the above-mentioned studies; therefore, in this paper, the image features at different scales were utilized.

This paper studied intelligent algorithms for pedestrian target detection on roads. This paper was written in the following structure. The abstract starts with a general statement of the full paper. The introduction gives a brief overview of the related literature. Then, the pedestrian target detection algorithm is described, including the traditional SVM algorithm and the improved Faster RCNN algorithm. Then, the simulation experiment is described. In the experiment, the SVM algorithm, traditional Faster RCNN, and improved Faster RCNN algorithms were compared. The final conclusion summarizes the results of this paper. The contribution of this paper is to optimize the Faster RCNN algorithm for pedestrian target detection, so that it can make full use of the convolutional feature maps at different scales, providing an effective reference for accurate and fast detection of pedestrian targets on the road. The limitation of this paper is that the types of images used in the training of the algorithm were not comprehensive enough, so the richness of the types of images required for algorithm training will be increased to improve the generalizability of the algorithm in the future.

II. Automatic Detection Algorithm for Pedestrians

A. Traditional Pedestrian Target Detection Algorithm

A video consists of multi-frame images, so the detection of pedestrians in the video can be considered as fast detection of pedestrians in the image. The traditional pedestrian target detection algorithm extracts features from the images in the candidate frames, uses a classification algorithm to classify and identify the images in the candidate frames according to the extracted features, and takes the candidate frames judged to be pedestrians as the output. Its specific steps are illustrated in Fig. 1.

1) An input image is pre-processed. A plural number of candidate boxes are added to the image. The size of the candidate boxes is determined according to the actual application.

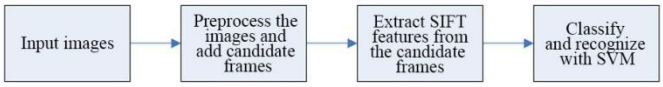

Fig. 1. Traditional pedestrian target detection algorithm.

2) Scale-invariant feature transform (SIFT) feature extraction is performed on the image in the candidate frame. The extraction of SIFT features requires the Difference of Gaussian (DoG) operator [11] to construct a Gaussian difference pyramid. The calculation formula of the DoG operator is:

$$D(x,y,\sigma)=(G(x,y,k\sigma)-G(x,y,\sigma))\times I(x,y), \quad (1)$$

where $D(x,y,\sigma)$ denotes the DoG operator, whose scale factor is σ, $G(x,y,\sigma)$ and $G(x,y,k\sigma)$ are the Gaussian filter functions, whose scale factor is σ and adjacent to σ, respectively, and $I(x,y)$ is the original image. Then, the local extreme points of the image in every scale in the Gaussian difference pyramid composed of DoG operators are searched. The gradient histogram is constructed by choosing the appropriate neighborhood range with the extreme point in every level of the pyramid as the center [12]. Eventually, the histograms corresponding to the main direction of every extreme point and the direction greater than 80% of the gradient peak of the main direction are merged as the SIFT feature.

3) The SIFT features of the collected image sample are separated into a training group and a test group. The training group is used to train and fit the SVM to get the classification function. After the training, the SVM classification function is used to determine whether the image in the candidate frame is a pedestrian according to the SIFT features of the image sample.

B. Pedestrian Target Detection Algorithm using Convolutional Neural Network

In the traditional pedestrian detection algorithm described in the previous text, the features of the image are firstly extracted before recognition by the SVM, which simply means that the extraction of the image features and the recognition of the image are independent of each other. Moreover, the extracted SIFT features are statistical, which are difficult to fully reflect the features of the image and will affect the detection accuracy of the algorithm.

A CNN, as a deep learning algorithm [13], can extract local features of images by convolutional kernels, and the plural features obtained from the plural convolutional kernels can be

combined into global features, taking into account the global and local. The Faster RCNN algorithm is a CNN algorithm for detecting pedestrian targets. It first extracts the convolution feature map of the image through the convolutional and pooling structures of a conventional CNN and obtains the candidate target frame from the map through a regional proposal network (RPN) [14]. After the convolutional features in the candidate target frame are pooled by region of interest (ROI), whether the target frame is a pedestrian is determined in the fully connected layer, and regressive calculation is also performed on the target frame that is judged as a pedestrian in the fully connected layer to get the coordinates of the target frame in the original image.

