Research on Pedestrian Detection Algorithm Based on Image

Boyi Wang
College of Automation, Beijing University of Science and Technology, Beijing, China
*Corresponding author e-mail: huntercet119@qq.com

Abstract. Pedestrian detection is an important application of target detection. It is an important foundation in many fields, such as unmanned vehicle driving, intelligent monitoring system, pedestrian analysis and robot development. Despite a lot of research, pedestrian detection methods still have many problems to be solved urgently. In order to sort out the existing pedestrian detection algorithms, this paper will summarize and analyze the advantages and disadvantages of these algorithms. The two methods of traditional algorithms and deep learning algorithms are used to discuss pedestrian detection, analyze and compare the results of the two types of algorithms, and finally target pedestrians. The detection algorithm is expected.

1. Introduction
Pedestrian detection, as an important part of intelligent transportation system, has broad prospects in scientific research and engineering applications. It is a key difficulty and hot spot in computer vision, an important application of target detection, and an important foundation in many fields such as unmanned vehicle driving, intelligent monitoring, pedestrian analysis and intelligent robots.

In recent years, pedestrian detection technology is becoming more and more perfect, the detection accuracy and speed are becoming more and more practical, and pedestrian database is also becoming large-scale. However, due to the diversity of pedestrian posture, size, clothing and deformation, as well as the different illumination conditions, background complexity and occlusion, there are still many problems to be solved in pedestrian detection. This paper will introduce, analyze and compare several mainstream pedestrian detection algorithms from the traditional methods and deep learning methods.

2. Pedestrian Detection Based on Traditional Method
Traditional pedestrian detection methods extract the main features describing pedestrians based on artificially designed feature extractors, then use these features to train classifiers to distinguish pedestrians and other things, and finally achieve the purpose of pedestrian detection. The specific process is shown in Figure 1:

![Fig.1. The process of pedestrians detection](image)

3. Pedestrian Detection Based on HOG Features and SVM
In 2005, Dalal and Triggs proposed a landmark algorithm for pedestrian detection—a detection algorithm based on HOG features and SVM classifiers [1].
3.1. HOG feature

The HOG features (Histogram of Oriented Gradients) are one of the most widely used pedestrian features. Because gradient can well describe the representation and shape of local objects (i.e. edges) in images, it is possible to construct features by calculating and statistic gradient direction histograms of local areas in images.

3.2. SVM classifier

The SVM algorithm (support vector machine) is a two-class algorithm proposed by Vapnik and divided into linear and nonlinear SVM algorithms. The idea of the SVM algorithm is to maximize the distance between the support vector and the hyperplane to maximize the reliability of classification.

For linear inseparable samples, the use of linear classification method can not avoid errors in classification. In this case, a penalty factor is introduced to maximize the interval when certain errors are allowed. For nonlinear problems, the data is first mapped to a linear problem in a high-dimensional space. Common kernel functions are as follows:

- Gaussian core:
  \[ K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \]

- Polynomial Kernel:
  \[ K(x, y) = (\langle x, y \rangle + c)^d \]

- Perceptron core:
  \[ K(x, y) = \tanh(\rho_1 \langle x, y \rangle + \rho_2) \]

The SVM training flow chart is shown in Figure 2:

Fig.2. An overview of training SVM chain
3.3. Algorithm effect
Dalal et al. use a 64 x 128 fixed size window to detect pedestrians, so in order to detect pedestrians of different sizes in the image, it is necessary to change the size of the target image. This method requires storing images of different scales, increasing the complexity of the space.

This algorithm achieves an accuracy of about 90% at a false detection rate of one in ten thousand in the pedestrian dataset INRIA created by Dalal et al., with high precision. However, it is not suitable for real-time detection because of its high dimension and slow computation speed. This paper presents an improvement of HOG algorithm [2,3,4].

3.4. Pedestrian Detection Based on HOG Features and AdaBoost
Since the calculation of the HOG+SVM algorithm is too large and the calculation speed is too slow, later researchers replaced the SVM with the AdaBoost cascade classifier.

3.5. Training AdaBoost Classifier
Boosting is an ensemble learning method to improve generalization by combining multiple weak classifiers to form strong classifiers. AdaBoost is the most popular and famous iteration algorithm in Boosting family. AdaBoost mainly made two adjustments to Boosting: how to train multiple weak classifiers, and how to linearly combine the trained weak classifiers into strong classifiers. AdaBoost linearly combines all weak classifiers according to the weight of the weak classifier, and "weighted voting" gives the final classification result. Figure 3 illustrates the idea of the AdaBoost algorithm.

