ISSN: 2414-1895 DOI: 10.6919/ICJE.202104_7(4).0026

Summary of Pedestrian Detection in Computer Vision

Zixuan Huang, Xinfeng Chen

Wenzhou Polytechnic, Wenzhou, Zhejiang, China.

Abstract

Pedestrian detection is the hot spot and difficult point in computer vision research, the main pedestrian detection technologies and research status at home and abroad were summarized in this paper. In this paper, first of all, pedestrian detection technologies were divided into traditional detection methods and detection methods based on deep learning, the main contents and application scopes of these methods were analyzed and compared, then the advantages and disadvantages of various detection methods were summarized, finally, the future of pedestrian detection technologies were forecasted.

Keywords

Pedestrian Detection; Traditional Methods; Deep Learning; CNN; YOLO.

1. Introduction

With the rapid development of computer technology, target detection [1] has become the research hot spot in the field of computer vision [2], and it has been widely used in national security, human-computer interaction [3] and information security. Pedestrian detection belongs to a type of target detection, and it is widely used in the intelligent monitoring, security, and assisted driving, etc. The tasks of pedestrian detection include two parts: target classification and target positioning, namely judging whether the image contains pedestrians, and finding out where the pedestrians are (region of interest, ROI), and use external rectangle frame.

According to whether it is necessary to manually extract features in the pedestrian detection algorithm, the detection algorithm can be divided into traditional methods and target detection algorithm based on deep learning [4]. Traditional methods have low requirements on computer hardware and are suitable for scenes where high-performance computers are not configured. On the other hand, when the pedestrian target is too small to extract enough features for learning, traditional methods can also identify pedestrians well [5]. The pedestrian detection algorithm based on deep learning solves the shortcomings of traditional target detection sliding window selection and manual feature extraction, the target detection accuracy and real-time performance are greatly improved by introducing self-learning target features of convolutional neural network (CNN) [6], area candidate frames or direct regression methods. In this paper, the traditional methods of pedestrian detection and the detection algorithm based on deep learning were introduced, respectively; finally, the existing problems in the field of pedestrian detection were summarized and developed.

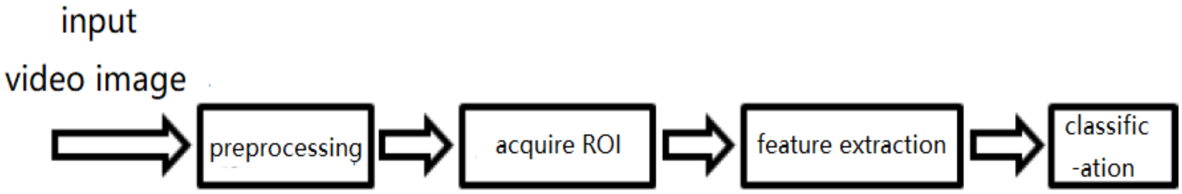

Fig. 1 The general sequence of the main modules in the pedestrian detection task

2. Traditional Pedestrian Detection Methods

Generally speaking, the frameworks of traditional pedestrian detection methods are divided into the following main parts, as shown in Fig.1: image preprocessing, selection and segmentation of region of interest (ROI), feature extraction [7], classification.

First, the input video image is preprocessed to improve the success rate of follow-up work.

Then, the region of interest (ROI) is acquired by the multi-scale sliding window method [8]. In order to reduce the number of ROI windows, the temporal difference method, background difference method, or target object information can be used. First, for the temporal difference method, the moving target will change the pixels in certain areas of the two adjacent frames of the video, the moving target in the video can be found out by making the difference among the pixels of the two adjacent frames. Second, the background difference method is based on the image difference method, which can realize fast segmentation of moving targets in complex backgrounds. The original image is preprocessed by median filtering, combined with the idea of background and temporal difference method, and the improved Surendra algorithm is used to quickly extract and update the background image, then ROI can be extracted more effectively [9]. Third, using the target object information, such as the edge information of the input image [10], it can also effectively segment the target and extract ROI.