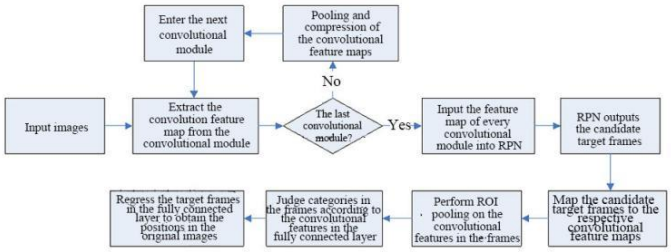

Fig. 2. Pedestrian detection process of the optimized Faster RCNN algorithm.

The convolutional and pooling structures of the CNN in the Faster RCNN algorithm will produce convolutional feature maps of different scales. In this paper, in order to make good use of the convolutional features of different scales to improve pedestrian detection accuracy, some improvements are made to the Faster RCNN algorithm. The optimized detection process is presented in Fig. 2.

1) A pre-processed image is input into the input layer.

2) Convolutional features are extracted in the convolutional module.

3) Whether the convolutional module is the last convolutional module in the conventional CNN structure is determined. If not, the convolutional features are pooled and compressed. The compressed convolutional feature map is input into the next convolutional module for the operation in step 2; if it is, the convolutional feature map of the last convolutional layer obtained in every convolutional module is input into the RPN.

4) The candidate target frame is obtained after calculation in RPN: In this structure, the pixel points in the feature map are regarded as anchor points, and every anchor point generates nine candidate frames with three scales and three length-width ratios with itself as the center [15]. The candidate frame score is calculated according to the convolution features in the candidate frame; the higher the score, the higher the probability of the candidate frame being the target frame. Some candidate frames with high probability are selected and mapped to the original image according to the ratio of the feature map where the candidate frame is located to the original image, and the candidate frames that are beyond the boundary of the original image are deleted. Some candidate frames with high probability are chosen from the remaining candidate frames again as the output of RPN.

5) The candidate target frames calculated by RPN are mapped to the respective convolutional feature maps to which they belong, i.e., the candidate target frames are mapped to the feature maps from which they are obtained.

6) ROI pooling operation [16] is performed on the convolutional features in the candidate target frame. The convolutional map in the target frame is divided into regions according to the required size for ROI pooling, and every region is processed by max-pooling. For example, if the feature map with a size of 9 × 9 in the target frame needs to be compressed into a size of 3 × 3, the feature map in the target frame is divided into regions in a size of 3 × 3, every region is processed by max-pooling, and the result is taken as the value of the corresponding region.

7) The convolutional features processed by ROI compression are input into the fully connected layer to determine whether they are pedestrians, and the position of the target frame in the original image is calculated [17].

The improvement of the optimized Faster RCNN algorithm compared to the traditional Faster RCNN algorithm is that instead of using only the convolutional feature map given by the last convolutional module, convolutional feature maps of different scales in the previous convolutional module are used, making full use of the convolutional features of different scales.

III. Simulation Experiments

A. Experimental Setup

The algorithm in Fig. 3 has five convolutional modules. Convolutional modules 1 and 2 both have two convolutional layers, and there are 32 convolutional kernels in a size of 3 × 3 in every layer [18]. Convolutional modules 3, 4 and 5 all have three convolutional layers, and there are 64 convolutional kernels in a size of 3 × 3 in every layer. Convolutional modules 1~4 have 1 pooling layer after every module, every pooling layer uses a pooling frame in a size of 2 × 2, and the mean-pooling is used in the pooling frame. The RPN module is a fully convolutional structure. Convolutional feature maps obtained from convolutional layers 2, 4, 7, 10, and 13 are all used to calculate the candidate target frames in the RPN. The ROI pooling layer compresses the convolutional features in the candidate target frames, and the compressed size is 6 × 6. The fully connected layer recognizes the category of convolutional features after ROI pooling to determine whether the image in the target frame is a pedestrian. Moreover, the regressive calculation is conducted on the candidate target frame to obtain the coordinates of the target frame in the original image.