![Fig.3. Schematic diagram of AdaBoost](image)

The classification accuracy of strong classifiers trained by AdaBoost algorithm depends on the quality of all weak classifiers and can be adapted according to the feedback of weak classifiers. Adjustment, high efficiency and robustness. However, due to the exponential growth of sample weights, which are difficult to classify in the training process, AdaBoost is easily disturbed by noise. Viola and Jones proposed Cascade AdaBoost classifiersto improve the problem of AdaBoost [5][6]. The structure is shown in Figure 4.
Because the former classifiers prefer the weak classifier during training, the best weak classifiers will be integrated. The required weak classifiers are small, the calculation is small and the speed is fast, and most negative samples can be excluded. The latter stages of classifiers use additional calculations to discriminate the remaining candidate areas.

3.6. Algorithm effect
The strategy based on the cascaded AdaBoost classifier greatly improves the final detection speed and basically realizes real-time monitoring. The results in show that the algorithm is 70 times faster than the HOG+SVM algorithm [7]. The paper proves that the algorithm is robust to pedestrian attitude changes, complex background and illumination changes [8,9].

3.7. Pedestrian detection based on ICF features and AdaBoost
The HOG feature is sensitive to noise due to the nature of the gradient, and only describes the edge and shape information of the object, so it does not reflect the apparent information of the object, and it is difficult to detect the occluded pedestrian. To solve these problems, Piotr et al. proposed Integral Channel Features (ICF). The idea is to quickly calculate some channel features of an image, such as histogram, Haar-like feature, intensity and color, by performing various linear and non-linear transformations of the input image [10].

3.8. ICF features
The optimal integration channel features obtained in the experiment include 10 channels, including gradient histograms in 6 directions, 3 LUV color spaces and 1 gradient amplitude, as shown in Figure 5:

These channels can be calculated quickly and efficiently, and reflect image information from different angles:

---

Fig.4. Schematic depiction of the cascade classifiers [6]

Fig.5. Example image and computed channels [10]

Fig.6. Rough visualization of spatial support of trained classifiers for all channels jointly(left) and separately for each channel type [10]
3.9. Algorithm effect
The article uses soft cascade AdaBoost to classify samples. Unlike Dalal et al., which uses fixed-size windows for detection, the author trained several commonly used size classifiers to interpolate with the prediction results of these typical size classifiers to approximate pedestrians of other scales, thus solving the accuracy of the image scaling. The problem of low, extra storage space is more efficient and accurate.

This paper proves that ICF performs better than HOG in pedestrian detection through a large number of experimental data, and has the advantages of more accurate spatial positioning and less parameter setting in the detection process. This method realizes real-time detection with GPU acceleration, and achieves the effect of fast and accurate coexistence.

3.10. Pedestrian detection based on DPM and latent SVM
DPM algorithm (Deformable Parts Models) is a component-based detection algorithm. It is proposed by Felzenszwalb et al. to improve the HOG characteristics. It is the best detection method in the pedestrian detection traditional algorithm since 2008. Algorithm [13].

3.11. DPM structure
The feature used is the improved HOG: the block in the original HOG is canceled, but the normalized cell directly normalizes the area composed of the current cell and the four cells around it, so the effect is not much different. As a general target detection method, DPM combines signed and unsigned gradients when calculating gradient direction, and principal component analysis (PCA) is used to obtain a dimension reduction effect.

![DPM model](image)

Fig.7. The whole calculation process of DPM [13]

DPM model consists of two components, each consisting of a root model and several parts of the model. Root-Filter mainly locates the potential area of the object and obtains the position of the possible
object, but whether there is an expected object needs to be further confirmed by calculating with Part-Filter. The flow chart of the algorithm is shown in Figure 7.

3.12. Latent SVM
Latent SVM adds latent variables to the SVM. The SVM algorithm is used to train the classifier as the latent values of each component, the parameters of the model are transformed into the hyperplane of the SVM, and the model is continuously updated in an iterative manner.

3.13. Algorithm effect
The DPM algorithm is simple and intuitive, can adapt to a certain degree of deformation of pedestrians and solve the occlusion problem. It has achieved good results in pedestrian detection, but it also has some limitations: First, it requires artificial design of excitation characteristics, large workload, and calculation speed. Not very fast; Secondly, this kind of artificial feature can not adapt to the object detection with large rotation, stretching and other visual angle changes, and has no universality for the detected scene, which is also a common fault based on the artificial feature algorithm.

