Array 19 (2023) 100318

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array

Occlusion and multi-scale pedestrian detection A review

Wei Chen [a,b,c], Yuxuan Zhu [a], Zijian Tian, CA [a,*], Fan Zhang [a], Minda Yao [b]

[a] School of Mechanical, Electrical & Information Engineering, China University of Mining and Technology (Beijing), Beijing, 100083, China
[b] School of Computer Science & Technology, China University of Mining and Technology, Xuzhou, 221116, China
[c] Key Laboratory of Intelligent Mining and Robotics, Ministry of Emergency Management, China

ARTICLE INFO

Keywords:
Pedestrian detection
Occlusion pedestrian detection
Multi-scale pedestrian detection

ABSTRACT

Pedestrian detection has a wide range of application prospects in many fields such as unmanned driving, intelligent monitoring, robot, etc., and has always been a hot issue in the field of computer vision. In recent years, with the development of deep learning and the proposal of many large pedestrian data sets, pedestrian detection technology has also made great progress, and the detection accuracy and detection speed have been significantly improved. However, the performance of the most advanced pedestrian detection methods is still far behind that of human beings, especially when there is occlusion and scale change, the detection accuracy decreases significantly. Occlusion and scale problems are the key problems to be solved in pedestrian detection. The purpose of this paper is to discuss the research progress of pedestrian detection. Firstly, this paper explores the research status of pedestrian detection in the past four years (2019–2022), focuses on analyzing the occlusion and scale problems of pedestrian detection and corresponding solutions, summarizes the data sets and evaluation methods of pedestrian detection, and finally looks forward to the development trend of the occlusion and scale problems of pedestrian detection.

1. Introduction

Pedestrian detection is a special task of target detection, which is closely related to the development of general target detection. Pedestrian detection aims to find pedestrians and mark the location and size of pedestrians from given images and videos. It is the basis of multiple tasks such as pedestrian rerecognition [1,2], pedestrian tracking [3,4], pedestrian trajectory prediction [5,6] and has a wide range of application prospects in intelligent monitoring [7,8], automatic driving [9,10], intelligent robots [11,12] and other fields. Therefore, pedestrian detection has always been a hot and difficult topic in the field of computer vision. Because pedestrian detection is easily affected by factors such as occlusion, scale, illumination, weather and similarity, the accuracy of pedestrian detection in complex scenes is poor and false positives are easy to occur. Therefore, every year, a large number of scholars focus on coming up with a more robust, faster and more advanced method. Fig. 1 shows the number of studies related to pedestrian detection between 2003 and 2022 (data from Web of Science search, year of publication selected, search results 20).

Pedestrian detection methods are mainly divided into two categories: hand-crafted feature-based methods and deep feature-based methods. The first kind of methods include VJ, HOG, ICF, etc. In 2003, Viola and Jones [13] first combined motion and appearance information together with Adaboost classifier to constitute VJ detector, which can robustly detect pedestrians from various angles with low false alarm rate. In 2005, Navneet Dalal and Bill Triggs [14] proposed HOG detector and used HOG features to train SVM classifier. HOG can describe local shape information and offset the influence of illumination to a certain extent, reducing the data dimension of the characterized image. In 2009, Piotr Dollar et al. [15] proposed the integral channel feature ICF. The second category of methods is mainly divided into one-stage detection methods YOLO series, SSD series, etc., and two-stage detection methods RCNN series. Joseph Redmon et al. [16] proposed YOLO method as a very representative one-stage detection method in 2016. YOLO treats detection as a regression task, divides the whole image into S*S grids, and each grid is responsible for detecting objects within the grid. It uses bounding box localization, conditional probability, confidence, etc., to complete the prediction in one go. Subsequently, YOLOv2 [17], YOLOv3 [18] and other versions were developed. Wei Liu et al. [19] proposed SSD as a one-stage detection method different from YOLO. SSD sets many bounding boxes with different aspect ratios and proportions for the image, and directly

* Corresponding author.
E-mail address: tianzj0726@126.com (Z. Tian).

https://doi.org/10.1016/j.array.2023.100318
Received 30 March 2023; Received in revised form 22 August 2023; Accepted 27 August 2023
Available online 3 September 2023

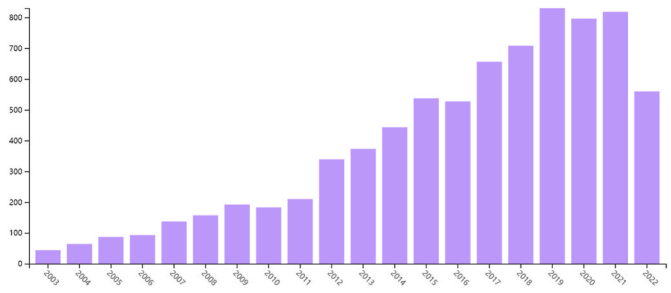

Fig. 1. Number of Web of Science searches for literature related to pedestrian detection, 2003–2022.

calculates the confidence of all bounding boxes in the regression. At the same time, multiple convolution layers with different scales are generated in the convolution layer to detect multi-scale objects and improve the detection accuracy. Ross Girshick et al. [20] introduced Convolutional Neural Network (CNN) into the field of object detection for the first time in 2014, and proposed R-CNN by combining CNN with traditional machine learning methods. R-CNN uses selective search to obtain as many pre-selected boxes as possible, uses CNN network to extract features instead of manually designed features for the first time, and uses multiple SVM classifiers for classification and regression. Then in 2015, Ross Girshick [21] designed a Region Proposal Network (RPN) to replace the selective search in R-CNN, and selected Softmax classifier to replace multiple SVM classifiers, which greatly improved the detection performance. With the rapid development of deep learning technology, many researchers have proposed many detection methods with better performance, such as Faster RCNN [22], MSCNN [23], YOLOv4 [24], CSP [25], SAF-RCNN [26] and other methods based on R-CNN. Although hand-designed features are effective in simple cases, their generalization ability is weak and it is difficult to obtain high-level semantic information. Deep features obtain feature representations based on the learning of a large number of samples, which can be learned end-to-end, with strong robustness and better generalization ability. In recent years, pedestrian detection has gradually evolved from hand-crafted features to deep features.

The detection speed and accuracy of pedestrian detection methods have improved significantly. However, the most advanced pedestrian detection methods are still not comparable to human perception. Pedestrian detection still faces a number of challenges, for example:

(1) Multiple gestures of pedestrian object: Pedestrian objects are both rigid and flexible and pedestrians may present a wide variety of postures, such as stationary, standing, squatting, bending, or walking. In addition, the appearance of pedestrians are all dressed differently, in various colors and styles. These variations pose a great challenge to current pedestrian detection. How to design a pedestrian detection method that is not affected by changes in pedestrian posture and clothing appearance is a challenge that needs to be solved for pedestrian detection nowadays.
(2) Weather and low-lighting: Rain, snow, fog, haze, dust storms and other phenomena seriously reduce the clarity of photos, the reduced visibility of the pedestrian object and the blurred contours severely limit the performance of pedestrian detection. At night when the light intensity is low, the pedestrian images have higher gray values, lack of color information, and increased interference information. Also low light conditions generally use infrared images, which have low resolution, contain less pedestrian information, and have blurred pedestrian object outlines. These problems seriously affect the accuracy of pedestrian detection. How to solve the problem of pedestrian object contouring due to bad weather and low light is a major challenge in pedestrian detection.
(3) Occlusion: Occlusion in pedestrian detection is divided into two main categories: One is intra-class occlusion between pedestrians, and the other is inter-class occlusion of pedestrian objects by other objects. The mutual occlusion between pedestrians brings a lot of interference information, which may lead to false detection and affects detection performance. When the pedestrian object is obscured by other objects, the structure of the pedestrian is incomplete and the information of the pedestrian object is missing, which can easily produce missed detection and reduce the accuracy of pedestrian detection.
(4) Multi-scale pedestrian: Different distances of pedestrian objects from the camera will bring different spatial scales of pedestrians, there may be multiple pedestrians at different scales in the same image. This has caused a significant negative impact on pedestrian detection. Large-scale pedestrians have richer information for better detection. But small-scale pedestrians tend to have lower pixels, blurred appearance and contours, contain less valid information, which easily misses detection and seriously affects the accuracy of pedestrian detection. It is a great challenge to accurately detect pedestrian objects at different scales.
(5) Real-time detection: The application of pedestrian detection in automatic driving, intelligent monitoring, etc. has high requirements for detection speed and needs to be able to detect in real time. With the development of deep learning, the pedestrian detection model is becoming more and more complex. While the detection accuracy is improved, it also brings a lot of calculations, which reduces the detection speed of pedestrian detection. How to achieve real-time detection while improving the accuracy of pedestrian detection is a great challenge in the field of pedestrian detection.

Faced with many challenges, pedestrian detection, although a category of general object detection, can still be studied as an independent problem. In recent years, in order to improve the detection accuracy and detection speed of pedestrian detection, researchers have proposed many new detection methods with better performance. This paper discusses and explores the research status of pedestrian detection from 2019 to 2022, summarizes commonly used pedestrian data sets and evaluation indicators, and focuses on the analysis of occlusion and scale problems and corresponding solutions of pedestrian detection.

2. Research status of pedestrian detection

Pedestrian detection is a special task of object detection and a basic task in real-world applications. It can be divided into two main categories: hand-crafted features based and deep learning features based. In recent years, many new pedestrian detection algorithms have been proposed, and the detection performance is constantly improved. The success relies heavily on large-scale datasets, such as, Caltech [27], KITTI [28], CityPersons [29].The following is a brief summary of the progress in pedestrian detection in 2019–2022.

2.1. Hand-crafted features based

Although the earliest hand-made features such as HOG, Hear, LBP and LUV are no longer so excellent with the development of pedestrian detection technology, they also provide ideas and innovative inspiration for later researchers. Over the past 19 years, researchers have designed many new feature descriptors with better performance. Wenshu Li [30] et al. designed eHOG feature by enhancing the contrast of HOG feature in order to balance the speed and accuracy of pedestrian detection. Ritesh Kumar Mishra [31] reduces the redundant information of HSG by adding a feature selection module to reduce the feature size and improve the calculation speed. Ming Yang [32]et al. proposed a Feature

Descriptor generation model (DGM), which can systematically generate various HOG descriptors for pedestrian detection and evaluate the detection performance of each descriptor. GuoYun Lian [33] proposed QGWLD descriptor by combining WLD feature and quaternion representation, which not only extracts features from gray images but also fuses color information and texture information, as shown in Fig. 2. Lienhard Pfeifer [34] believes that Shearlets feature can obtain the direction information well to detect the edge direction in the target image. Two image features, shearlet amplitude and shearlet histogram, are designed.

Some researchers believe that the fusion of different feature operators can effectively use the characteristics of different features and improve the performance. Yu Jiang [35] et al. extracted a new fusion feature HOG-LBP by means of series fusion of HOG features and LBP features. Zijie Xie [36] et al. serially concatenated HOG features and three-layer LBP features extracted in three times to form a new multi-level concatenated feature, and performed dimensionality reduction operation (PCA) before feature fusion. The concatenated features are shown in Fig. 3. Kaushal Kumar [37] et al. combined HSG and NRULBP to generate new descriptors, which incorporated shape and texture information and contained more valuable feature information. Hongzhi Zhou [38] et al. fused the feature vectors of color image features and depth image features into a complex vector by weighted fusion method. Daxue Liu [39] et al. proposed Shallow and Deep feature Fusion (SDFF) based on ACF, which uses shallow features to quickly generate pedestrian proposals and effectively removes false positives through deep features.

2.2. Deep features based

Pedestrian detection based on deep features is further divided into one-stage and two-stage. Recently, a number of single-stage pedestrian detection methods have been proposed. Wei Liu [40] et al. proposed CSP detector to directly predict pedestrian center point and scale through high-level semantic feature detection, which opened up a new idea for later researchers. The model framework is shown in Fig. 4. Tao Zhang [41] et al. proposed F-CSP network. F-CSP introduces FPN and BFP to improve the original feature fusion module. FPN can reduce the number of channels of the feature map, and BFP fuses multiple feature maps into a feature map, reducing other parameters. Based on the original YOLOv3 network model, Jingwei Cao [42] et al. improved the multi-scale bounding box prediction based on receptive field and introduced the Soft-NMS algorithm, which improved the performance of the detection algorithm in various complex scenes. Haohui Lv [43] et al. proposed YOLOv5-AC based on YOLOv5s, introduced L1 regularization in BN layer to remove networks with small impact factors to reduce model size and improve speed, introduced CEM module to extract more features and used two attention modules CxAM and CnAM to filter useless features. Correct the position offset to improve the accuracy. Wei Liu [44] et al. designed the Progressive Localization Fitting (ALF) module, which

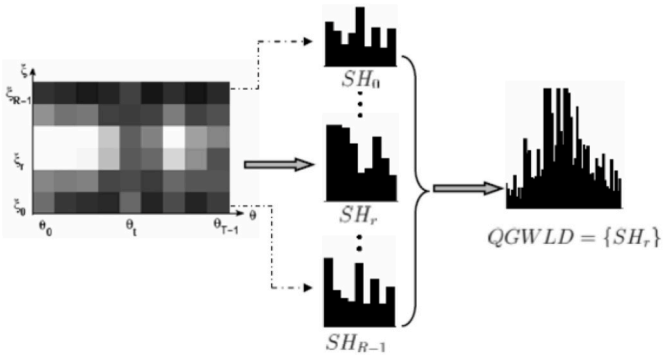

Fig. 2. QGWLD pedestrian detector(image from literature [33]).