The images collected by the author were used as the dataset for the simulation experiment. The images came from a variety of scenes, not limited to traffic intersections. After preliminary removing images with too blurred pedestrians and too dark backgrounds, 15,210 images were left, and the scenes included traffic intersections, parks, supermarkets, subway stations, neighborhoods, etc. Sixty percent of the images were used as the training samples, and the remaining 40% as the test samples.

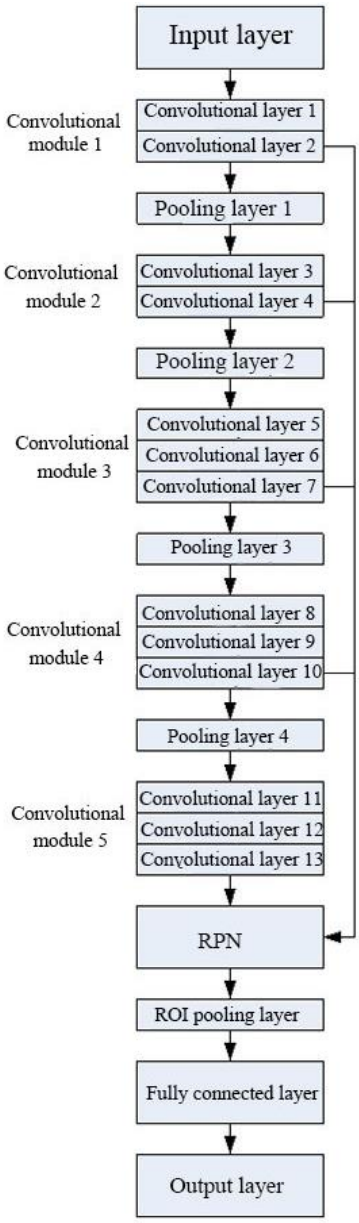

Fig. 3. Basic structure of the improved faster RCNN algorithm.

In the simulation experiment, two detection algorithms, the SVM algorithm and the traditional Faster RCNN algorithm, were also tested to further verify the performance of the improved Faster RCNN algorithm. The SVM algorithm identified pedestrians in the images with SIFT features, and the size of the target frame used for extracting SIFT features was 6 × 6. The basic structure of the traditional Faster RCNN algorithm was similar to that of the optimized Faster RCNN algorithm, and their only difference was that the convolutional feature maps in convolutional layers 2, 4, 7, and 10 were not input into the RPN.

B. Evaluation Criteria

Target detection for pedestrians is a binary classification problem, i.e., to determine whether the target in an image target frame is a pedestrian, so the performance of the detection algorithm can be evaluated using a confusion matrix [19], as shown in Table I. The detection precision and recall rate are calculated using the following equations:

$$\begin{cases} P = \dfrac{TP}{TP + FP} \\ R = \dfrac{TP}{TP + FN} \end{cases}, \tag{2}$$

where P is the precision and R is the recall rate. In addition, the detection speed of the pedestrian target detection algorithm is also quite important. Here, frame per second (FPS) was used to measure the detection speed of the algorithm, i.e., the number of images detected per unit time.

In addition to the above evaluation criteria, the author also used Intersection over Union (IoU) to measure the target frame positioning accuracy of the algorithm. The calculation formula of IoU is:

$$IoU = \frac{DR \cap GT}{DR \cup GT}, \tag{3}$$

where IoU stands for the degree of target frame overlap, DR denotes the target frame predicted by the algorithm, and GT denotes the actual target frame in the image.