4. Pedestrian Detection Based on Deep Learning
Traditional pedestrian detection methods based on machine learning can achieve good results in some scenarios, but they can not meet the actual needs more and more, so people turn their attention to deep learning. Deep learning is a special kind of machine learning, and machine vision is the first field in which deep learning achieves breakthroughs. Target detection algorithms based on deep learning can be roughly divided into two categories: Region Proposal-based algorithms, which represent R-CNN, SPP-NET, Fast R-CNN, and regression-based algorithms such as YOLO and SSD. This article will focus on the R-CNN class, YOLO, and SSD.

4.1. CNN (Region with CNN feature)
R-CNN was proposed by Ross Girshick and the algorithm can be divided into three steps [16]:

1. Candidate region extraction. R-CNN adopts a higher quality region extraction method, namely region proposal based on selective search (SS). Generally, 1k~2k candidate frames are extracted from the image to be detected.

2. CNN feature extraction. After the candidate frame is scaled to a fixed size, the feature of each candidate frame is extracted by CNN to obtain a feature vector of a fixed dimension.

3. Classification and boundary regression. The obtained eigenvectors are input to SVM for classification, and the category information is obtained and then sent to the fully connected network for boundary regression.

Although R-CNN is cleverly designed to break the bottleneck of DPM algorithm for many years, it seems that it has many shortcomings: First, although the "exhaustive method" of sliding window is avoided, R-CNN still has repeated operations, thousands of candidates. The box needs to extract features through CNN, the amount of calculation is still large, and the overlap of candidate frames brings a lot of unnecessary double counting; Secondly, the training process of R-CNN is very complicated. Candidate region extraction, feature extraction, classification and regression are operated separately, and the data between steps are saved separately, which results in its slow speed. Finally, the scaling of input before feature extraction of CNN will cause image distortion, and the detection performance of the algorithm will be greatly reduced.

Fig.8. The architecture of R-CNN [16]
4.2. Fast R-CNN
To improve the above shortcomings, Ross proposed Fast R-CNN inspired by SPP. The main innovation is the ROI (Region of interest) Pooling layer [14]. One of the disadvantages of R-CNN is that it requires scaling of the image before CNN to cause deformation, which has a huge impact on feature selection. Fast R-CNN removes the limitation of data input by the algorithm by introducing the ROI pooling layer. The function of this layer is to sample the candidate frame convolution feature maps of different sizes into fixed-size features.

In addition, Fast R-CNN solves the disadvantage that R-CNN separates the steps and consumes time space to a certain extent. It integrates neural network feature extraction and subsequent classification regression through a new network, classifies it with softmax instead of SVM, and replaces boundary regression with multi-task loss function border regression.

These improvements greatly reduce the operation time of Fast R-CNN and improve the detection rate. However, the algorithm still uses SS to extract candidate regions, and the computing speed is still not enough for real-time applications.

4.3. Yolo
YOLO was proposed by Redmon et al. in 2016 [17], which is a typical representative of the “end-to-end” (end-to-end) method. The highlight of the algorithm that differs from the R-CNN series is:

① YOLO is a single-stage algorithm. It discards the steps of candidate frame region extraction, and simultaneously achieves target location and recognition through CNN, and extracts feature features, candidate frames, and classifications into the same branchless convolution network, carry out.

② Target detection is treated as a regression problem. After one inference, the position, category and confidence rate of the object in the input image can be obtained. The main advantages of YOLO are: fast detection speed, nearly 10 times higher than Faster RCNN, reaching the requirements of real-time detection; using the whole image as input, less background errors; strong generalization ability.

YOLO’s network structure draws on Google Net [23] and contains 24 convolutional layers and 2 fully connected layers, as shown in Figure 10:

First, the input image is scaled to the same size and divided into s×s grid cells, which grid is detected in which grid, and which grid is responsible for predicting the position of the target.
The work of each grid is to predict B bounding boxes, to return the location information of each border and the corresponding confidence value, and to predict C categories. Each border has four position parameters (x, y, w, h), representing the center coordinate of the target, the width and height of the border, and a confidence value Therefore, a border contains 5 data values, and YOLO's final fully connected layer output is a tens of s × s × (5 * B + C).