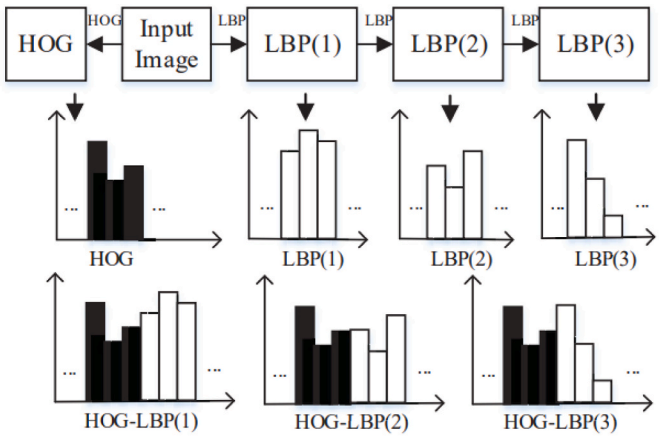

Fig. 3. Multilayer cascading features (image from literature [36]).

uses a series of predictors to gradually convolutional evolve the default anchor box of SSD. Residual learning and multi-scale context coding are combined to improve the accuracy of localization and enhance the prediction ability. Mahmoud Saeidi [45] et al. designed a DM-PPP pedestrian detection model based on the estimation of different parts. The improved SSD is used to generate initial candidates and expand the region with safety boundaries, and the pedestrian proposal is divided into nine parts and evaluated independently. Chintakindi Balaram Murthy [46] et al. proposed an optimized MobileNet+SSD network, which makes feature extraction, deformation, and classification components work together to reduce the number of parameters to improve the detection speed. Yuang Zhang [47] et al. regard single-stage pedestrian detection as a variational inference problem, and propose a customized Auto-coded variational Bayes (AEVB) algorithm by modeling dense proposals as latent variables. The algorithm process is shown in Fig. 5.

There are many two-stage pedestrian detection algorithms with excellent performance that have been proposed in recent years. Garrick Brazil [48] et al. proposed AR-Ped pedestrian detection method. AR-RPN can autoregressively generate and improve features and predictions by retaining more context information and improving performance by using convolutional resampling layer, and its framework is shown in Fig. 6. Zhihang Fu [49] et al. think that although rich context information is important, the core area of the target instance is more important, and propose the ExtAtt two-stage pedestrian detection framework. Jin Ren [50] et al. improved Faster RCNN, designed a method of pooling Context Environment (PCE) around the candidate window, and fused different levels of features to improve the detection accuracy. Fiseha B. Tesema [51] et al. believe that manual features can be a good supplement to CNN features, and combine manual features and CNN features for detection on the basis of RPN+BF. RuihaoYin [52] et al. proposed a Multi-resolution Generative Adversarial Network (MRGAN) to generate high-resolution images from low-resolution images for multi-resolution pedestrian detection, introducing a novel perceptual loss. Peijia Yu [53] et al. performed quality assessment by fusing multi-channel visual features and proposed a new reduction Adjustment module (RA). RA can improve the connection between feature channels and enhance the features with less information. Zebin Lin [54] et al. proposed an example Guided Contrastive learning (EGCL) model, which learns a feature transformation module through contrastive learning, projects the initial feature space into a new feature space, minimizes the semantic distance between pedestrians to eliminate the appearance diversity of pedestrians, and maximizes the semantic distance between pedestrians and the background. Han Xie [55] et al. designed a deformable convolution with an attention module to sample from non-rigid locations and extract attention feature maps with context information.

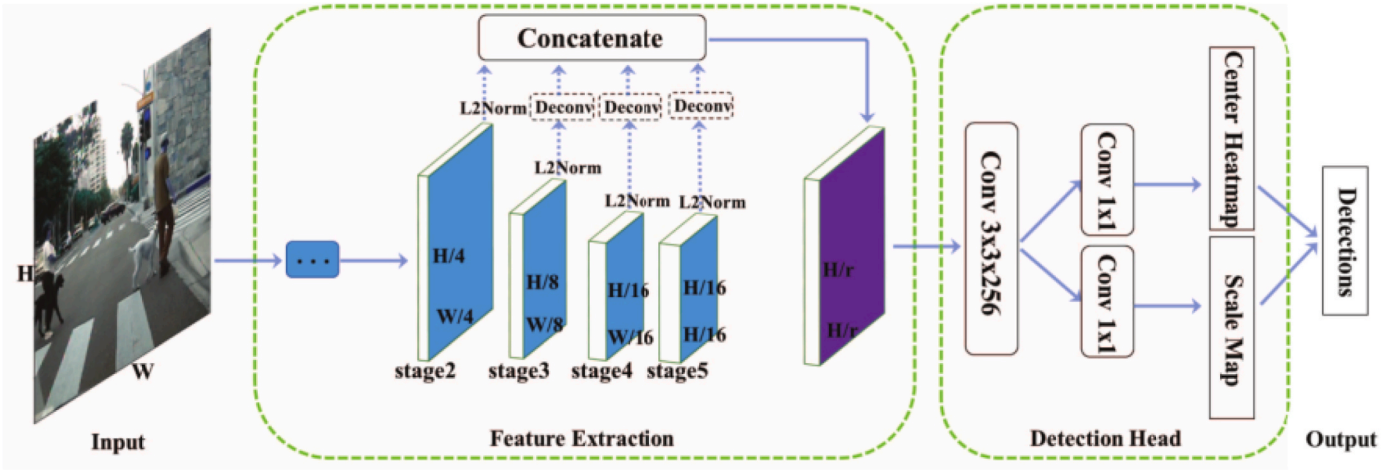

Fig. 4. CSP algorithm (image from literature [40]).

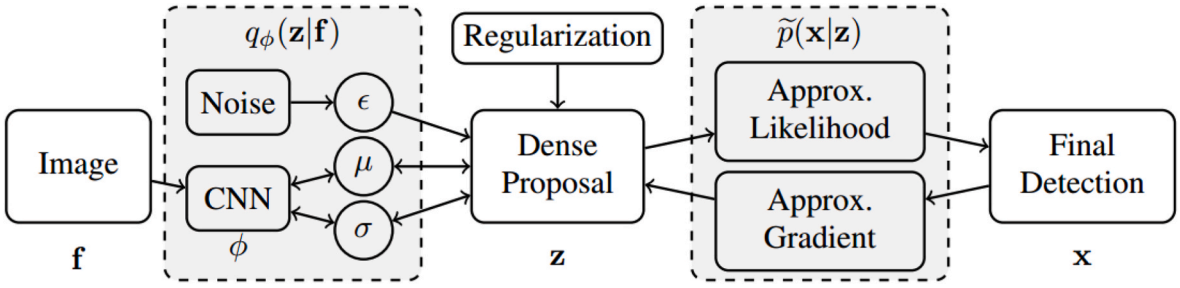

Fig. 5. AEVB algorithm (image from literature [47]).

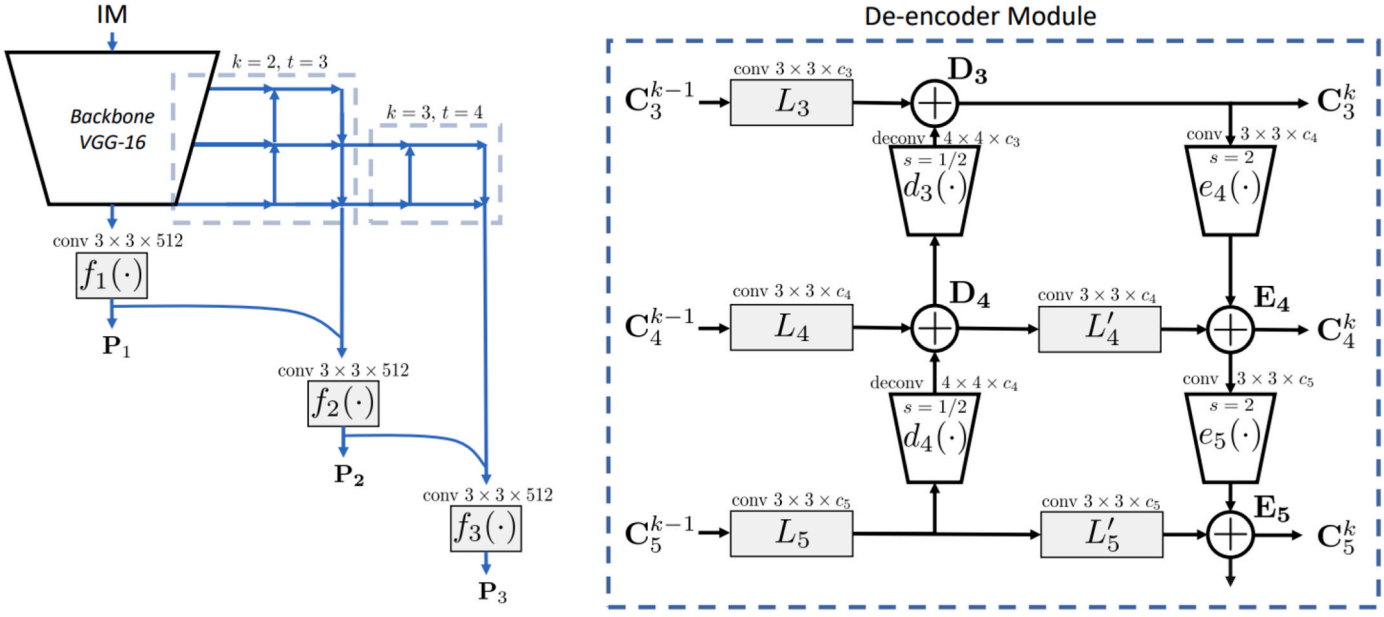

Fig. 6. AR-RPN framework and de-encoder module (image from literature [48]).

3. Datasets and performance evaluation

In the past few decades, the performance of pedestrian detection has been greatly improved, and a large-scale research system has been formed. At the same time, some datasets dedicated to pedestrian detection and some evaluation methods for pedestrian detection performance evaluation have emerged in the field of pedestrian detection.

3.1. Datasets

The dataset is the basis of the research in the field of object detection. It is not only a general means to measure the performance of similar algorithms, but also a power to promote the development of object detection research. The size of the dataset and the quality of the labeling information are very important for the detector training, and accurate

labeling information can help the detector to learn the specified content. To date, published pedestrian datasets include MIT [56], INRIA [14], KAIST [57], ECP [58], Caltech [27], Cityperson [29], KITTI [28], Daimler [59], NICTA [60], ETH [61], CrowdHuman [62], WidePedestrian and WiderPerson [63] and others. The attributes of these datasets are summarized in Table 1. These datasets each have different characteristics depending on their content. Such as CrowdHuman, WidePedestrian and WiderPerson. The content of these datasets is more diversified and diverse, which helps to improve the robustness and generalization of the network, while Caltech, CityPersons and KITTI, which have more perfect annotation information, have better effects for occlusion and multi-scale scenes, and therefore have a wide range of applications. Three datasets Caltech, CityPersons and KITTI are detailed here.

The Caltech dataset was published by Caltech in 2009. It is mainly captured by a car camera running normally in a regular traffic scene. The video is about 10 h long and has about 250 k frames, which contains 350,000 bounding boxes and 2300 pedestrian annotations. And the annotations include bounding boxes and correspondences for different occlusion levels.

The CityPersons dataset is a subset of the Cityscapes dataset and contains only pedestrian annotations. There are 2975 images for training and 500 and 1575 images for validation and testing. On average, each image contains 7 pedestrians, and the density of pedestrians is high and there are many occlusions.

KITTI is one of the important datasets in the field of autonomous driving, which provides a large amount of real scene data, about 50 min of outdoor scenes. It has 7481 images for training, 7518 images for testing, and their corresponding point cloud data with ~200,000 3D labeled boxes. There are up to 15 vehicles and 30 pedestrians in each image, and various degrees of occlusion and truncation are included.

3.2. Evaluation method

The detection performance of pedestrian detector is generally reflected by the corresponding evaluation method, and an excellent evaluation method can objectively reflect the excellence of pedestrian detector.

Table 1
Summary of pedestrian datasets.

Dataset name	Publish year	Images total	Images size	Description
MIT	2000	924	64 × 128	An earlier publicly available pedestrian dataset
KAIST	2015	95328	640 × 480	A multispectral pedestrian detection dataset
Caltech	2009	250000	640 × 480	A widely used dataset with a large amount of data and rich annotation information
Cityperson	2017	5050	2048 × 1024	A subset of the Cityscapes dataset whose image background is an urban landscape
NICTA	2008	30,758		A large scale pedestrian data set with static images
ETH	2007	1803	640 × 480	A pedestrian dataset based on binocular vision
USC	2005	359		A small pedestrian dataset with images mostly from surveillance video
CrowdHuman	2018	24370	608 × 608	A dense occluded pedestrian dataset
WiderPerson	2019	13382	1400 × 800	A dense pedestrian detection in the wild dataset
ECP	2019	47300	1920 × 1024	A pedestrian dataset with images from multiple European countries
KITTI	2012	14,999	1240 × 370	An important dataset in the field of autonomous driving

Average precision AP is one of the commonly used indicators in the detection field. Generally, the performance of the model is dynamically evaluated by plotting the Precious-Recall(P-R) curve, where the abscissa is the recall rate and the ordinate is the accuracy rate. Their calculations can be shown in (1) and (2). Where TP (True Positive) predicts positive samples and the prediction is correct. The closer this metric is to the annotated number of pedestrians in the validation set, the higher detection rate of the detector is indicated. FP (False Positive) predicts a positive sample but is wrong. This index reflects the false detection rate, and the lower the false detection rate, the better. FN (False Negative) The prediction result is negative samples but the prediction result is wrong, this index reflects the missed detection rate, the smaller the index, the better. Based on the TP and FP detections, the precision P(β) and recall R(β) can be computed as a function of the confidence threshold β. P-R curve can be obtained by varying the confidence threshold, and then the Average Precision (AP) can be found. The mean Average Precision (mAP) is a commonly used evaluation metric in multi-object detection and multi-label image classification. The mAP is the sum of the average accuracy AP on the multi-class classification task and the average, which measures how good the model is on all classes, its calculations can be shown in (3), c represents the total number of categories.

$$Recall = \frac{N_{TP}}{(N_{TP} + N_{FN})} \tag{1}$$

$$Precision = \frac{N_{TP}}{(N_{TP} + N_{FP})} \tag{2}$$

$$mAP = \frac{\sum_{k=0}^{c} AP_k}{c} \tag{3}$$

Intersection Over Union (IOU) is a measure of the overlap between the prediction box and the Ground Truth, which can determine whether the predicted detection box meets the conditions. Its calculations can be shown in (4), where B_{gt} represents the Ground Truth, B_p represents the predicted bounding box.

$$IOU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \tag{4}$$

The log-average Miss Rate(MR) is an indicator to describe the miss rate in the detection results. False Positive Per Image(FPPI) describes the average false detection rate per image. Their calculations can be shown in (5) and (6). The MR-FPPI curve is similar to the P-R curve used for object detection and can reflect the overall performance of the detector. Mr^{-2} uses a curve with FPPI as the abscissa and log(MR) as the ordinate. Nine FPPS in the range of [0.01,1] are uniformly selected to obtain their corresponding nine log(MR) values, and these vertical values are averaged. Finally, the above average values are restored to the percentage form of MR by exponential operation, and the MR^{-2} index is obtained, the smaller the index is, the better the performance of the detector.

$$MR = \frac{N_{FN}}{N_{gt}} = 1 - Recall \tag{5}$$

$$FPPI = \frac{N_{FP}}{N} \tag{6}$$

4. Multi-scale pedestrian detection

Although pedestrian detection technology has made great progress in recent years, multi-scale problem has always been a big challenge for pedestrian detection. In the convolutional feature map, the features of large-scale and small-scale pedestrian instances are very different. In addition, large-scale pedestrians also have more abundant information, while small-scale pedestrians have low resolution and blurred edges,

which are difficult to distinguish from the background. Multi-scale pedestrian detection should be able to detect and accurately locate pedestrian instances of different scales without being affected. This is a great challenge for pedestrian detection, and this chapter reviews the 19 years that some researchers have proposed new methods for the scale problem, as shown in Fig. 7. Tables 2–4 show the detection capability of the introduced pedestrian detection method.