C. Experimental Results

The SVM algorithm and traditional Faster RCNN algorithm were compared with the optimized Faster RCNN algorithm. Due to the limitation of space, only some of the detection results are displayed. Fig. 4 shows the pedestrian target detection results of three algorithms for the same image. It was seen from Fig. 4 that the SVM algorithm marked the relatively obvious pedestrians in the image but missed smaller pedestrians, and moreover, it identified two pedestrians as one pedestrian among the marked pedestrians, so it was not very effective in recognizing pedestrian targets overall. In the result of the conventional Faster RCNN algorithm, more pedestrians were detected than in the SVM algorithm, and the two pedestrians that overlap in the picture were also distinguished, but it also missed smaller pedestrian targets. The improved Faster RCNN algorithm not only detected and distinguished relatively significant pedestrians but also detected smaller pedestrian's targets, so its detection performance was the best.

Fig. 5 shows the precision and recall rate of the SVM, traditional Faster RCNN, and optimized Faster RCNN algorithm for the test set. The precision of the SVM algorithm for pedestrian target detection was 75.3%, and the recall rate was 73.8%; the precision of the traditional Faster RCNN algorithm for pedestrian target detection was 86.7%, and the recall rate was 85.7%; the precision of the improved Faster RCNN algorithm had a precision of 96.6% and a recall rate of 95.4%.

TABLE I. CONFUSION MATRIX

	Pedestrian actually	**Background actually**
Judged as pedestrian	*TP*	*FP*
Judgment as background	*FN*	*TN*

Fig. 4. Partial detection results of three pedestrian target detection algorithms.

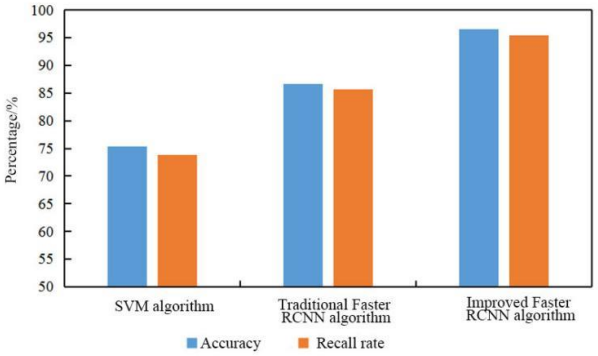

Fig. 5. Detection performance of the three algorithms.

Fig. 6 shows the detection speed of SVM, traditional Faster RCNN, and improved Faster RCNN algorithms for the test set. The detection speed of the SVM, traditional Faster RCNN, and improved Faster RCNN algorithms for pedestrian targets was 10.36 FPS, 21.33 FPS, and 33.51 FPS, respectively. It was seen from Fig. 6 that the SVM algorithm had the lowest detection speed, the traditional RCNN algorithm had a detection speed higher than the SVM algorithm, and the improved Faster RCNN algorithm had a detection speed higher than the traditional RCNN algorithm.

Fig. 7 shows the target frame localization accuracy of the three pedestrian target detection algorithms. The IoU of the target frame of the SVM, traditional Faster RCNN, and improved Faster RCNN algorithms was 67.3%, 79.8%, and 93.4%, respectively. It was observed in Fig. 7 that the target frame obtained by the SVM algorithm in the process of pedestrian detection had the lowest degree of overlap with the actual target frame, the degree of overlap between the target frame obtained by the traditional Faster RCNN and the actual target frame was higher than that of the SVM algorithm, and the degree of overlap between the target frame calculated by the improved algorithm and the actual target frame was higher than that of the traditional Faster RCNN algorithm.

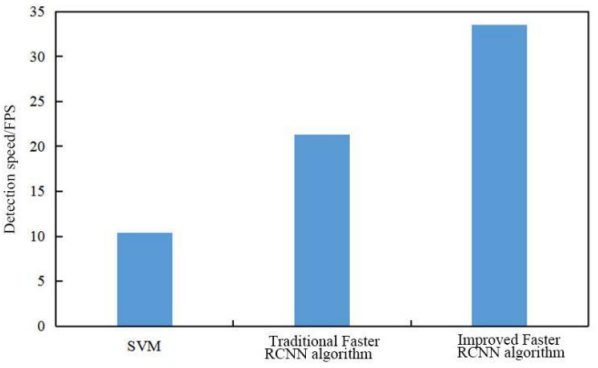

Fig. 6. Detection speed of three pedestrian target detection algorithms.