YOLO converts the target inspection task into a regression problem, greatly speeding up the inspection, enabling YOLO to process 45 images per second. Moreover, since each network predicts the target window using full-picture information, the false positive ratio is greatly reduced. In 2016, Redmon and others put forward YOLO 9000 [24], which makes up for the shortcoming of low detection accuracy of YOLO algorithm: combining target detection and target training, and making a series of improvements to the detection framework of YOLO.

4.4. SSD
The SSD algorithm was proposed by Liu et al. in 2016 [18]. It follows the method of direct regression of borders and classification probability in YOLO. At the same time, it also refers to Faster R-CNN, and uses anchor to improve recognition accuracy. By combining these two structures, the SSD maintains a high recognition speed and raises the MAP to a higher level, achieving speed and accuracy to a certain extent.

Figure 12 shows that SSD, like YOLO, adopts the idea of meshing, but it abandons the full connection layer in YOLO and has no limitation on the size of the detected object: small targets are detected in the output image of the front convolution layer, and large targets are detected in the output image of the back convolution layer.

4.5. Comparison of Key Pedestrian Detection Algorithms
Based on the performance of the above algorithms and some other representative algorithms on INRIA and PASCAL, the performance, advantages and disadvantages of these algorithms are compared and summarized [1,26]. The traditional algorithm compares the log-average miss rate and the two FPS evaluation indicators, as shown in Table 1. The deep learning algorithm compares the two evaluation
indicators of mAP and FPS, as shown in Table 1. Finally, these algorithms are excellent. The summary of the shortcomings fully illustrates the performance of each algorithm, as shown in Table 2.

Table 1. Experimental results of different traditional pedestrian detection algorithm

| Method                          | Datasets | log-average miss rate | FPS  |
|---------------------------------|----------|-----------------------|------|
| HOG+liner SVM [1]               | INRIA    | 68.46%                | .239 |
| PoseInv [19] (HOG+AdaBoost)     | INRIA    | 86.32%                | .474 |
| VJ [6] (Haar-like+AdaBoost)     | INRIA    | 94.73%                | .447 |
| Shapelet+AdaBoost [21]          | INRIA    | 91.37%                | .051 |
| ICF+AdaBoost [10]               | INRIA    | 14% at 1 fppi         | /    |
| LatSvm-V1 [22] (DPM+Latent SVM) | PASCAL   | 79.78%                | .392 |
| LatSvm-V2 [23] (DPM+Latent SVM) | INRIA    | 63.26%                | .629 |

Table 2. Experimental results of different deep learning methods on PASCAL VOC testing data set.

| Detection Frameworks           | Train      | mAP  | FPS  |
|--------------------------------|------------|------|------|
| Fast R-CNN [14]               | 2007+2012  | 70.0 | 0.5  |
| Faster R-CNN (ResNet) [15]    | 2007+2012  | 76.4 | 5    |
| Faster R-CNN (VGG-16) [15]    | 2007+2012  | 73.2 | 7    |
| Faster R-CNN (ZF) [15]        | 2007+2012  | 62.1 | 18   |
| YOLO [17]                     | 2007+2012  | 63.4 | 45   |
| Fast YOLO [17]                | 2007+2012  | 52.7 | 155  |
| SSD300 [18]                   | 2007+2012  | 74.3 | 59   |
| SSD500 [18]                   | 2007+2012  | 76.8 | 23   |

Table 3. The highlights and disadvantages of different pedestrian detection algorithm [27]