4.1. Based on context information

Context information can effectively improve the accuracy of pedestrian detection. Small-scale pedestrian targets are difficult to provide rich ROI features, and additional context information can effectively supplement the original ROI features. In 2019, Han Xie and Yunfan Chen [64] et al. introduced deconvolution and porous modules into the Faster R-CNN network to add more semantic context information to enhance the feature map. The new synthetic feature map can provide richer visual detail information and semantic context representation. Junhao Hu and Lei Jin [65] et al. proposed FPN++ on the basis of FPN. In order to extract deeper features with more semantic information and increase the local receptive field, the convolution is replaced by dilated convolution and the step size of the convolution is reduced in the backbone of FPN. Chen Zhang [66] et al. proposed a multi-scale pedestrian detector. The detector fuses pooled features with different resolutions and context information in the fully connected layer to make full use of their unique features, and designs a recurrent convolutional layer with feature connection (RCL-FC) to extract strong and deep context information by iteratively integrating large RF features and small RF features. In 2021, Xiaoting Yang [67] et al. proposed Scale-sensitive Feature Recombination Network (SSNet), which uses a multi-parallel branch sampling module to flexibly adjust the receptive field and anchor stride to extract scale-sensitive features, and introduces a context enhancement fusion module to reduce the information loss of medium and high-level features. The framework is shown in Fig. 8. The method proposed by Han Xie and Xiaoting Yang et al. performs well for small-scale pedestrian detection, but the detection speed is not fast enough to meet the real-time requirements. The processing speed of Xiaoting Yang's method is relatively fast, and it shows good compatibility in different driving scenarios.

4.2. Based on attentional mechanisms

The attention mechanism can find the correlation between the original data and highlight some important features. The introduction of the attention mechanism in pedestrian detection can better fuse different features and improve the robustness of pedestrian detection at different scales. In 2019, Zhichang Chen [68] et al. believed that the competitive attention module can be used to improve the hard negative samples and multi-scale problems of pedestrian detection, and the competitive attention module helps the network architecture to obtain

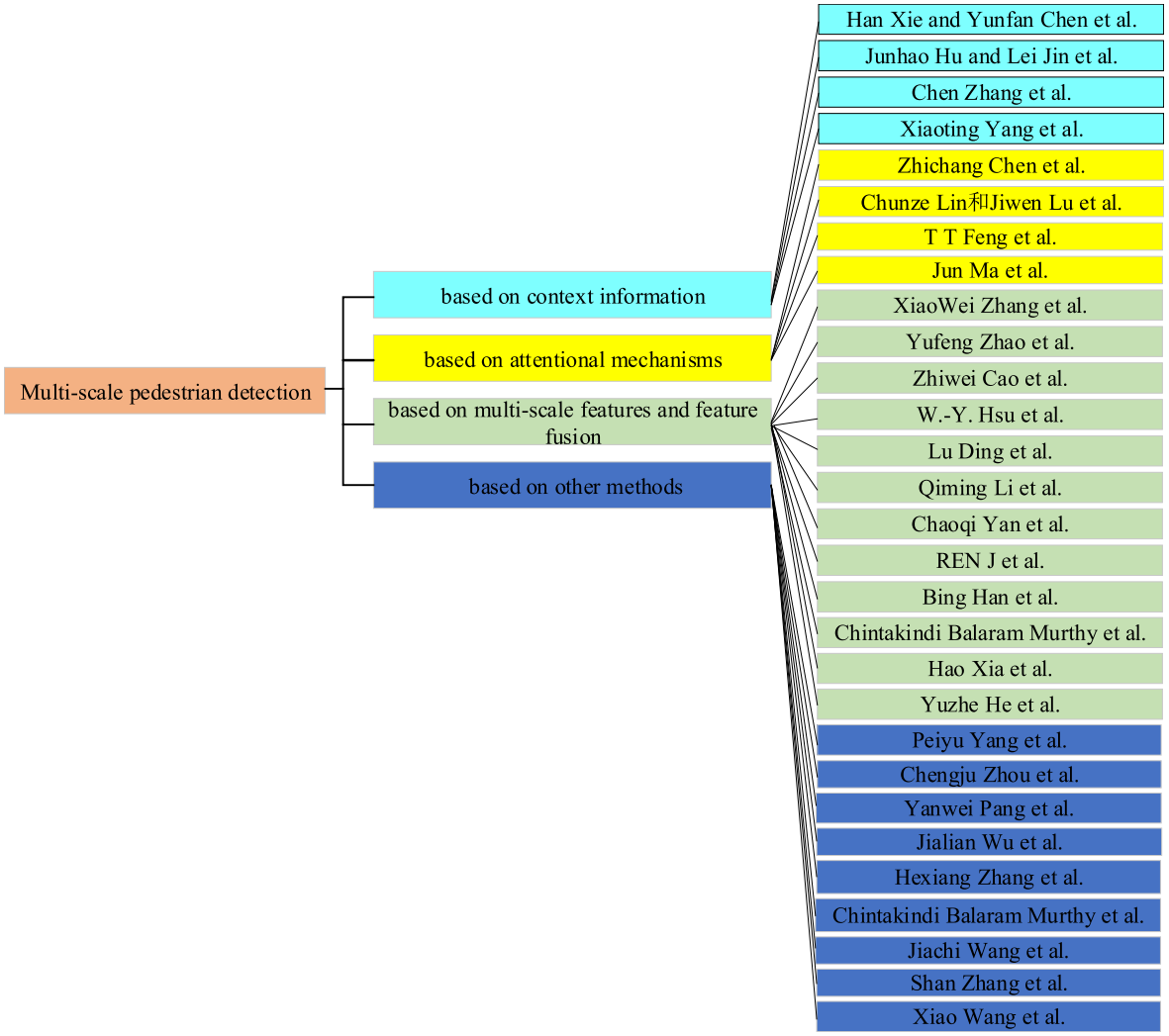

Fig. 7. Summary of algorithms in this chapter.

Table 2
Detection performance of the presented method on the Caltech (R represents Reasonable setting, M^0 represents Medium setting, F represents Far setting, MR^{-2} as evaluation index in %) and KITTI (E represents Easy setting, M^1 represents Moderate setting, H represents Hard setting, Red numbers indicate the evaluation index in % with average precision, Black numbers indicate the evaluation index in % with the mean average precision) dataset.

	Backbone	Year	Caltech			KITTI			Journal
			R	M^0	F	E	M^1	H	
[86]	VGG-16	2019	8.81	/	/	85.62	74.99	69.65	TIFS
[80]	VGG-16	2020			91.08				T-ITS
[66]	VGG-16	2019	/	/	/	83.51	73.29	67.67	MTA
[69]	VGG-16	2020	7.84			85.05	68.87	66.93	TIP
[92]	VGG-16	2020				77.94	65.6	60.45	TCSVT
[76]	VGG-16	2021	14	38	71	84.68	77.59	69.21	NCA
[72]	ResNet	2020	7.41	28.77	70.69				Ieee Access
[87]	ResNet-50	2020	6.8	21.2	63.6				MM
[67]	ResNet-50	2021	6.3						Sensors
[84]	ResNet-50	2021				78.43	67.2	61.88	T-ITS
[64]	Inception ResNet	2019	7.79	/	/	/	/	/	APIN
[91]	/	2020	7.32	24.34	51.36	85.63	70.52	69.31	Neurocomputing

Table 3
Detection performance of the presented method on the Cityperson dataset (R represents Reasonable setting, H represents Heavy setting, P represents Partial setting, B represents Bare setting, S represents Small setting, M represents Medium setting, L represents Large setting, A represents ALL setting, MR^{-2} as evaluation index in %).

	Backbone	Year	Cityperson								Journal
			R	H	P	B	S	M	L	A	
[65]	ResNet-50	2019								37.36	ICME
[71]	ResNet-50	2021	11.95	50.1	11.1	7.9					JRTIP
[74]	ResNet-50	2021	9.5	48.4	9.3	6.2	15.5	3.5	6.2		WCMC
[67]	ResNet-50	2021	11.9				18.0	6.9	7.5		Sensors
[82]	ResNet-50	2022	11.47	43.84	10.05	6.29					APIN
[77]	Res2Net50	2021	9.4	44.5	9.1	6.2	14.0	3.3	5.2		Inf. Sci
[78]	ResNeXt-50	2022	11.6	57.3	12.6	6.5					APIN
[85]	Tailored VGG-16	2022	10.3	50.2	10.6	6.5	16.5	5.9	6.1	35.8	T-ITS

Table 4
Detection performance of the presented method on the other dataset (VOC series dataset with mAP as evaluation index, 2D-light dataset with MR^{-2} as evaluation index, all units are in %).

	Backbone	Year	VOC2007	VOC2012	VOC07 + 12	2D-light	Journal
[79]	VGG-16	2020	79.6				Ieee Access
[70]	VGG-16	2020	78				CCISP
[73]	ResNet-50	2020				30.09	Ieee Access
[89]	CSPResNet50	2022	87.3				MVA
[88]	Darknet-53	2021			87.616		TVC
[75]	CSPDarknet-53	2021		77.3			Ieee Access

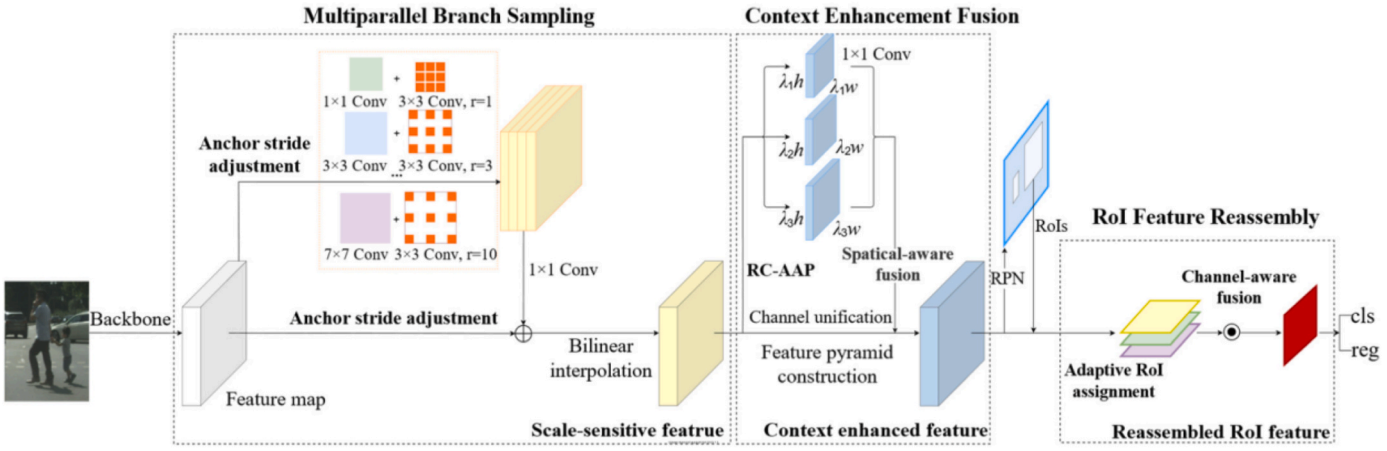

Fig. 8. SSNet framework(image from literature [67]).

valuable specific information from multi-scale feature maps. They designed a novel architecture (CompAt). In 2020, Chunze Lin and Jiwen Lu [69] et al. proposed a granular aware deep feature learning method (CA-GDFL), which can use the convolution backbone to generate multiple feature maps to represent pedestrian targets at different scales. Then, the scale-aware pedestrian attention module is used to generate the attention map, and the attention map and the feature map are fused into a granular aware feature map to better distinguish the background and pedestrians. Finally, the zoom and zoom module is used to combine context information to enhance shallow features. T T Feng [70] et al. introduced the feature enhancement strategy and attention mechanism module into SSD, and proposed a new pedestrian detection algorithm. In 2021, the MSCM ANet network proposed by Jun Ma [71] et al. introduced a multi-scale convolution module, and added an attention module to make the detection network focus on the characteristics of pedestrians. The feature maps extracted from the attention module are merged by the pyramid fusion structure, and then the CRM is used for regression and classification. T T Feng et al. 's method has certain recognition ability for small targets, but the model parameters and calculation are not small enough. Although the detection accuracy of Jun Ma et al.'s method is improved, the detection speed is reduced.