After that, feature extraction mainly uses manual selection. The features of pedestrian detection are mainly divided into three categories: gradient, color, and outline, specifically include Haar-like [11] feature, shape context [12], Edgelet operator of head and shoulder detection [13], histogram of gradient(HOG) [14], and depth information [15], etc. Among them, HOG (Histogram of Oriented Gradients) is the most widely used pedestrian feature descriptor at present. The HOG feature constitutes the features by calculating and counting the histogram of gradient of the local area of the image, and it can still effectively describe the edge features of the human body when the illumination changes and the target deviate very small. Dollar [16] compared several pedestrian detection methods at the highest level at present, and findings indicated that no single feature can exceed HOG. In allusion to the detection limitations of HOG in the case of unclear pedestrian target outline, etc., some scholars choose to take local binary mode LBP [17], motion features [18], color features [19] and posterior features of sample universality [20] as additional feature to combine with the HOG operator, and provide effective supplementary information. Wojek [21] combines multiple features to train a new detection model, although its performance exceeds various single detection operators, it still cannot meet the detection requirements in scenes where pedestrian targets are severely overshadowed.

In addition to the above features, pedestrian features also include scale invariant feature transform (SIFT) feature [22], Gabor feature, DPM feature [23] and so on.

The commonly used classifiers are support vector machine (SVM) [24] and AdaBoost [25]. The extracted HOG features combines the algorithm composed of SVM classifier, which has been widely used in the field of image processing target recognition, especially good results have been obtained in pedestrian detection [26], and it is also widely used in engineering projects.

3. Pedestrian Recognition Technology Based on Deep Learning Methods

The concept of deep learning was proposed by Hinton in 2006, deep learning has been widely used in recent years, and it has been widely used in computer vision, speech recognition, and natural language processing. The target detection algorithms based on deep learning use CNN replace the traditional manual feature selection, which can be divided into the two-stage target detection algorithm and the single-stage target detection algorithm. The two-stage detection algorithm treats the object detection in accordance with classification problem, first, the region containing the object is generated, and then the candidate region is classified and calibrated to obtain the final detection result. The single-stage detection algorithm directly gives the final detection result; there is no explicit step of generating candidate frames, the main representative algorithm is the YOLO algorithm.

3.1 Two-stage target detection algorithm

3.1.1 R-CNN model

The R-CNN model proposed by Girshick R et al. made the target detection accuracy of the PASCAL VOC 2010 dataset increased by 18.6% [27], and it became the pioneering work of the follow-up R-CNN series of target detection. The framework process of R-CNN is shown in Fig.2, first, selective search is used to extract about 2,000 candidate frames; then the extracted candidate frames are preprocessed to the same size, and send them to the Alex-Net network for regional feature extraction; finally, the regional features extracted by CNN are classified and frame calibrated by SVM. The performance of the R-CNN algorithm is greatly improved in comparison with the traditional algorithm, but there is also a serious time-consuming of candidate frames generated by SS algorithm, tailoring will cause disadvantages: loss of information or the introduction of too much background, the large amount of repetitive calculation of convolution features, network training need to be carried out in steps.

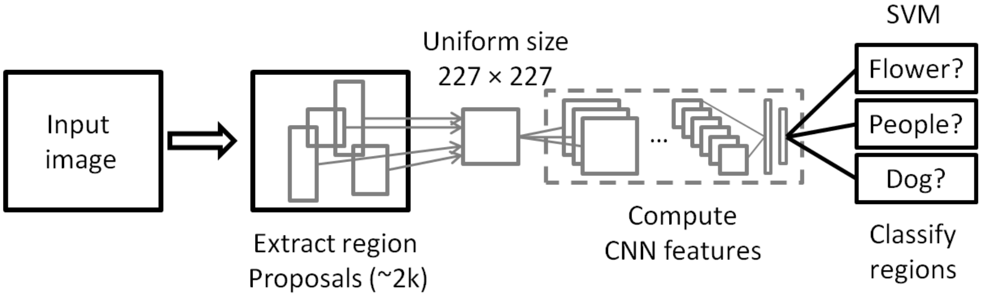

Fig. 2 R-CNN model

3.1.2 Fast R-CNN model

In 2015, Ross G et al. [28] proposed Fast R-CNN target detection algorithm. This algorithm combined SPP-Net to improve R-CNN, used the VGG16 backbone network replace the Alex-Net network, simplified the pyramid pooling layer in the SPP algorithm to a single scale, so that all layer parameters can be adjusted, changed the SVM classifier to the Soft Max classifier, introduced multi-task learning mode, and solved the problem of classification and location regression simultaneously. Compared with R-CNN and SPP-Net, Fast R-CNN integrates many steps into one model, the training process is no longer conducted step-by-step, reduces disk space usage, and accelerates training while improving network performance. But the shortcoming of Fast R-CNN is that it still needs special algorithm for generating candidate frames.