| traditional algorithm          | Method                          | highlights and disadvantages                                                                 |
|--------------------------------|---------------------------------|-----------------------------------------------------------------------------------------------|
|                                 | HOG+liner SVM [1]               | Advantages: It can keep good invariance to image geometry and optical distortion, neglect pedestrian's subtle limb movements without affecting the detection effect, and effectively depict human features. Disadvantages: the gradient is sensitive to noise; high latitude, slow calculation speed, unsuitable for real-time monitoring; poor handling ability for occluded pedestrians; unused color, shape and texture features. |
|                                 | PoseInv [19] (HOG+AdaBoost)     | Advantages: Basic real-time monitoring is achieved, which is 70 times faster than the previous HOG+SVM algorithm; it is robust to pedestrian attitude changes, complex background and illumination changes. Disadvantages: Easy to be disturbed by noise. |
| Algorithm | Advantages | Disadvantages |
|-----------|------------|---------------|
| **VJ**[^6] (Haar-like+AdaBoost) | Advantages: VJ was first proposed in these methods (2003). Integral images for fast feature computation and AdaBoost cascade structure for detection are introduced, which are the basis of subsequent detection algorithms. | Disadvantage: Compared with the later algorithm, the proposed algorithm is not mature, and has a high false detection rate and a slow detection speed. |
| **Shapelet+AdaBoost**[^22] | Advantages: Shapelet features improve the shortcomings of Edgelet features that need to be manually set. Automatically generate local features through machine learning. It has good adaptability; it is more suitable for human contours; it is reduced by 10 times compared with HOG+SVM. | Disadvantages: Training samples with machine learning algorithm twice, with large amount of calculation and high time complexity; tend to extract the smallest amount of information that can be classified, and will lose many data sets that may be useful. |
| **ICF+AdaBoost**[^10][^28] | Advantages: collecting feature information from different angles; locating space more accurately and setting parameters less in the detection process; combining ICF with cascaded AdaBoost classifier, real-time detection is realized through GPU acceleration; high accuracy. | Disadvantage: The feature pool extracted is redundant; each feature scale needs to be calculated separately, which takes a long time; and GPU is needed to assist in acceleration. |
| **LatSvm-V1**[^22] (DPM+Latent SVM) | Advantages: It can adapt to a certain degree of deformation of pedestrians; solve the occlusion problem. | Disadvantages: It is necessary to artificially design the excitation characteristics, the workload is large, and the calculation speed is slow; the artificial features can not adapt to the object detection of large rotation and stretching angles, and the detection scene is not universal. |
| **LatSvm-V2**[^23] (DPM+Latent SVM) | Advantages: It can keep good invariance to image geometry and optical distortion, and can ignore the pedestrian's subtle limb movements without affecting the detection effect. | Disadvantage: Because the gradient is sensitive to the noise, it is not suitable for real-time monitoring because of its high latitude and slow calculation speed. |

**Deep Learning Algorithms**

| Algorithm | Advantages | Disadvantages |
|-----------|------------|---------------|
| **Fast R-CNN**[^14] | Advantages: The first truly end-to-end algorithm (ignoring the RP generation process); the RoI pooling layer (a special case of the SSP layer) is designed; it is faster and more accurate than the SPP algorithm; feature extraction does not require disk storage. | Disadvantages: External RP computing has become a new bottleneck; speed is still not fast enough for real-time detection. |
| **Faster R-CNN (ResNet)**[^15] | Advantages: RP is proposed to replace selective search to produce almost cost-free and high-quality RPs; translation invariants and multi-scale anchor boxes are introduced as reference in RPN; RPN and Fast R-CNN are unified into a single network by sharing CONV layer; detection speed is one order of magnitude faster than Fast R-CNN without performance loss. | |
Disadvantage: The training process is complex; it is not a streamlined method; it still cannot meet the real-time detection standard.

YOLO \cite{17}  
Advantages: The first efficient first-order detector; the step of completely abandoning the RP; an elegant and efficient detection framework; the detection speed is qualitatively better than the previous algorithm. 
Disadvantages: Accuracy is far behind with advanced detectors; it is difficult to locate smaller targets.

Fast YOLO \cite{17}  
Advantages: the first accurate and efficient first-order detector; effective combination of RPN and YOLO ideas, detection on multi-scale CONV layer; faster and more accurate than YOLO. 

SSD300 \cite{18}  
Advantages: the first accurate and efficient first-order detector; effective combination of RPN and YOLO ideas, detection on multi-scale CONV layer; faster and more accurate than YOLO. 
Disadvantages: Not suitable for small target detection

SSD500 \cite{18}  
Advantages: the first accurate and efficient first-order detector; effective combination of RPN and YOLO ideas, detection on multi-scale CONV layer; faster and more accurate than YOLO. 
Disadvantages: Not suitable for small target detection

5. Conclusion
In this paper, representative pedestrian detection algorithms are summarized, and discussed from the perspective of traditional algorithms and deep learning algorithms. The algorithms and improvements are described. The effectiveness, advantages and disadvantages of each algorithm are also analyzed. Through the description and comparison of the pedestrian detection algorithm, it can be seen that with the continuous improvement of computer hardware and algorithms, the detection effect has been greatly improved, but the occluded pedestrians and small-scale pedestrians are still not well detected. There is still a lot of room for development in pedestrian testing, and there are still many challenges waiting to be completed.