4.3. Based on multi-scale features and feature fusion

The extracted features are very important in pedestrian detection, and the amount of information contained in the features directly affects the detection performance of pedestrian detection. For small-scale pedestrian detection, single-scale feature maps are not robust. Some researchers propose to extract multi-scale feature maps from images of different scales for prediction, and multi-scale features can combine semantic features and geometric features to obtain richer information. In 2020, XiaoWei Zhang [72] et al. proposed a scale-aware hierarchical detection network, which introduced a cross-scale feature aggregation module on FPN to fuse the semantics and localization of pedestrians at different scales, and enhanced the feature pyramid representation. Yufeng Zhao [73] et al. proposed a multi-focus pedestrian detection network (MFDN) based on Faster R-CNN. MFDN uses multiple refocused images as input to solve the problem of small-scale pedestrian detection and confusion with anthropomorphic negative samples. The Cumulative Probability Selection (CPS) layer was introduced to combine the results of multiple detection branches. In 2021, Zhiwei Cao [74] et al. proposed a multi-scale anchor-free region proposal network. Adaptive Channel Feature Fusion (ACFF) was used to select features at different scales, and multi-scale anchor-free region Proposal (MSAF) was used to predict the position, center offset and height of the person to obtain the boundary box. Reducing hyperparameters improves the scale imbalance problem. W.-Y. Hsu [75] et al. introduced the idea of segmentation on the basis of YOLOv4. After extracting the feature information of the entire image, the non-overlapping pedestrians were divided into new subimages and the feature information of the subimages was extracted. Lu Ding [76] et al. proposed a new framework based on SSD. FPN was introduced into the SSD framework to effectively utilize features of different scales, and NIN strategy was used to fuse global information and local details to enhance feature maps. Qiming Li [77] et al. proposed an Anchor-free Multi-scale Pedestrian Detection Network (MPAF-Net) to simplify pedestrian detection into the prediction task of center and scale, and improve the multi-scale representation strength of the detector for pedestrians by designing a new bottleneck block and improving Res2Net. And the CRF-based message passing mechanism is used to enhance the multi-scale features, whose framework is shown in Fig. 9. In 2022, Chaoqi Yan [78] et al. proposed R-SSD based on the SSD architecture. During feature fusion, feature maps of different scales are fused, and then the fusion block is fused with other layers to generate six prediction layers from different depths. A residual block is added to each prediction layer of SSD to improve prediction performance. The method proposed by XiaoWei Zhang, Zhiwei Cao, Qiming Li et al. has good performance for small-scale pedestrian detection. Lu Ding's method has relatively fast detection speed, but it is not effective for small-scale pedestrian detection. Some hyperparameters of W.-Y. Hsu's method need to be selected manually and the segmentation function leads to an increase in reasoning time. Chaoqi Yan's method does not need to configure anchors for different data sets, but it does not perform well for crowded pedestrian scenes.

At the same time, feature maps of different depths contain different semantic information. Shallow feature maps have strong semantic information that can well activate small-scale pedestrians, while deep feature maps have rich spatial information. In 2020, REN J [79] et al. proposed the IF-RCNN method, which adjusted the size of the anchor box in the RPN network and used three convolution kernels with different sizes to generate candidate regions from the last layer feature map. In order to obtain the deep semantic features and shallow detail features at the same time, the feature maps generated by the Conv5_3 and Conv4_3 layers of the feature extraction network are fused. Bing Han [80] et al. proposed a small-scale perception network (SSN). SNN generates more proposal regions from lower convolutional layers, uses the highest convolutional layer to supplement the global information of the image, and uses deconvolution to merge convolutional layers to generate new feature maps. Chintakindi Balaram Murthy [81] et al. proposed an improved YOLOv2 pedestrian detection algorithm (YOLOv2PD), which used multi-layer feature fusion (MLFF) strategy to improve the feature extraction ability of the model and removed a convolutional layer in the last layer to reduce the computational complexity. Normalization is applied to improve the loss function to improve the detection performance. In 2022, Hao Xia [82] et al. proposed the MAFA-Net pedestrian detection method, which uses deep expansion blocks to extract deeper features, uses pedestrian attention

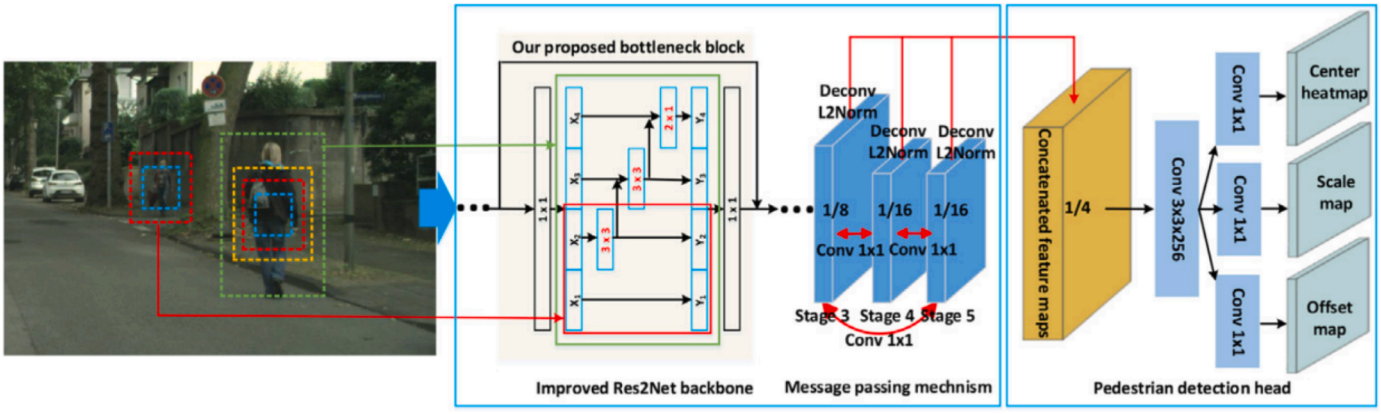

Fig. 9. MPAF-Net framework(image from literature [77]).

blocks to obtain more mutual information between features, and introduces a feature aggregation module to combine high-level features and low-level features. Yuzhe He [83] et al. proposed a multi-scale feature balance enhancement network. The network extends a bottom-up short-circuit path to improve the utilization of low-level feature information, designs a feature balance module to obtain the same amount of semantic information from different resolutions, and uses a feature enhancement module to expand the receptive field to retain more information. The method proposed by REN J makes good use of the deep features of the network and has good detection effect. Bing Han's method has good performance for small-scale pedestrians, but the detection speed is relatively slow. Chintakindi Balaram Murthy's method can detect in real time, but it has poor performance for small-scale pedestrians and occluded pedestrians. The method proposed by Hao Xia and Yuzhe He both explored the use of feature fusion to improve detection accuracy, but did not balance accuracy and speed well.

4.4. Based on other methods

In recent years, some researchers have proposed some other methods to deal with the scale problem in pedestrian detection. Some researchers have introduced semantic segmentation into pedestrian detection. Semantic segmentation actually understands the image from the pixel level, which can separate the pedestrian target and background to help small-scale pedestrian detection achieve better detection performance. In 2021, Peiyu Yang and Guofeng Zhang [84] et al. proposed a partially aware Multi-scale fully convolutional Network (PEMS-FCN). Pems-fcn designs a multi-scale FCN to detect pedestrians at different scales and utilizes partially aware RoI pooling and instance RoI pooling to combine local and global information to deal with occlusion. In 2022, Chengju Zhou [85] et al. proposed a detection framework based on R-FCN, which proposed several new components to improve the detection performance of small-scale pedestrians by using semantic segmentation. Some researchers believe that large-scale pedestrians can obtain good detection accuracy, while small-scale pedestrian detection is difficult, so the relationship between small-scale pedestrians and large-scale pedestrians can be used to help small-scale pedestrian detection. In 2019, Yanwei Pang and Jiale Cao [86] et al. used the connection between large-scale and small-scale pedestrians to recover the details of small-scale pedestrians, and designed JCS-Net for small-scale pedestrian detection through joint optimization classification and super-resolution. In 2020, Jialian Wu [87] et al. believed that the weak representation of small-scale pedestrians was the main reason for the classifier to miss them, so they proposed component of Self-simulation Learning (SML). SML forces the feature representation of small-scale pedestrians to be close to that of large-scale pedestrians by designing a simulation loss. Peiyu Yang's method has relatively slow detection speed and little help for normal scale pedestrian detection. The method proposed by Chengju Zhou is not effective for small-scale pedestrian detection with height less than 40 pixels. The method proposed by Yanwei Pang only considers the similarity between reconstructed pedestrians and large-scale pedestrians, and ignores the dissimilarity between reconstructed background and large-scale pedestrians. The method proposed by Jialian Wu is a general component, which can be effectively applied to other detectors with backbone networks.

Some researchers have improved the robustness of multi-scale pedestrian detection through multi-scale detection. In 2021, Hexiang Zhang [88] et al. proposed YOLOv3-Z based on the Retinex algorithm and the improved YOLOv3 algorithm. The Retinex algorithm is used as a preprocessing algorithm to improve the brightness and contrast of pedestrians, and the YOLOv3 algorithm is improved by adding multiple scale detection. In 2022, Chintakindi Balaram Murthy [89] et al. proposed a lightweight real-time detection algorithm (EfficientLiteSet), which introduced a three-scale transformer prediction head (TPH) in Tiny-YOLOv4 to replace the original detection head, which could improve the detection performance of small targets. At the same time, the attention mechanism is introduced to weaken the influence of messy information and only focus on the key information. The method proposed by Hexiang Zhang is not good enough for small-scale data sets. Chintakindi Balaram Murthy's method is a lightweight real-time detector, which is easier to adapt to edge devices.

In addition, Jiachi Wang [90] et al. proposed SICNN network, which adds a spatial pyramid pooling layer on the basis of Faster RCNN to pool all feature vectors, and uses pooling Windows of different scales to pool different target areas to improve the performance of multi-scale pedestrian detection. Shan Zhang and Xiaoshan Yang [91] et al. believe that the body shape of pedestrians is always rectangular, so they designed an asymmetric Multi-stage network (AMSNet). AMSNet uses rectangular convolution kernels instead of square convolutions to extract features, and designs a three-stage framework based on RPN to exclude non-pedestrian prediction boxes according to coarse-to-fine features. Xiao Wang [92] et al. discovered a discriminant surface (SSS) in the score scale space by studying the functional relationship between pedestrian scores and scales. According to the position of pedestrians on the discriminant surface, pedestrians can be distinguished at different scale levels. Based on this, the S3D pedestrian detection method is proposed, and its framework is shown in Fig. 10. The method proposed by Jiachi Wang and Xiao Wang performs well for small-scale pedestrian detection.

5. Occluded pedestrian detection

Most of the pedestrian detectors have good detection capability for pedestrian objects with high visibility and small percentage of occlusion. However, when the pedestrian target is occluded, the performance of the pedestrian detector may degrade, especially when there is severe occlusion the performance of the detector can be significantly degraded. And occlusion can appear in different places, and the degree of occlusion is difficult to control, so how to deal with occlusion is a hot topic of research in recent years. This chapter reviews the new methods proposed by researchers for the occlusion problem in the past 19–22 years, as shown in Fig. 11. Table 5 shows the detection performance of the occluded pedestrian detection algorithms reviewed in this chapter.

5.1. Based on body parts

Because of the lack of sufficient information of occluded pedestrians, the detector cannot accurately identify pedestrians and effectively locate the part of pedestrians. Some researchers construct local detectors to assist whole-body detection according to the characteristics of different parts of the human body. The detection method based on body parts can effectively improve the accuracy of occluded pedestrian detection. In 2019, LIN C-Y [93] et al. proposed PedJointNet, a dual-branch architecture, which uses the feature pyramid module to predict the head-shoulder region and the full-body region respectively, and proposes a new sharing mechanism that can dynamically and adaptively adjust the weights between the head-shoulder region and the full-body region. Junhua Gu [94] et al. designed a new method for joint pedestrian and body part detection. The BPIF representation was used to represent the semantic relationship between each body part and highlight the characteristics of each part, which had good robustness to partial occlusion. Chunluan Zhou [95] et al. used a multi-label learning method to construct a joint learning part detector. Part detectors share a set of decision trees, which reduce the computational complexity by exploiting partial correlations and obtain an overall score for all parts. Muhammad Mobaidul Islam [96] et al. proposed a new training strategy, which learns a set of body parts during training, and each part corresponds to the defined occluded pedestrian, detects pedestrians from the body parts, and turns the occluded pedestrian detection problem into a multi-class classification problem. In 2021, Shanshan Zhang [97] et al. found that some channel features can not only be located but also

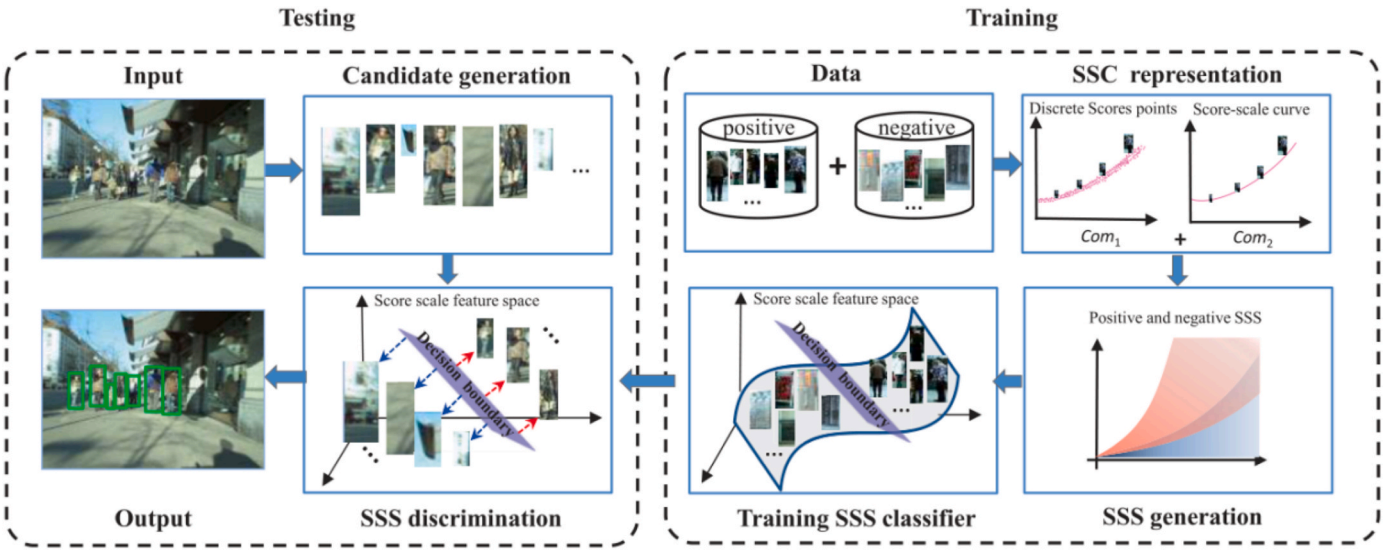

Fig. 10. S^{3}D framework(image from literature [92]).

correspond to certain body parts, and the occlusion situation can be represented as a specific combination of body parts. Therefore, a channel attention mechanism was designed to obtain a more efficient representation of occluded pedestrians. In 2022, Jiali Ding [98] et al. proposed a head-aware pedestrian detection network (HAPNet) based on the structural relationship between the human body and the head. HAPNet detects both the head and the body of pedestrians, and proposes a head-side affinity module to represent the association between the head and the body. Ameen Abdelmutalab [99] et al. proposed a multi-branch pedestrian detection algorithm (MB-CSP), which used four branches to extract features from the pedestrian's ground, middle, lower and whole body respectively, and at the same time to mark the visibility of different pedestrian parts for training. The model framework is shown in Fig. 12. The methods proposed by Chunluan Zhou, Shanshan Zhang, Jiali Ding etc. have good performance in severe occlusion, and the attention network proposed by Shanshan Zhang can be applied to other detection tasks. Ameen Abdelmutalab's method achieves good detection performance, but it does not work well when the pedestrian height is below 50 pixels.