3.1.3 Faster R-CNN model

In 2015, the Ross Girshick team proposed Faster-Rcnn based on Fast-Rcnn, simple network target detection (ZF model) can reach 17f/s frame rate in speed, and the accuracy rate on the PASCAL VOC data set is 59.9%; the frame rate of complex networks is around 5 f/s, and the accuracy can reach 78.8% [29].

3.2 YOLO algorithm

The two-stage target detection algorithm has greatly improved the detection success rate, but still has the disadvantages of many training parameters and long training time. The YOLO v1 [30] algorithm based on regression directly uses a CNN complete classification and regression tasks simultaneously. However, because the YOLO v1 algorithm takes each lattice as the center point, it has the problem of low accuracy and low detection accuracy for small-scale objects and densely arranged objects. In order to overcome the problem of fast detection speed but low detection accuracy of YOLO v1, the YOLO v2 [31] algorithm introduced BN (batch normalization), multi-scale training, anchor frame mechanism and fine-grained features to improve the YOLO v1 algorithm. On the basis of YOLO v2, the YOLO v3 [32] algorithm uses a better backbone network, multi-scale prediction and 9 anchor frames for detection, it makes the detection algorithm more accurate while ensuring real-time performance. YOLO v4[33] has been modified in the four major sections: YOLO v3's input, BackBone backbone network, Neck, and Prediction, which improves the detection accuracy of blocked targets. On the COCO data set, when the detection speed is 83FPS, the AP value of YOLO v4 reaches 43%, which is 10% higher than that of YOLO v3.

4. Summaries and Prospect

The adaptive detector can be designed for traditional pedestrian detection methods. In special scenes, especially in monitoring situations where the camera is stationary, how to use incremental learning, online learning and other algorithms to migrate the general pedestrian detector to the special scene, improving the performance of pedestrian detectors through self-learning in the detection process will be the focus of future research.

The detection method based on deep learning, the two-stage target detection algorithm and single-stage target detection algorithm are the mainstream frameworks based on deep learning target detection currently. Compared with the single-stage target detection algorithm, the advantages of two-stage target detection algorithm are higher accuracy in positioning and detection rate, the anchor frame mechanism is used to consider regions of different scales to improve target detection performance. But its disadvantages are slow speed and long training time. Compared with the two-stage target detection algorithm, the advantages of single-stage target detection algorithm are fast speed and can learn the generalized features of the object, but the disadvantages is low accuracy of positioning and detection rate, and the detection effect of small objects is not good. The target detection algorithms based on deep learning have greatly improved their accuracy and real-time performance in comparison with traditional detection algorithms; however, due to the complexity and variability of real scenes, it still faces many problems. How to reduce the influence of complex background on target detection and how to reduce the accuracy decline caused by the change of target scale and shape has become hot spots in the field of target detection.

Future research on pedestrian detection technologies will continue to face the following problems: blocking problems and multiple viewing angles. A stable pedestrian detection must work in severe weather conditions (such as rain and snow, etc.), the system must be able to accurately detect partially blocked, low-resolution, long-distance pedestrians carrying large areas of objects, and maintain low errors report rate. In order to solve this problem, the pedestrian test database specifically for blocking, low resolution and long distance can be established. In addition, multi-cameras or depth information can be used to detect pedestrians, and the posture-based pedestrian detection technology in multi-eye vision can be explored. In short, pedestrian detection is the core difficult problem in the field of computer vision today, its solution has important theoretical significance and good application prospects, and it has also attracted a large number of researchers to invest in this field. Although certain results have been achieved, effective solutions to pedestrian detection problems in real complex scenes need further research.

Acknowledgments

Fund Project: General Scientific Research Project of Wenzhou Polytechnic (WZY2020026).

References

[1] Vailaya, Aditya, HongJiang, et al. Automatic Image Orientation Detection.[J]. IEEE Transactions on Image Processing, 2002.

[2] Voulodimos A, Doulamis N, Doulamis A, et al. Deep Learning for Computer Vision: A Brief Review[J]. Computational Intelligence and Neuroscience, 2018, 2018:1-13.

[3] Arroyo E, Selker T. Attention and Intention Goals Can Mediate Disruption in Human-Computer Interaction[J]. Lecture Notes in Computer Science, 2017, 6947:454-470.

[4] Yanming, Guo, Yu, et al. Deep learning for visual understanding: A review[J]. Neurocomputing, 2016, 187(Apr.26):27-48.