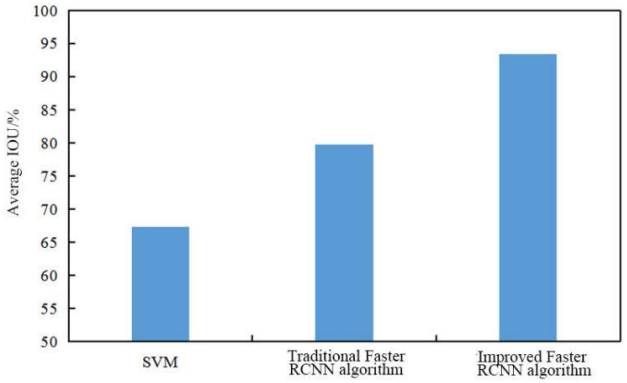

Fig. 7. Target frame localization accuracy of three pedestrian target detection algorithms.

Based on the comparison results of detection accuracy and speed among the three pedestrian target detection algorithms, it was found that the improved Faster RCNN algorithm had the best detection performance, followed by the traditional Faster RCNN algorithm, and the SVM algorithm performed the poorest. The reasons are shown below. The SVM algorithm extracted SIFT features when detecting pedestrian targets, which were statistical features that could not fully reflect the features in the image, so it was difficult to distinguish overlapped pedestrians in the image in the detection process, and the fixed size of the target frame also made it inflexible to distinguish smaller pedestrian targets. The traditional Faster RCNN algorithm used the convolutional structure of a CNN to extract the local and global features of the image, so it performed better than the SVM algorithm in recognition, but only the convolutional features of the last convolutional layer were used in the calculation of the candidate target frame. Even if the target frames with three length-width ratios were used, it was difficult to effectively use the multi-scale features, so the traditional algorithm missed the pedestrians. In the improved Faster RCNN algorithm, the convolutional features of different scales were fully utilized in the calculation of candidate frames, so it effectively recognized smaller pedestrian targets in the images. Moreover, the SVM algorithm used fixed-size target frames for recognition and identified all the target frames, leading to a low detection speed and decreased localization accuracy of the computed target frames; the traditional Faster RCNN algorithm used RPN to pre-compute candidate frames, which reduced the repetitiveness of target frame selection, and different scales of candidate frames improved the localization accuracy of small targets; the improved Faster RCNN algorithm used convolutional features of different scales in the computation of target candidate frames and further improved the localization accuracy of target frames for small pedestrians by using target frames of different scales.

IV. Conclusion

This paper compared the SVM, traditional Faster RCNN, and improved Faster RCNN algorithms in simulation experiments after improving the traditional Faster RCNN algorithm. The experimental results are shown below. (1) The detection results of some images showed that the improved Faster RCNN algorithm effectively distinguished the pedestrians in the image as well as the background and also

achieved better detection results when facing small target pedestrians in the image compared to the other two algorithms. (2) In terms of detection accuracy for pedestrians in images, the detection accuracy and recall rate of the improved Faster RCNN was 75.3% and 73.8%, respectively; the traditional Faster RCNN algorithm was 86.7% and 85.7%, respectively; the SVM algorithm was 96.6% and 95.4%, respectively. (3) In terms of the detection speed, the detection speed of the SVM algorithm was 10.36 FPS, the traditional Faster RCNN algorithm was 21.33 FPS, and the improved Faster RCNN algorithm was 33.51 FPS. (4) In terms of the localization frame accuracy, the IoU of the target frame of the SVM algorithm was 67.3%, the IoU of the traditional Faster RCNN algorithm was 79.8%, and the IoU of the improved Faster RCNN algorithm was 93.4%.

REFERENCES

[1] C. B. Murthy, M. F. Hashmi, G. Muhammad, and S. A. AlQahtani, "YOLOv2PD: An Efficient Pedestrian Detection Algorithm Using Improved YOLOv2 Model," Comput. Mater. Contin., pp. 3015-3031, January 2021.

[2] Z. Xu, W. Zhao, L. Peng, and J. Chen, "Research on Pedestrian Detection Algorithm Based on Deep Learning," J. Phys. Conf. Ser., vol. 1646, pp. 1-6, September 2020.