5.2. Based on context information

Some researchers use context information to supplement the features of occluded pedestrians to better represent occluded pedestrians and improve the robustness of occluded pedestrian detection. In 2019, Chi Fei [100] et al. proposed Context-aware feature Learning (CAFL). CAFL uses the pixel-level context embedding module to integrate the context information of multiple surrounding regions into the feature layer, improving the discrimination ability of the detector and enhancing the robustness to occlusion. Wittawin Susutti [101] et al. believe that local information is as important as global information, and propose a multi-channel pedestrian detection method based on appearance. The complete image is divided into multiple regions, and the features of each region are extracted and analyzed. ACF and uLBP features are used to hierarchically combine to represent pedestrians. In 2020, Zhaoqing Li [102] et al. proposed the SCN model by combining segmentation ideas and context information. SCN uses segmentation to obtain pedestrian contours to generate more accurate pedestrian information and uses LSTM for information exchange. Sheping Zhai [103] et al. proposed FCF R-CNN, which adopts a progressive cascade strategy in the backbone network to extract features from different layers for fusion, and the feature information of shallow layers is utilized. A multi-layer LSTM module is also designed to extract the global context information of the image. In 2022, Hangzhi Jiang [104] et al. proposed the SMPD pedestrian detector, designed a semantic integration module, independently of the CSP detector, used the semantic context of urban scenes for detection, and fused the output of CSP detector and semantic integration module as the final detection result. Zhenxing Liu [105] et al. proposed a global context-aware feature extraction module, which can combine context information with local and global pedestrian characteristics. At the same time, a visual feature enhancement module was designed to introduce unblocked upper body information into the network to enhance the representation ability of extracted features. Chi Fei's method performs well for severe occlusion and the network is relatively lightweight. Wittawin Susutti's method is fast and can be applied in real time. A Although the detection accuracy of Zhaoqing Li's method is good, the training and detection are slow. The auxiliary module proposed by Hangzhi Jiang can be applied to the rest of the general detectors. Zhenxing Liu's method has unstable performance in different datasets and weak generalization ability.

5.3. Based on GAN network

Some researchers believe that GAN networks can be used to cope with occlusion in pedestrian detection. In 2020, Songyan Liu [106] et al. proposed attribute Preserving Generative Adversarial Network (APGAN), which introduced a new occlusion reconstruction loss on the basis of CycleGAN to improve image quality and complete attribute preservation after style changes. Firstly, pedestrian images of different scales and occlusive were simulated to improve the diversity of data. Then, the style changes are performed to make the generated pedestrian images more realisti. In 2021, Jin Xie and Yanwei Pang [107] et al. proposed a mask-guided attention network (MGAN). The standard pedestrian detection branch of MGAN uses whole-body annotations to generate features for supervision, while the MGA branch uses unoccluded body regions to generate spatial attention maps that enhance the features of unoccluded regions and suppress the background. MGAN also uses OSEM to introduce occlusion level into the sampling process and designs OSL occlusion brightness loss. Yi Jin [108] et al. proposed an end-to-end Super-Resolution Detection (SRD) network. SRD performed Super-resolution reconstruction (SRGAN) on the image to better distinguish pedestrians from the background, improved the loss function of Faster R-CNN, and fixed the size of the input image. Occluded pedestrians are re-detected by using the improved Faster R-CNN. Yongqiang Zhang [109] et al. proposed a keypoint-guided Super-resolution network (KGSNet). Firstly, the corresponding SR image blocks are

Fig. 11. Summary of algorithms in this chapter.

Table 5
Detection performance of the presented method on the Caltech(R^0 represents Reasonable setting, P represents Partial setting, H represents Heavy setting, the log-average miss rate as evaluation index in %)and CityPersons(R^1 represents Reasonable setting, H represents Heavy setting, P represents Partial setting, B represents Bare setting, the log-average miss rate as evaluation index in %)dataset.

	Backbone	Year	Caltech			CityPersons				Journal
			R^0	P	H	R^1	H	P	B	
[132]	VGG-16	2019	8.0	12.2	37.9	11.1	44.3	11.2	6.9	ICCV
[130]	VGG-16	2019	7.54	22.25						Ieee Access
[135]	VGG-16	2019			38.53					TCSVT
[103]	VGG-16	2020	8.02	11.27	50.9					Ieee Access
[117]	VGG-16	2020	8.4	14.3	41.5	13.61	46.17			PRL
[107]	VGG-16	2021	6.4		36.4	9.3	36.7			TIP
[137]	VGG-16	2021	4.7		34.6	12.37	49.81			TIP
[136]	VGG-16	2021	6.5		36.4	11.3	50.5			TIP
[120]	VGG-16	2021	6.38	12.98	37.54	9.88	38.12	10.37	5.23	T-ITS
[98]	VGG-16	2022				12.9	54.3	13.6		SCIS
[119]	VGG-16	2022	6.4		34.8	9.3	37.0			SCIS
[138]	VGG-19	2021				10.4	47.4	9.4	7.0	TCSVT
[95]		2019	9.8	15.6	48.6					Pattern Recognit
[100]	Resnet-50	2019		15.02	38.71	11.4	50.4	12.1	7.6	Ieee Access
[116]	ResNet-50	2020					51.6			Ieee Access
[106]	ResNet-50	2020				10.9	43.7	10.3	6.7	Neurocomputing
[111]	ResNet50	2021	5.8		45.7	11.2	47.9	10.4	7.2	IET Image Process
[97]	ResNet-50	2021	4.52		33.28	11.9	39.06			IJCV
[109]	ResNet-50	2021	3.89		34.17	10.96	39.68			TNNLS
[112]	ResNet-50	2021	4.2			9.7	47.3	8.8	6.5	ICPR
[122]	ResNet-50	2022	4.4			11.6	42.8	11.9		T-ITS
[127]	ResNet-50	2022	3.8		35.7	9.8	46	9.2	6.8	CAIS
[104]	ResNet-50	2022	4.2		44.8	9.9	36.6	9.0	6.5	Neurocomputing
[139]	ResNet-50	2022	5.2		35.3	10.7	40.9	9.9	6.9	Neurocomputing
[121]	ResNet-50	2022	3.45		30.18	9.7				T-ITS
[115]	ResNet-50	2022				10.1	47.4	9.6	6.7	Tjsc
[99]	ResNet-50	2022	5.3		30.55	10.08	47.29	10.22	6.12	T-ITS
[105]	ResNet-50	2022				11.0	41.7			NPL
[129]	ResNet-101	2022	4.6		42.7	11.03	47.82	10.43	7.26	APIN
[126]	ResNeSt101	2022				9.71	42.53	8.89	5.83	Electronics
[93]	DetNet	2019				13.45	52.17			Ieee Access
[110]	CircleNet	2020	10.21	20.27	44.53	11.77	50.22	12.21	7.14	T-ITS
[123]	DLA-34	2021				8.5	44.7	8.4	5.7	Neurocomputing
[124]	DLA-34	2021				8.8	46.6	8.3	5.8	TMM
[125]	HRNet-W32	2022			32.1	9.4	43.1	8.3	5.6	T-ITS
[128]	HRNet-W32	2022				9.3	45.3	7.8	5.6	Neurocomputing

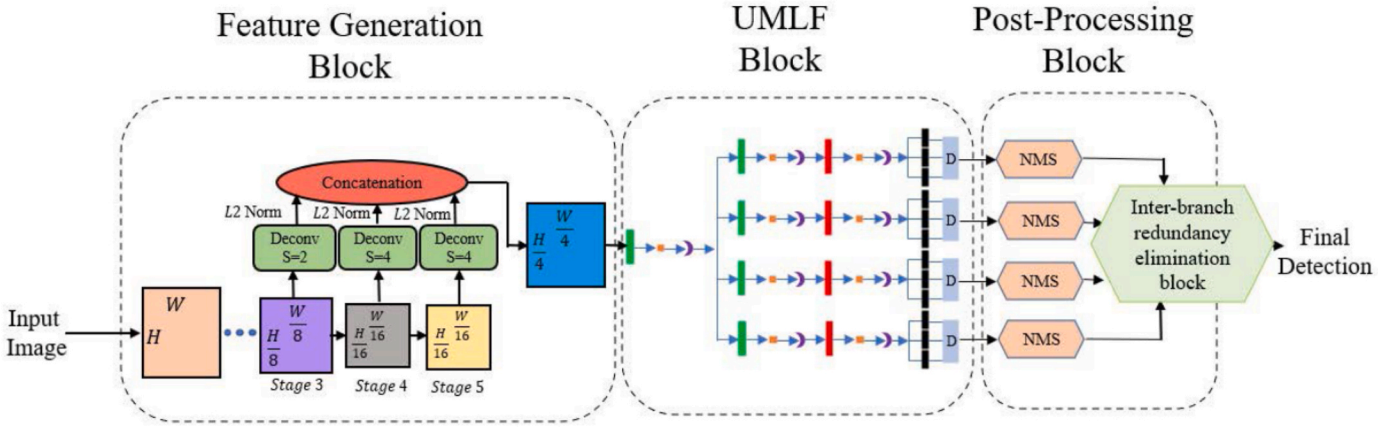

Fig. 12. MB-CSP Framework(image from literature [99]).

generated according to the key points of the pedestrian in the small-scale pedestrian image through the super-resolution network. The method proposed by Songyan Liu and Jin Xie has strong generalization ability. Yi Jin's method still has good performance in low-quality image detection.

5.4. Based on feature enhancement

Researchers believe that the representation of occluded pedestrians can be improved by enhancing features, so that the detector can detect the occlusion form more easily. Some researchers enhance the occluded pedestrian features by feature fusion. In 2020, the feature learning model CircleNet proposed by Tianliang Zhang [110] et al., uses multiple top-down and bottom-up paths to reciprocally fuse features. The top-down path can strengthen semantic information and the bottom-up path can expand the receptive field and combine contextual information. And the instance decomposition training strategy is used to detect pedestrian instances with different occlusions in each cycle. In 2021,

Binjie Ruan [111] et al. proposed the MF-CSP detector combined with semantic features on the basis of CSP, designed a semantic feature enhancement module to enhance semantic features by fusing feature maps at different levels, and then detected pedestrians according to their position and proportion. Yuzhi Tan [112] et al. proposed the PRF-Ped anchorage free pedestrian detector, designed the Bidirectional Feature Enhancement Module (BFEM) to fuse the features of different levels, and proposed the prior based receptive field Block (PRFB) to guide the network to pay attention to pedestrians and reduce the interference of background information. The model framework is shown in Fig. 13. In 2022, Guiyi Yang [113] et al. proposed a Multi-scale Feature Attention Fusion Network (PFF-CB). PPF can enhance the key important features through parallel feature fusion, and the convolutional attention module can enhance the effective feature information of space and channel and adjust its weight. Jing Wang [114] et al. proposed a pedestrian detection technology in crowded scenes, which used the unoccluded part to assist pedestrians to improve the detection accuracy as a whole. Dual Region Feature Generation (DRFG) and Selective Kernel feature fusion (SKFF) were designed to eliminate false positives in crowded pedestrian detection, and Paired Multiple Instance Prediction (PMIP) was introduced. Yugang Qin [115] et al. proposed FE-CSP single-stage pedestrian detection algorithm, which combines GCB and attention and selects deformable convolution to improve the feature extraction ability of the backbone network. At the same time, the feature pyramid network is used to fuse the low-level and high-level features to obtain more semantic information. The method of Binjie Ruan and Jing Wang does not consider the detection speed. Yugang Qin's method is difficult to accurately detect a single pedestrian when pedestrians are very crowded. Guiyi Yang's method is a general module, which can be well applied to other detectors.

Some other researchers enhance occluded pedestrian features by attention and improving feature extraction networks. In 2020, Ruihong Yin [116] et al. proposed the DA-Net network. In order to better deal with occlusion and highlight the unoccluded part of the pedestrian, DA-Net uses CWAM to weight each channel feature, and uses GAM to supplement the global information for the occluded part. Tengtao Zou [117] et al. proposed an Attention Guided Neural network model (AGNN) to deal with occlusion. AGNN is actually a pedestrian classification model, and the most important thing is that the attention guidance network weights the local features to enhance the features. In 2021, Xiaotao Shao [118] et al. proposed an occluded pedestrian detection algorithm (MFPN). MFPN designs a new feature extraction network DFR, which can effectively enhance the semantic information and contour of occluded pedestrians while reducing the computational complexity. Jin Xie [119] et al. introduced a Local Spatial Co-occurrence (PSC) module based on Faster R-CNN. The PSC module can obtain the intra-part and inter-part spatial co-occurrence of different body parts through graph convolution to enhance the body part feature representation and the final feature representation. Ye He [120] et al. proposed the Pedestrian Detection Network (DMSFLN), which performs pedestrian detection through a standard whole-body detection branch and an additional visible body branch, and the two branches are supervised by full-body annotations and visible body annotations, respectively. In 2022, Yan Luo [121] et al. proposed the Occluded Pedestrian Detection Network (SA-DPM), which used sequential attention to extract local features and introduced Frobenius norm for constraints to ensure the diversity of local features. Its framework is shown in Fig. 14. The Feature Calibration Network (FC-Net) proposed by Tianliang Zhang [122] et al. can adaptively detect pedestrians under various occlusions. Its core is a self-activation module, which can simply and effectively enhance features by reusing classification weights to estimate pedestrian activation maps. The method of Xiaotao Shao and Tengtao Zou still has good performance in severe occlusion. Ye He's method does not perform well enough in crowded scenes. The network module proposed by Yan Luo is complex and has too many parameters. Tianliang Zhang's method is suitable for general object detection tasks.