References
[1] N. Dalal and B. Triggs. “Histograms of oriented gradients for human detection”. In CVPR’05, 2005.
[2] Qu, and Liu. "Non-background HOG for pedestrian video detection." Eighth International Conference on Natural Computation IEEE, 2012.
[3] Wang, Xiaoyu, T. X. Han, and S. Yan. "An HOG-LBP human detector with partial occlusion handling." IEEE International Conference on Computer Vision IEEE, 2010.
[4] Cheng Guangtao, Chen Xue, and Guo Zhaozhuang. "Pedestrian Visual Detection Based on HOG Characteristics." Sensors and Microsystems 30.7 (2011).
[5] Viola, Paul, M. Jones, and D. Snow. "Detecting Pedestrians Using Patterns of Motion and Appearance." IEEE International Conference on Computer Vision IEEE Computer Society, 2003.
[6] Viola, Paul, and M. J. Jones. "Robust Real-Time Face Detection." International Journal of Computer Vision57.2(2004):137-154.
[7] Zhu, Qiang, et al. "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients." Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on IEEE, 2006.
[8] Fanchunmian, Du Weiping, and Liu Yanrong. "Pedestrian detection based on HOG features combined with AdaBoost algorithm." Automation technology and application v.37; No.277.07 (2018): 93-95.
[9] Cheng, Wen Chang, and D. M. Jhan. "A cascade classifier using AdaBoost algorithm and support vector machine for pedestrian detection." IEEE International Conference on Systems IEEE, 2011.
[10] Dollar, Piotr and Tu, Zhuowen and Perona, Pietro and Belongie, Serge "Integral Channel Features" in Proceedings of the British Machine Vision Conference, 2009
[11] Mohan, A., C. Papageorgiou, and T. Poggio . "Example-based object detection in images by components." IEEE Transactions on Pattern Analysis and Machine Intelligence 23.4(2001):349-361.
[12] Wu, Bo, and R. Nevatia . "Detection of multiple, partially occluded humans in a single image by
Bayesian combination of edgelet part detectors.” Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on IEEE, 2005.

[13] Pedro F. Felzenszwalb Ross B. Girshick David McAllester and Deva Ramanan "Object Detection with Discriminatively Trained Part Based Models" in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.

[14] Girshick R. Fast R-CNN[J]. Computer Science, 2015.

[15] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]// International Conference on Neural Information Processing Systems. 2015.

[16] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for object detection and semantic segmentation[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014.

[17] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2016.

[18] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[J]. 2015.

[19] Forsyth D, Torr P, Zisserman A. [Lecture Notes in Computer Science] Computer Vision – ECCV 2008 Volume 5305 || A Pose-Invariant Descriptor for Human Detection and Segmentation[C]// European Conference on Computer Vision. Springer-Verlag, 2008:423-436.

[20] P. Dollár, S. Belongie, and P. Perona, “The Fastest Pedestrian Detector in the West, ” Proc. British Machine Vision Conf., 2010.

[21] Sabzeydani P, Mori G. Detecting Pedestrians by Learning Shapelet Features[C]// 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), 18-23 June 2007, Minneapolis, Minnesota, USA. IEEE, 2007.

[22] Felzenszwalb P F, Mcallester D A, Ramanan D. A Discriminatively Trained, Multiscale, Deformable Part Model[C]// 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA. IEEE, 2008.

[23] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, “Object Detection with Discriminatively Trained Part Based Models, ” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627-1645, Sept. 2010.

[24] Zhong Z, Jin L, Xie Z. High Performance Offline Handwritten Chinese Character Recognition Using GoogLeNet and Directional Feature Maps[J]. 2015.

[25] Redmon J, Farhadi A. [IEEE 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Honolulu, HI (2017.7.21-2017.7.26)] 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - YOLO9000: Better, Faster, Stronger[J]. 2017:6517-6525.

[26] Everingham M, Eslami S M A, Gool L V, et al. The Pascal Visual Object Classes Challenge: A Retrospective[J]. International Journal of Computer Vision, 2015, 111(1):98-136.

[27] Liu L, Ouyang W, Wang X, et al. Deep Learning for Generic Object Detection: A Survey[J]. 2018.

[28] Huang Peng, Yu Fengqin, Chen Ying. Fast multi-scale pedestrian detection with improved integral channel characteristics [J]. Computer Engineering and Application, 2017 (14).