[5] Hu Yazhou, Zhou Yali, Zhang Qizhi. High Point Monitoring Pedestrian Detection Based on Background Modeling and Inter Frame Difference Method[J]. Research and Exploration in Laboratory, 2018, 037(009): 12-16.

[6] C U R A A B, A S L O, A Y H, et al. A deep convolutional neural network model to classify heartbeats[J]. Computers in Biology and Medicine, 2017, 89:389-396.

[7] Nixon M S. Feature Extraction and Image Processing[M]. Publishing House of Electronics Industry, 2013.

[8] Xu Y, Xu D, Lin S, et al. Sliding Window and Regression Based Cup Detection in Digital Fundus Images for Glaucoma Diagnosis[J]. 2011.

[9] Gou Juanying. Moving Object Segmentation Based on Background Subtraction[J]. Industrial Control Computer, 2013, 26(008): 36-37.

[10] Jin Peifei, Zhou Li, Liu Jian, Ge Zhiwei, Chen Jie. Pedestrian detection Based on ROI extraction[J]. Computer Engineering and Design, 2016, 37(11): 3035-3039.

[11] Viola P, Jones M J. Robust Real-Time Face Detection[J]. International Journal of Computer Vision, 2004, 57(2):137-154.

[12] Ding Y, Xiao J. Contextual boost for pedestrian detection[C]// Computer Vision & Pattern Recognition. IEEE, 2012.

[13] B Wu, R Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edge-let part detectors [A}. IEEE Conference on Computer Vision and Pattern Recognition [C]. Beijing, China: IEEE, 2005. 90-97.

[14] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE, 2005.

[15] Ningbo Wang, Xiaojin Gong, Jilin Liu. A new depth descriptor for pedestrian detection in RGB-D images [A}. IEEE Conference on International Conference on Pattern Recognition [C]. Tsukuba, Japan: IEEE, 2012. 3688-3691.

[16] Dollar P, Wojek C, Schiele B, et al. Pedestrian detection: an evaluation of the state of the art [J]. IEEE Transactions on Pattern Analysis and M achine Intelligence, 2012, 99: 1-20.

[17] Xiaoyu W, Han T X, Shuicheng Y. An HOG-LBP human detector with partial occlusion handling [A]. IEEE Conference on International Conference on Computer Vision[C}. Kyoto, JAPAN: IEEE, 2009: 32-39.

[18] Viola P, Jones M J, Snow D. Detecting pedestrians using patterns of motion and appearance[J}. IEEE Transactions on International Journal of Computer Vision, 2005, 63(2): 153-161.

[19] S Walk, N Majer, K Schindler, et al. New features and insights for pedestrian detection[A}. IEEE Conference on Computer Vision and Pattern Recognition[C}. San Francisco, USA: IEEE, 2010. 1030-1037.

[20] Liu Wei, Duan Cheng-wei, Yu Bing, et al. Method research on vehicular infrared pedestrian detection based on local features[J]. Acta Electronica Sinica, 2015, 43 (2): 217-224. (in Chinese)

[21] Wojek C, Schiele B. A Performance Evaluation of Single and Multi-Feature People Detection[M]. Pattern Recognition. Springer Berlin Heidelberg, 2008: 82-91.

[22] Cheng Chunyang, Zhang Jingya. An Improved Sift Algorithm Based on Multi-Scale Corner Point[J]. Computer Applications and Software, 2017(7).

[23] Zeng Jiexian, Cheng Xiao. Pedestrian Detection Co mbined with Single and Couple Pedestrian DPM Models in Traffic Scene[J]. Acta Electronica Sinica, 2016, 44(11): 2668-2675.

[24] Xue Mingdong, Guo Li. Image Classification Based on SVM[J]. Computer Engineering and Applications, 2004(30):233-235.

[25] Cao Ying, Miao Qiguang, Liu Jiachen, Gao Lin. Advance and Prospects of AdaBoost Algorithm [J]. Acta Automatica Sinica, 2013, 39(06): 745-758.

[26] Huang Chengdu, Huang Wenguang, Yan Bin. Pedestrian Detection Based On Codebook Background Modeling In Video[J]. Transducer and Microsystem Technologies, 2017, 36(03): 144-146.

[27] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// CVPR. IEEE, 2014.

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

[29] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6).

[30] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Computer Vision & Pattern Recognition. IEEE, 2016.

[31] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017:6517-6525.

[32] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.

[33] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020.