[3] H. Xu, M. Guo, N. Nedjah, J. Zhang, and P. Li, "Vehicle and Pedestrian Detection Algorithm Based on Lightweight YOLOv3-Promote and Semi-Precision Acceleration," IEEE T. Intell. Transp., vol. 23, pp. 19760-19771, January 2022.

[4] H. Xia, J. Ma, J. Ou, X. Lv, and C. Bai, "Pedestrian detection algorithm based on multi-scale feature extraction and attention feature fusion," Digit. Signal Process. vol. 121, pp. 1-13, November 2022.

[5] S. S. Liu, "Self-adaptive scale pedestrian detection algorithm based on deep residual network," Int. J. Intell. Comput., vol. 12, pp. 318-332, August 2019.

[6] J. Yang, W. Y. He, T. L. Zhang, C. L. Zhang, L. Zeng, and B. F. Nan, "Research on subway pedestrian detection algorithms based on SSD model," IET Intell. Transp. Sy., vol. 14, pp. 1491-1496, November 2020.

[7] Y. Zhang, K. Guo, W. Guo, J. Zhang, and Y. Li, "Pedestrian crossing detection based on HOG and SVM," J. Cyber Secur., vol. 2, pp. 79-88, January 2021.

[8] D. Pei, M. Jing, H. Liu, L. Jiang, and F. Sun, "A fast retinanet fusion framework for multi-spectral pedestrian detection," Infrared Phys. Techn., vol. 105, pp. 1-8, January 2020.

[9] G. Shojaei and F. Razzazi, "Semi-supervised domain adaptation for pedestrian detection in video surveillance based on maximum independence assumption," Int. J. Multimed. Inf. R., vol. 8, pp. 241-252, December 2019.

[10] C. Wang, D. Luo, Y. Liu, B. Xu, and Y. Zhou, "Near-surface pedestrian detection method based on deep learning for UAVs in low illumination environments," Opt. Eng., vol. 61, pp. 1-19, February 2022.

[11] J. Wang, C. Zhao, Z. Huo, Y. Qiao, and H. Sima, "High quality proposal feature generation for crowded pedestrian detection," Pattern Recogn., vol. 128, pp. 1-10, February 2022.

[12] H. Zhou, and G. Yu, "Research on Fast Pedestrian Detection Algorithm Based on Autoencoding Neural Network and AdaBoost," Complexity, vol. 2021, pp. 1-17, March 2021.

[13] Z. J. Wang, Y. Q. Zhao, and C. L. Zhao, "Improved MSER Pedestrian Detection Algorithm based on TOF Camera," J. Phys. Conf. Ser., vol. 1576, pp. 1-6, June 2020.

[14] S. Zhai, S. Dong, D. Shang, and S. Wang, "An Improved Faster R-CNN Pedestrian Detection Algorithm Based on Feature Fusion and Context Analysis," IEEE Access, vol. 8, pp. 138117-138128, January 2020.

[15] S. Y. Cho, J. H. Lee, and C. G. Park, "A Zero-Velocity Detection Algorithm Robust to Various Gait Types for Pedestrian Inertial Navigation," IEEE Sens. J., vol. 22, pp. 4916-4931, March 2021.

[16] J. Ren, C. Niu, and J. Han, "An IF-RCNN Algorithm for Pedestrian Detection in Pedestrian Tunnels," IEEE Access, vol. 8, pp. 165335-165343, January 2020.

[17] G. Li, C. Zong, G. Liu, and T. Zhu, "Application of Convolutional Neural Network (CNN)–AdaBoost Algorithm in Pedestrian Detection," Sensor. Mater., vol. 32, pp. 1997-2006, June 2020.

[18] D. Liu, S. Gao, W. Chi, and D. Fan, "Pedestrian detection algorithm based on improved SSD," Int. J. Comput. Appl. T., vol. 65, pp. 25-35, January 2021.

[19] B. Wang, "Research on Pedestrian Detection Algorithm Based on Image," J. Phys. Conf. Ser., vol. 1345, pp. 1-12, November 2019.