5.5. Based on center point detection

Some researchers regard the center point of pedestrians as a high-level semantic feature, and transform the pedestrian detection problem into the detection problem of high-level semantic features, and improve the performance of occluded pedestrian detection through center point detection. In 2021, Yang Wang [123] et al. proposed the MAPD detector for crowd detection, which obtained the attributes of center, scale, offset, density and id through five output graphs, and proposed a multi-attribute NMS based on the segmented NMS algorithm and id information to distinguish mutually occluded pedestrians. Jialiang Zhang [124] et al. proposed an attribute-aware pedestrian detector, which transformed pedestrian detection into high-level feature detection. In addition to semantic information and location, pedestrian attribute features were also introduced to represent the differences between pedestrians. In 2022, the OAF-Net proposed by Qiming Li and Yijing Su [125] et al., uses the occlusion perception detection head to detect pedestrians with different occlusion levels. The occlusion perception detection head contains three central prediction branches with different occlusion levels, and selects the most appropriate branch according to the pedestrian occlusion level. Yu Zhang [126] et al. proposed an anchor-free method (CSPRS) combining pedestrian detector CSP, region resolution learning and segmentation learning. Xinchen Lin [127] et al. proposed a scale-refined CSP model (SASR-CSP). SASR-CSP establishes a regression branch refinement scale through the distance between the candidate box center and the edge of the image, which can improve the prediction results and accuracy when the predicted center deviates from the actual situation. The model is shown in Fig. 15. Qiming Li [128] et al. proposed Bi-Center Detector to solve the occlusion

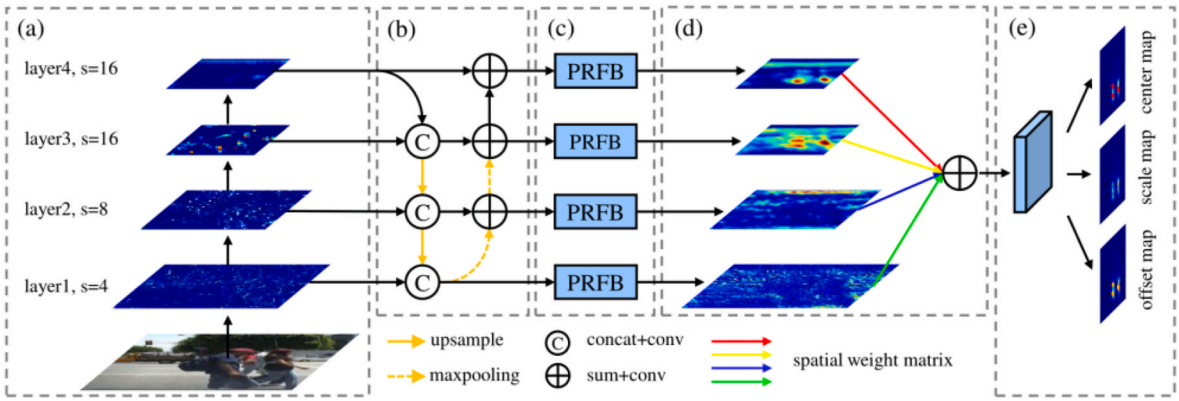

Fig. 13. PRF-Ped Framework(image from literature [112]).

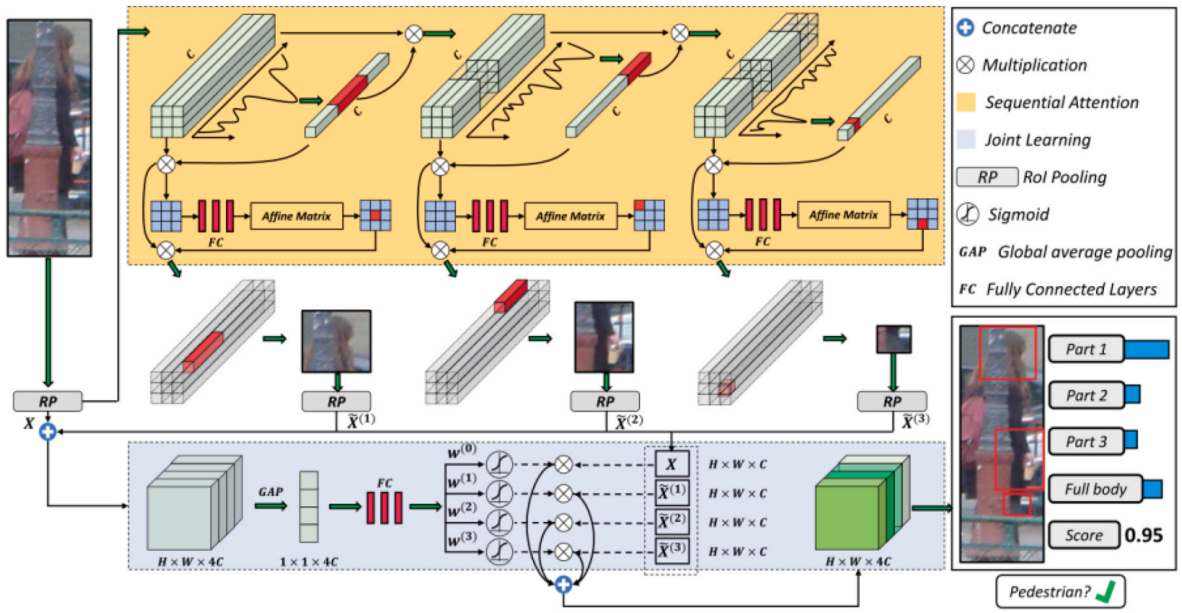

Fig. 14. SA-DPM Framework(image from literature [121]).

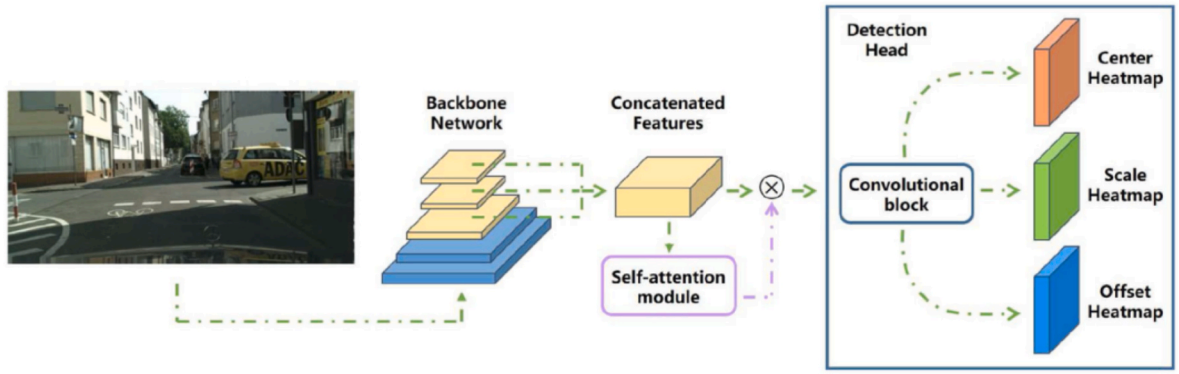

Fig. 15. SASR-CSP Framework(image from literature [127]).

problem. The first center prediction branch downweights the occluded part around the center of the occluded pedestrian according to the degree of occlusion. The second center prediction branch integrates the pedestrian occlusion ratio into the Gaussian mask of center ground truth. Zhuowei Wang [129] et al. proposed a multi-branch detection network based on triggered attention (MBDN). MBDN predicts the center points of the upper, middle and lower pedestrian respectively through three prediction branches. Yang Wang and Qiming Li's method still has good results in crowd scenes. Yu Zhang and Qiming Li's method still performs well in severe occlusion.

5.6. Based on other methods

In addition to the methods introduced in the previous sections, some other researchers have proposed new occluded pedestrian detection methods. In 2019, some researchers generated saliency maps to eliminate background interference on occluded pedestrians to improve detection accuracy. Inyong Yun [130] et al. used saliency and bounding box alignment for pedestrian detection. The model uses saliency to eliminate false positives such as trees, knots FCN and CAM to improve the resolution of the confidence map and successfully recall missing body parts. Wei Wei [131] et al. used MobileNet for detection and positioning to balance detection speed and detection accuracy, introduced binocular depth information to reduce the influence of clutters, and used the saliency of depth map to distinguish pedestrians to improve detection accuracy. Chunluan Zhou [132] et al. proposed a discriminative feature transformation to solve the pedestrian occlusion problem. In the feature space, the pedestrian example was close to the center of the easily classified non-occlusion pedestrian example and the background example was pushed to the easily classified background example, and a new network was constructed by combining Faster RCNN. Lijun Xu [133] et al. proposed to improve the accuracy of pedestrian detection by using edge perception, and proposed a new edge perception pool module to extract edge maps from trained EDS and fuse them with features to obtain more detailed pedestrian contour information. Yuting Xu [134] et al. used deep Omega shape feature to represent pedestrians. The multi-path detection and online hard example mining were introduced to reduce the impact of scale change and clutter background, and a non-maximum suppression method with boot-up strategy was designed to improve the detection performance under partial occlusion. Chunze Lin [135] et al. parallel multi-scale networks and human parsing generator. The multi-scale network uses multi-granularity features to detect and select the most appropriate feature map to predict the pedestrian at the corresponding scale. The human parsing generator generates fine-grained attention maps to guide the multi-scale network to pay more attention to the pedestrian visible area. In 2021, Yi Tang [136] et al. used data augmentation to improve pedestrian detection in crowded environments. Firstly, the data

augmentation strategy and loss function were represented by the probabilities of different hyperparameters respectively, and then an automatic pedestrian scheme was designed according to the automatic ML principle, and the double-loop scheme with importance sampling was used to optimize the data augmentation and loss function types. Tianrui Liu [137] et al. proposed a pedestrian detection method with a coupling network. The deformable occlusion processing sub-network adaptively adjusts the relative position of the body parts of the pooling grid by using the deformable region of interest pooling, so as to better adapt to the occluded pedestrians. Yifan Jiao [138] et al. proposed the Pose Embedded Pedestrian Detection Network (PEN), which generates a large number of candidate boxes and corresponding confidence scores through the region proposal network, and the pedestrian recognition network filters candidate boxes through pedestrian posture information and visual information. In 2022, Xiaolin Song [139] et al. proposed the Progressive Refinement Network (PRNet). The proposed confident perception anchor calibration method can adaptively initialize anchor points in the presence of occlusion. Wei Wei's method has low latency and fast response, and can be deployed in embedded devices. Chunluan Zhou used discriminative feature transformation for pedestrian detection and Yifan Jiao improved detection performance by embedding human pose information, both of which provided new ideas for others. Yuting Xu's method has some limitations when detecting small-scale pedestrians. Yi Tang's method can be well applied to other detectors. Xiaolin Song's method is effective for different occlusions in different fields.

6. Conclusion

Pedestrian detection has always been a research hotspot in the field of computer vision. From manually designed feature descriptors combined with machine learning to the now widely used deep learning methods, pedestrian detection has made great progress in detection accuracy and detection speed, but there are still many problems and challenges waiting for us. This paper discusses the pedestrian detection methods proposed in the past, and introduces some newly proposed pedestrian detection algorithms after the 2019 years in terms of handcrafted features and deep features. In addition, dozens of detection methods proposed to solve the scale problem and occlusion problem in pedestrian detection are discussed. In the past 4 years, the detection accuracy and detection speed of pedestrian detection have been significantly improved. Researchers have done a lot of work. Aiming at the scale problem, some researchers use context information and attention mechanism to improve multi-scale pedestrian detection, some researchers improve the robustness of pedestrian detection for pedestrians at different scales by fusing multi-scale features, and some researchers fuse deep features and shallow features to enrich the semantic information and spatial features of features. The MR-2 index of MPAF-Net network with better performance on the Small, Medium and Large subsets of Cityperson dataset reaches 14.0%, 3.3% and 5.2%. For the occlusion problem, some researchers use pedestrian detection based on body parts to reduce the impact of occlusion. Some researchers use center point detection to transform pedestrian detection into high-level semantic feature detection, which effectively improves the detection accuracy in the presence of occlusion. Some researchers enhance feature representation through attention mechanism, context information and other methods to significantly improve the detection performance of occluded pedestrians. Some researchers introduce saliency map, edge perception and other methods to cope with occlusion. The MR-2 index of SA-DPM network with better performance reaches 30.18% in the severely occluded subset of Caltech dataset. However, there is still a big gap between the detection ability of pedestrian detection and human ability. In order to better apply pedestrian detection to unmanned driving, intelligent video surveillance and other fields, there is still a lot of work to be done:

Scale problem: Despite the rapid development of pedestrian detection technology, the detection accuracy and detection speed have been greatly improved, but the scale problem of pedestrians is still a big challenge. Constructing multi-scale feature maps is a common method to deal with the scale problem of pedestrian detection. Deep feature maps often have rich semantic information, while shallow feature maps contain accurate positioning information, and fusing multi-scale feature maps can obtain more effective information. Exploring a better fusion method of multi-scale feature maps and designing a more reasonable and efficient pyramid structure are important ways to solve the multi-scale problem of pedestrian detection. At the same time, improving the representation ability of features by fusing features is also an effective way to deal with the scale problem. In the future, multiple attention mechanisms such as competitive attention can be explored for feature fusion. Multi-scale representation is an important solution to the problem of pedestrian detection scale. This method improves the detection performance while inevitably increasing the amount of calculation. How to balance the accuracy and calculation needs further exploration, and designing a more reasonable FPN structure is also worth studying. Context information improves the performance of pedestrian detection by enriching feature information, but not all context information can play a positive role in improving detection performance. How to select effective context information and be able to better insert it in the network is a direction worthy of research. Adding pre-processing and post-processing modules to the pedestrian detection framework can significantly improve the detection performance, such as Mosaic, flip, etc. Exploring more efficient pre-processing and post-processing methods is also an important way to improve the performance of the detector. Multi-task joint learning is also a good research direction, which can obtain richer visual information through multiple visual tasks. Therefore, how to effectively use multi-task joint learning to improve the accuracy of multi-scale pedestrian detection is a good research direction in the future. The method based on super-resolution is a new direction to solve the pedestrian scaling problem in recent years, and its biggest difficulty is the training of GAN network. How to balance the generator and discriminator is a direction to be explored in the future. Data augmentation strategies can reduce the negative impact of imbalanced data to improve detection performance. How to design efficient data augmentation strategies is also worth studying.

Occlusion problem: Pedestrian detection technology develops very rapidly, and has a high detection accuracy under normal circumstances. However, when there is occlusion in the image, the detection accuracy is significantly reduced, especially in severe occlusion, pedestrian detection can not maintain good robustness, which is still a huge challenge for us. Deep convolutional networks can extract higher dimensional pedestrian information, but some shallow information will be lost with the increase of dimension. The fusion of feature information of different dimensions can effectively improve the adverse impact of occlusion. Feature fusion has always been a common method to deal with occlusion. The fusion feature has rich information and strong representation ability. Exploring how to design an efficient feature fusion network is also an effective method to solve the occlusion problem. Pedestrian detection is only one task of computer vision, which also includes semantic segmentation and human pose estimation. Using other vision tasks to assist pedestrian detection is a way to improve the performance of occluded pedestrian detection, and the integration of multiple vision tasks is also an important research direction. In the future, the accuracy of occlusion pedestrian detection can be improved by exploring the degree of help of different body parts of pedestrians for occlusion detection to select the most appropriate body part for detection. At the same time, different training strategies will also affect the detection effect, and it is very important to explore efficient training strategies. Context information can make up for the lack of information in the invisible part, and how to design a more efficient context information extraction network can be explored in the future. Center-based pedestrian detection can effectively improve the adverse effects caused by occlusion. In the future, whether the body part-based pedestrian

detection and center-based pedestrian detection can be combined, and the pedestrian body part is used as the center point to improve the robustness of occlusion pedestrian detection. At the same time, the ability of data augmentation, edge perception and other methods to improve occlusion can also be explored in the future, and the detection accuracy of occluded pedestrians can be improved by combining these methods.

Funding

National Natural Science Foundation of China (52274160, 52074305, 51874300); National Natural Science Foundation of China-Shanxi Joint Fund for Coal-Based Low-Carbon Technology (U1510115).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Wang X, Liu M, Raychaudhuri DS, et al. Learning person Re-identification models from videos with weak supervision. IEEE Trans Image Process 2021;30:3017–28.
[2] Wang H, Du H, Zhao Y, et al. A comprehensive overview of person Re-identification approaches. IEEE Access 2020;8:45556–83.
[3] Xu X, Li X, Zhao H, et al. A real-time, continuous pedestrian tracking and positioning method with multiple coordinated overhead-view cameras. Measurement 2021:178.
[4] Dimitrievski M, Veelaert P, Philips W. Behavioral pedestrian tracking using a camera and LiDAR sensors on a moving vehicle. Sensors 2019;19(2).
[5] Wang CX, Cai SF, An G, et al. GraphTCN: spatio-temporal interaction modeling for human trajectory prediction. In: Proceedings of the IEEE Winter conference on applications of computer vision (WACV). Electr Network, F Jan 05-09; 2021 [C].
[6] Xue H, Huynh DQ, Reynolds M. PoPPL: pedestrian trajectory prediction by LSTM with automatic route class clustering. IEEE Transact Neural Networks Learn Syst 2021;32(1):77–90.
[7] Mhalla A, Chateau T, Gazzah S, et al. An embedded computer-vision system for multi-object detection in traffic surveillance. IEEE Trans Intell Transport Syst 2019;20(11):4006–18.
[8] Yang L, Hu G, Song Y, et al. Intelligent video analysis: a Pedestrian trajectory extraction method for the whole indoor space without blind areas. Computer Vision And Image Understanding 2020:196.
[9] Du Y, Hetherington NJ, Oon CL, et al. Group surfing: a pedestrian-based approach to sidewalk robot navigation. In: Proceedings of the IEEE international conference on robotics and automation (ICRA). Montreal: CANADA, F May 20-24; 2019 [C].
[10] Li T, Ma Y, Shen H, et al. FPGA implementation of real-time pedestrian detection using normalization-based validation of adaptive features clustering. IEEE Trans Veh Technol 2020;69(9):9330–41.
[11] Robin C, Lacroix S. Multi-robot target detection and tracking: taxonomy and survey. Aut Robots 2016;40(4):729–60.
[12] Qiao Z, Zhao L, Gu L, et al. Research on abnormal pedestrian trajectory detection of dynamic crowds in public scenarios. IEEE Sensor J 2021;21(20):23046–54.
[13] Viola P, Jones MJ, Snow D. Detecting pedestrians using patterns of motion and appearance. Int J Comput Vis 2005;63(2):153–61.
[14] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition, CVPR 2005, June 20, 2005 - June 25, 2005. San Diego, CA, United states, F: IEEE Computer Society; 2005 [C].
[15] Dollar P, Tu Z, Perona P, et al. Integral channel features. In: Proceedings of the 2009 20th British machine vision conference, BMVC 2009, september 7, 2009 - september 10, 2009,. London, United kingdom, F: British Machine Vision Association, BMVA; 2009 [C].
[16] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. In: Proceedings of the 29th IEEE conference on computer vision and pattern recognition, CVPR 2016, June 26, 2016 - July 1, 2016, Las Vegas, NV, United States, F. IEEE Computer Society; 2016 [C].
[17] Redmon J, Farhadi A. Ieee. YOLO9000: better, faster, stronger; proceedings of the 30th IEEE/CVF conference on computer vision and Pattern recognition (CVPR), Honolulu, HI, F Jul 21-26. 2017 [C].
[18] Redmon J, Farhadi A. YOLOv3: an incremental improvement. 2018 [M]. arXiv.
[19] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. In: Proceedings of the 14th European conference on computer vision, ECCV 2016, October 8, 2016 - October 16, 2016, Amsterdam, Netherlands, F. Springer Verlag; 2016 [C].
[20] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the 27th IEEE conference on computer vision and pattern recognition, CVPR 2014, June 23, 2014 - June 28, 2014, columbus, OH, United States, F. IEEE Computer Society; 2014 [C].
[21] Girshick R. Fast R-CNN. In: Proceedings of the 15th IEEE international conference on computer vision, ICCV 2015, December 11, 2015 - December 18, 2015, santiago, Chile, F. Institute of Electrical and Electronics Engineers Inc; 2015 [C].
[22] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017;39(6):1137–49.
[23] Cai Z, Fan Q, Feris RS, et al. A unified multi-scale deep convolutional neural network for fast object detection. In: Proceedings of the 21st ACM conference on computer and communications security, CCS 2014, November 3, 2014 - November 7, 2014, scottsdale, AZ, United States, F. Springer Verlag; 2016 [C].
[24] Bochkovskiy A, Wang C-Y, Liao H-YM. YOLOv4: optimal speed and accuracy of object detection. 2020 [M]. arXiv.
[25] Liu W, Hasan I, Liao S. Center and scale prediction: anchor-free approach for pedestrian and face detection. 2019 [M]. arXiv.
[26] Li J, Liang X, Shen S, et al. Scale-aware fast R-CNN for pedestrian detection. IEEE Trans Multimed 2018;20(4):985–96.
[27] Dollar P, Wojek C, Schiele B, et al. Pedestrian detection: a benchmark; proceedings of the 2009 IEEE conference on computer vision and Pattern recognition, CVPR 2009, June 20, 2009 - June 25, 2009. Miami, FL, United states, F: IEEE Computer Society; 2009 [C].
[28] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the KITTI vision benchmark suite. In: Proceedings of the 2012 IEEE conference on computer vision and pattern recognition, CVPR 2012, June 16, 2012 - June 21, 2012, providence, RI, United States, F. IEEE Computer Society; 2012 [C].
[29] Zhang S, Benenson R, Schiele B. CityPersons: a diverse dataset for pedestrian detection. In: Proceedings of the 30th IEEE conference on computer vision and pattern recognition, CVPR 2017, July 21, 2017 - July 26, 2017, Honolulu, HI, United States, F. Institute of Electrical and Electronics Engineers Inc; 2017 [C].
[30] Li W, Ruan M, Guo X, et al. A novel architecture of pedestrian detection. In: Proceedings of the IEEE int conf on parallel and Distributed processing with applications, big data and cloud computing, sustainable computing and communications, social computing and networking (ISPA/BDCloud/SocialCom/SustainCom), Xiamen, PEOPLES R China, F Dec 16-18; 2019 [C].
[31] Kumar K, Mishra RK. A robust mRMR based pedestrian detection approach using shape descriptor. Trait Du Signal 2019;36(1):79–85.
[32] Yang M, Qian Y, Xue L, et al. G2P: a new descriptor for pedestrian detection. Neural Comput Appl 2020;32(9):4665–74.
[33] Lian G. Pedestrian detection using quaternion gradient based weber local descriptor. IEEE Access 2021;9:43675–83.
[34] Pfeifer L. Shearlet features for pedestrian detection. J Math Imag Vis 2019;61(3):292–309.
[35] Jiang Y, Tong G, Yin H, et al. A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters. IEEE Access 2019;7:118310–21.
[36] Xie Z, Yang R, Guan W, et al. A novel descriptor for pedestrian detection based on multi-layer feature fusion. In: Proceedings of the IEEE international conference on real-time computing and robotics (IEEE-RCAR), electr network, F sep 28-29; 2020 [C].
[37] Kumar K, Mishra RK. A heuristic SVM based pedestrian detection approach employing shape and texture descriptors. Multimed Tool Appl 2020;79(29–30):21389–408.
[38] Zhou H, Yu G. Research on pedestrian detection technology based on the SVM classifier trained by HOG and LTP features. Future Generation Computer Systems-the International Journal Of Escience 2021;125:604–15.
[39] Liu D, Zang K, Shen J. A shallow-deep feature fusion method for pedestrian detection. Applied Sciences-Basel 2021;11(19).
[40] Liu W, Liao S, Ren W, et al. High-level semantic feature detection: a new perspective for pedestrian detection. In: Proceedings of the 32nd IEEE/CVF conference on computer vision and pattern recognition (CVPR), long Beach, CA, F Jun 16-20; 2019 [C].
[41] Zhang T, Cao Y, Zhang L, et al. Efficient feature fusion network based on center and scale prediction for pedestrian detection. Visual Computer; 2022.
[42] Cao J, Song C, Peng S, et al. Pedestrian detection algorithm for intelligent vehicles in complex scenarios. Sensors 2020;20(13).
[43] Lv H, Yan H, Liu K, et al. YOLOv5-AC: attention mechanism-based lightweight YOLOv5 for track pedestrian detection. Sensors 2022;22(15).
[44] Liu W, Liao S, Hu W. Efficient single-stage pedestrian detector by asymptotic localization fitting and multi-scale context encoding. IEEE Trans Image Process 2020;29:1413–25.
[45] Saeidi M, Ahmadi A. High-performance and deep pedestrian detection based on estimation of different parts. J Supercomput 2021;77(2):2033–68.
[46] Murthy CB, Hashmi MF, Keskar AG. Optimized MobileNet plus SSD: a real-time pedestrian detection on a low-end edge device. International Journal Of Multimedia Information Retrieval 2021;10(3):171–84.
[47] Zhang Y, He H, Li J, et al. Variational pedestrian detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). Electr Network, F Jun 19-25; 2021 [C].

[48] Brazil G, Liu X, Soc IC. Pedestrian detection with autoregressive network phases. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), long Beach, CA, F Jun 16-20; 2019 [C].
[49] Fu Z, Chen Y, Jiang R. See extensively while focusing on the core area for pedestrian detection. IEEE Access 2019;7:27017–25.
[50] Ren J, Han J. A new multi-scale pedestrian detection algorithm in traffic environment. Journal Of Electrical Engineering & Technology 2021;16(2):1151–61.
[51] Tesema FB, Wu H, Chen M, et al. Hybrid channel based pedestrian detection. Neurocomputing 2020;389:1–8.
[52] Ruihaoyin IEEE. MULTI-RESOLUTION generative adversarial networks for tiny-scale pedestrian detection. In: Proceedings of the 26th IEEE international conference on image processing (ICIP), Taipei, TAIWAN, F sep 22-25; 2019 [C].
[53] Yu P, Zhao Y, Zhang J, et al. Pedestrian detection using multi-channel visual feature fusion by learning deep quality model. J Vis Commun Image Represent 2019:63.
[54] Lin Z, Pei W, Chen F, et al. Pedestrian detection by exemplar-guided contrastive learning. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society; 2022.
[55] Xie H, Zheng W, Shin H. Occluded pedestrian detection techniques by deformable attention-guided network (DAGN). Applied Sciences-Basel 2021;11(13).
[56] Papageorgiou C, Poggio T. A trainable system for object detection. Int J Comput Vis 2000;38(1):15–33.
[57] Hwang S, Park J, Kim N, et al. Multispectral pedestrian detection: benchmark dataset and baseline. In: Proceedings of the IEEE conference on computer vision and pattern recognition, CVPR 2015, June 7, 2015 - June 12, 2015, Boston, MA, United States, F. IEEE Computer Society; 2015 [C].
[58] Braun M, Krebs S, Flohr F, et al. EuroCity persons: a novel benchmark for person detection in traffic scenes. IEEE Trans Pattern Anal Mach Intell 2019;41(8):1844–61.
[59] Enzweiler M, Gavrila DM. Monocular pedestrian detection: survey and experiments, F. IEEE Computer Society; 2009 [C].
[60] Overett G, Petersson L, Brewer N, et al. A new pedestrian dataset for supervised learning; proceedings of the 2008 IEEE Intelligent Vehicles Symposium, IV, June 4, 2008 - June 6, 2008. Eindhoven, Netherlands, F: Institute of Electrical and Electronics Engineers Inc; 2008 [C].
[61] Ess A, Leibe B, Van Gool L. Depth and appearance for mobile scene analysis. In: Proceedings of the 2007 IEEE 11th international conference on computer vision, ICCV, October 14, 2007 - October 21, 2007, Rio de Janeiro, Brazil, F. Institute of Electrical and Electronics Engineers Inc; 2007 [C].
[62] Shao S, Zhao Z, Li B, et al. CrowdHuman: a benchmark for detecting human in a crowd. 2018 [M]. arXiv.
[63] Zhang S, Xie Y, Wan J, et al. WiderPerson: a diverse dataset for dense pedestrian detection in the wild. IEEE Trans Multimed 2020;22(2):380–93.
[64] Xie H, Chen Y, Shin H. Context-aware pedestrian detection especially for small-sized instances with Deconvolution Integrated Faster RCNN (DIF R-CNN). Appl Intell 2019;49(3):1200–11.
[65] Hu J, Jin L, Gao S, et al. FPN plus plus : a simple baseline for pedestrian detection. In: Proceedings of the IEEE international conference on multimedia and expo (ICME), Shanghai, PEOPLES R China, F Jul 08-12; 2019 [C].
[66] Zhang C, Kim J. Multi-scale pedestrian detection using skip pooling and recurrent convolution. Multimed Tool Appl 2019;78(2):1719–36.
[67] Yang X, Liu Q. Scale-sensitive feature reassembly network for pedestrian detection. Sensors 2021;21(12).
[68] Chen Z, Zhang L, Khattak AM, et al. Deep feature fusion by competitive attention for pedestrian detection. IEEE Access 2019;7:21981–9.
[69] Lin C, Lu J, Wang G, et al. Graininess-aware deep feature learning for robust pedestrian detection. IEEE Trans Image Process 2020;29:3820–34.
[70] Feng TT, Ge HY. Pedestrian detection based on attention mechanism and feature enhancement with SSD; proceedings of the 5th international conference on communication, image and signal processing (CCISP), chengdu, PEOPLES R China. F Nov 2020;13–15 [C].
[71] Ma J, Wan H, Wang J, et al. An improved one-stage pedestrian detection method based on multi-scale attention feature extraction. Journal of Real-Time Image Processing 2021;18(6):1965–78.
[72] Zhang X, Cao S, Chen C. Scale-aware hierarchical detection network for pedestrian detection. IEEE Access 2020;8:94429–39.
[73] Zhao Y, Shi F, Zhao M, et al. Detecting small scale pedestrians and anthropomorphic negative samples based on light-field imaging. IEEE Access 2020;8:105082–93.
[74] Cao Z, Yang H, Xu W, et al. Multiscale anchor-free region proposal network for pedestrian detection. Wireless Commun Mobile Comput 2021;2021.
[75] Hsu W-Y, Lin W-Y. Adaptive fusion of multi-scale YOLO for pedestrian detection. IEEE Access 2021;9:110063–73.
[76] Ding L, Wang Y, Laganiere R, et al. Learning efficient single stage pedestrian detection by squeeze-and-excitation network. Neural Comput Appl 2021;33(23):16697–712.
[77] Li Q, Qiang H, Li J. Conditional random fields as message passing mechanism in anchor-free network for multi-scale pedestrian detection. Inf Sci 2021;550:1–12.
[78] Yan C, Zhang H, Li X, et al. R-SSD: refined single shot multibox detector for pedestrian detection. Appl Intell 2022;52(9):10430–47.
[79] Ren J, Niu C, Han J. An IF-RCNN algorithm for pedestrian detection in pedestrian tunnels. IEEE Access 2020;8:165335–43.
[80] Han B, Wang Y, Yang Z, et al. Small-scale pedestrian detection based on deep neural network. IEEE Trans Intell Transport Syst 2020;21(7):3046–55.
[81] Murthy CB, Hashmi MF, Muhammad G, et al. YOLOv2PD: an efficient pedestrian detection algorithm using improved YOLOv2 model. Cmc-Computers Materials & Continua 2021;69(3):3015–31.
[82] Xia H, Wan H, Ou J, et al. MAFA-net: pedestrian detection network based on multi-scale attention feature aggregation. Appl Intell 2022;52(7):7686–99.
[83] He Y, He N, Zhang R, et al. Multi-scale feature balance enhancement network for pedestrian detection. Multimed Syst 2022;28(3):1135–45.
[84] Yang P, Zhang G, Wang L, et al. A part-aware multi-scale fully convolutional network for pedestrian detection. IEEE Trans Intell Transport Syst 2021;22(2):1125–37.
[85] Zhou C, Wu M, Lam S-K. Enhanced multi-task learning architecture for detecting pedestrian at far distance. IEEE Trans Intell Transport Syst 2022;23(9):15588–604.
[86] Pang Y, Cao J, Wang J, et al. JCS-net: joint classification and super-resolution network for small-scale pedestrian detection in surveillance images. IEEE Trans Inf Forensics Secur 2019;14(12):3322–31.
[87] Wu J, Zhou C, Zhang Q, et al. Self-mimic learning for small-scale pedestrian detection. In: Proceedings of the 28th ACM international conference on multimedia (MM), electr network, F Oct 12-16; 2020 [C].
[88] Zhang H, Hu Z, Hao R. Joint information fusion and multi-scale network model for pedestrian detection. Vis Comput 2021;37(8):2433–42.
[89] Murthy CB, Hashmi MF, Keskar AG. EfficientLiteDet: a real-time pedestrian and vehicle detection algorithm. Mach Vis Appl 2022;33(3).
[90] Wang J, Li H, Yin S, et al. Research on improved pedestrian detection algorithm based on convolutional neural network. In: Proceedings of the IEEE int congr on cybermat/12th IEEE int conf on cyber, phys and social comp (CPSCom)/15th IEEE int conf on green computing and communications (GreenCom)/12th IEEE int conf on internet of things (iThings)/5th IEEE int conf on smart data, Atlanta, GA, F Jul 14-17; 2019 [C].
[91] Zhang S, Yang X, Liu Y, et al. Asymmetric multi-stage CNNs for small-scale pedestrian detection. Neurocomputing 2020;409:12–26.
[92] Wang X, Liang C, Chen C, et al. (SD)-D-3: scalable pedestrian detection via score scale surface discrimination. IEEE Trans Circ Syst Video Technol 2020;30(10):3332–44.
[93] Lin C-Y, Xie H-X, Zheng H. PedJointNet: joint head-shoulder and full body deep network for pedestrian detection. IEEE Access 2019;7:47687–97.
[94] Gu J, Lan C, Chen W, et al. Joint pedestrian and body Part Detection via semantic relationship learning. Applied Sciences-Basel 2019;9(4).
[95] Zhou C, Yuan J. Multi-label learning of part detectors for occluded pedestrian detection. Pattern Recogn 2019;86:99–111.
[96] Islam MM, Newaz AAR, Gokaraju B, et al. Pedestrian detection for autonomous cars: occlusion handling by classifying body parts. In: Proceedings of the IEEE international conference on systems, man, and cybernetics (SMC). Electr Network, F Oct 11-14; 2020 [C].
[97] Zhang S, Chen D, Yang J, et al. Guided attention in CNNs for occluded pedestrian detection and Re-identification. Int J Comput Vis 2021;129(6):1875–92.
[98] Ding J, Liu T, Zhao Y, et al. HAPNet: a head-aware pedestrian detection network associated with the affinity field. Sci China Inf Sci 2022;65(6).
[99] Abdelmutalab A, Wang C. Pedestrian detection using MB-CSP model and boosted identity aware non-maximum suppression. IEEE Trans Intell Transport Syst 2022;23(12):24454–63.
[100] Fei C, Liu B, Chen Z, et al. Learning pixel-level and instance-level context-aware features for pedestrian detection in crowds. IEEE Access 2019;7:94944–53.
[101] Susutti W, Lursinsap C, Sophatsathit P. Pedestrian detection by using weighted channel features with hierarchical region reduction. Journal Of Signal Processing Systems for Signal Image And Video Technology 2019;91(6):587–608.
[102] Li Z, Chen Z, Wu QMJ, et al. Pedestrian detection via deep segmentation and context network. Neural Comput Appl 2020;32(10):5845–57.
[103] Zhai S, Dong S, Shang D, et al. An improved faster R-CNN pedestrian detection algorithm based on feature fusion and context analysis. IEEE Access 2020;8:138117–28.
[104] Jiang H, Liao S, Li J, et al. Urban scene based semantical modulation for pedestrian detection. Neurocomputing 2022;474:1–12.
[105] Liu Z, Song X, Feng Z, et al. Global context-aware feature extraction and visible feature enhancement for occlusion-invariant pedestrian detection in crowded scenes. Neural Process Lett 2023;55(1):803–17.
[106] Liu S, Guo H, Hu J-G, et al. A novel data augmentation scheme for pedestrian detection with attribute preserving GAN. Neurocomputing 2020;401:123–32.
[107] Xie J, Pang Y, Khan MH, et al. Mask-guided attention network and occlusion-sensitive hard example mining for occluded pedestrian detection. IEEE Trans Image Process 2021;30:3872–84.
[108] Jin Y, Zhang Y, Cen Y, et al. Pedestrian detection with super-resolution reconstruction for low-quality image. Pattern Recogn 2021:115.
[109] Zhang Y, Bai Y, Ding M, et al. KGSNet: key-point-guided super-resolution network for pedestrian detection in the wild. IEEE Transact Neural Networks Learn Syst 2021;32(5):2251–65.
[110] Zhang T, Han Z, Xu H, et al. CircleNet: reciprocating feature adaptation for robust pedestrian detection. IEEE Trans Intell Transport Syst 2020;21(11):4593–604.
[111] Ruan B, Zhang C. Occluded pedestrian detection combined with semantic features. IET Image Process 2021;15(10):2292–300.
[112] Tan Y, Yao H, Li H, et al. PRF-ped: multi-scale pedestrian detector with prior-based receptive field. In: Proceedings of the 25th international conference on pattern recognition (ICPR). Electr Network, F Jan 10-15; 2021 [C].
[113] Yang G, Wang Z, Zhuang S, et al. PFF-CB: multiscale occlusion pedestrian detection method based on PFF and CBAM. Comput Intell Neurosci 2022;2022.

[114] Wang J, Zhao C, Huo Z, et al. High quality proposal feature generation for crowded pedestrian detection. Pattern Recogn 2022:128.
[115] Qin Y, Qian Y, Wei H, et al. FE-CSP: a fast and efficient pedestrian detector with center and scale prediction. J Supercomput 2023;79(4):4084–104.
[116] Yin R, Zhang R, Zhao W, et al. DA-net: pedestrian detection using dense connected block and attention modules. IEEE Access 2020;8:153929–40.
[117] Zou T, Yang S, Zhang Y, et al. Attention guided neural network models for occluded pedestrian detection. Pattern Recogn Lett 2020;131:91–7.
[118] Shao X, Wang Q, Yang W, et al. Multi-scale feature pyramid network: a heavily occluded pedestrian detection network based on ResNet. Sensors 2021;21(5).
[119] Xie J, Pang Y, Cholakkal H, et al. PSC-Net: learning part spatial co-occurrence for occluded pedestrian detection. Sci China Inf Sci 2021;64(2).
[120] He Y, Zhu C, Yin X-C. Occluded pedestrian detection via distribution-based mutual-supervised feature learning. IEEE Trans Intell Transport Syst 2022;23(8):10514–29.
[121] Luo Y, Zhang C, Lin W, et al. Sequential attention-based distinct Part Modeling for balanced pedestrian detection. IEEE Trans Intell Transport Syst 2022;23(9):15644–54.
[122] Zhang T, Ye Q, Zhang B, et al. Feature calibration network for occluded pedestrian detection. IEEE Trans Intell Transport Syst 2022;23(5):4151–63.
[123] Wang Y, Han C, Yao G, et al. MAPD: an improved multi-attribute pedestrian detection in a crowd. Neurocomputing 2021;432:101–10.
[124] Zhang J, Lin L, Zhu J, et al. Attribute-aware pedestrian detection in a crowd. IEEE Trans Multimed 2021;23:3085–97.
[125] Li Q, Su Y, Gao Y, et al. OAF-net: an occlusion-aware anchor-free network for pedestrian detection in a crowd. IEEE Trans Intell Transport Syst 2022;23(11):21291–300.
[126] Zhang Y, Wang H, Liu Y, et al. Region resolution learning and region segmentation learning with overall and body Part Perception for pedestrian detection. Electronics 2022;11(6).
[127] Lin X, Zhao C, Zhang C, et al. Self-attention-guided scale-refined detector for pedestrian detection. Complex & Intelligent Systems; 2022.
[128] Li Q, Bi Y, Cai R, et al. Occluded pedestrian detection through bi-center prediction in anchor-free network. Neurocomputing 2022;507:199–207.
[129] Wang Z, Lin W, Cheng L, et al. Multi-branch detection network based on trigger attention for pedestrian detection under occlusion. Applied Intelligence; 2022.
[130] Yun I, Jung C, Wang X, et al. Part-level convolutional neural networks for pedestrian detection using saliency and boundary box AlignmentD. IEEE Access 2019;7:23027–37.
[131] Wei W, Cheng L, Xia Y, et al. Occluded pedestrian detection based on depth vision significance in biomimetic binocular. IEEE Sensor J 2019;19(23):11469–74.
[132] Zhou C, Yang M, Yuan J, et al. Discriminative feature transformation for occluded pedestrian detection. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV), seoul, South Korea, F Oct 27-Nov 02; 2019 [C].
[133] Xu L, Yan S, Chen X, et al. Motion recognition algorithm based on deep edge-aware pyramid pooling network in human-computer interaction. IEEE Access 2019;7:163806–13.
[134] Xu Y, Zhou X, Liu P, et al. Rapid pedestrian detection based on deep omega-shape features with partial occlusion handing. Neural Process Lett 2019;49(3):923–37.
[135] Lin C, Lu J, Zhou J. Multi-grained deep feature learning for robust pedestrian detection. IEEE Trans Circ Syst Video Technol 2019;29(12):3608–21.
[136] Tang Y, Li B, Liu M, et al. AutoPedestrian: an automatic data augmentation and loss function search scheme for pedestrian detection. IEEE Trans Image Process 2021;30:8483–96.
[137] Liu T, Luo W, Ma L, et al. Coupled network for robust pedestrian detection with gated multi-layer feature extraction and deformable occlusion handling. IEEE Trans Image Process 2021;30:754–66.
[138] Jiao Y, Yao H, Xu C. PEN: pose-embedding network for pedestrian detection. IEEE Trans Circ Syst Video Technol 2021;31(3):1150–62.
[139] Song X, Chen B, Li P, et al. PRNet plus plus : learning towards generalized occluded pedestrian detection via progressive refinement network. Neurocomputing 2022;482:98–115.