Application of Railway Passenger Flow Statistics Based on Mask R-CNN

Di Meng¹², Shangzhi Xu¹²*,

¹Information Processing and Intelligent Transportation System Laboratory, Department of Information and Communication Engineering, Tongji University, Shanghai, 201804 China
²The Center of National Railway Intelligent Transportation System Engineering and Technology, China

*Corresponding author e-mail: xushangzhi@tongji.edu.cn

Abstract This paper discusses and implements the application of Mask R-CNN in railway passenger flow statistics. Mask R-CNN is a target detector based on candidate region, combining deep learning, neural network and image recognition technology, and has a wide range of application value in the field of intelligent monitoring. The paper focuses on the algorithm named Mask-Flow that Mask R-CNN uses the same color for tracking on the same target, and counts the number of pedestrians. And the passenger flow statistics are compared with Retina-Net in the railway scene of the dense crowd, which proves the accuracy of the method in the paper.

1. Introduction
Since R-CNN was proposed, the candidate region-based algorithm has been successfully applied to a larger variety of computer vision tasks [1], for example to object-detection, segmentation, object tracking and superresolution. These examples are but a few of all the computer vision to which R-CNN have been very successfully applied ever since. Passenger flow statistics [2] technology based on artificial intelligence is an important branch in the field of computer vision. It mainly allows computers to simulate human brain work, intelligently reading, analyzing and processing image information. Such as face recognition and license plate recognition, which have entered industrial applications. Among them, railway passenger flow statistics technology is one of hot research topics, and it has broad application prospects in the field of intelligent monitoring and security management.

Passenger flow statistics technology based on Retina-Net [3] has serious missed detection problems for pedestrian recognition in dense scenes. Because the scale changes of target are not the same, and there are different degrees of occlusion between people and scenery. Also because of crowded passengers and limited shooting angles. Many basic algorithms [4] cannot accurately detect and count. With the gradual improvement of R-CNN, Mask R-CNN is one of the better detectors based on candidate detectors [5] in the candidate region. In this paper, the algorithm named Mask-Flow is designed to achieve higher accuracy rate in the application of railway passenger flow statistics and avoid counting repeatedly.
2. Passenger flow statistics algorithm

2.1. Mask R-CNN target detector

Mask R-CNN is an algorithm of instance segmentation proposed by Facebook, which can be used for target detection. Mask R-CNN is the result of improvements to Faster R-CNN. On the one hand, ROIAlign is used instead of ROIPooling. The principle is to replace the two quantization operations with a linear difference algorithm, so that there is no quantization error caused by quantization, which improves the performance of detection. On the other hand, it adds a Mask network branch based on Faster R-CNN to generate the corresponding Mask. Since the previous network uses Faster R-CNN/ResNet for multiple convolutions and pooling, the corresponding resolution is reduced. The mask branch starts to use deconvolution to improve the resolution and reduce the number of channels, and finally output the mask template. This will generate a Mask on each target in the processed image, and fill the color to achieve more accurate target detection.

2.2. Mask-Flow

2.2.1. Fundamental

Mask-Flow is an improvement of Mask R-CNN in the application of railway passenger flow statistics. The specific method of applying Mask-Flow to the railway passenger flow statistics scenario is divided into four stages: First, the network model parameters are obtained from training on the COCO training set. Next, the identification of other categories is deleted in the network, and only the detection of pedestrian target is detected, so that the detection efficiency is improved. Third, the color filling in Mask R-CNN uses the method of randomly generating colors, which is not applicable to the application scenario of railway passenger flow. Different color generation of each frame of image will make the same target in the video cannot be accurately tracked. For this reason, the algorithm is improved in this paper, and the pedestrian tracking in the railway passenger flow video is realized. Finally, through the loop detection of each frame in the video, combined with the high accuracy of the Mask R-CNN and pedestrian tracking, more accurate passenger flow statistics are achieved.

2.2.2. Tracking algorithm

The tracking algorithm [6, 7] uses the comparison detection between the adjacent frames. First of all, generate the corresponding Bounding Box (BB), color and pedestrian probability for the pedestrian mask obtained in each frame and put the above parameters of each pedestrian in the corresponding order into a list.

Later on, the coordinates of the adjacent frames BB are determined by IOU [8] algorithm. If the result of the determination is the same pedestrian in the adjacent frames BB, the color corresponding to the current BB of the previous frame is assigned to the next frame, thus realizing that the same pedestrian can always be tracked using the same color.

The algorithm principle of the IOU is as follows. Each BB is obtained by looping through the list obtained by the image of the latter frame, and the coordinates of the upper left corner and the lower right corner of the BB \( x_1, x_2, y_1, y_2 \) are obtained. Calculate the area \( S \) of the BB by using Equation 1;

\[
S = (x_2 - x_1) \times (y_2 - y_1)
\] (1)

Recycling each BB of the previous frame to obtain the coordinate \( x_1', x_2', y_1', y_2' \) and \( S' \). Calculates the intersection area \( S_{\text{inter}} \) of the two BBs by using Equation 2;

\[
S_{\text{inter}} = (\min\{x_2', x_2\} - \max\{x_1, x_1'\}) \times (\min\{y_2', y_2\} - \max\{y_1, y_1'\})
\] (2)

Finally, the calculation of the IOU is performed. As shown in Equation 3, if the obtained value satisfies the threshold value greater than the set value, then the operation of assigning the color value is performed.
\[ IOU = \frac{S_{\text{inter}}}{S + S' - S_{\text{inter}}} \] (3)

3. Implementation

The experimental results [9] obtained in this paper are compared with the experimental results of the Retina-Net algorithm in the same scene, as shown in Figure 1. Figure (a)(c)(e) are the results obtained by Retina-Net, which show that most of the crowds in the middle dense part are missed. Therefore, the pedestrian probability of the test results is not very high. However, figure (b)(d)(f) is the method of this paper. It can be seen that the Mask-Flow algorithm has much higher accuracy than the Retina-Net in the railway passenger flow statistics application scenario.

![Figure 1: Comparative experiment results schematic diagram](image)

Figure 2 shows the tracking algorithm. When the same pedestrian is continuously detected in the video, the pedestrian will get the same color of Mask to track, and then achieve more accurate passenger flow statistics.

![Figure 2: Mask-Flow tracking pedestrian schematic diagram](image)
4. Conclusion

This paper introduces the application of Mask-Flow in the field of railway passenger flow statistics. And compare the experimental results with the statistical results based on Retina-Net, which verifies the accuracy of the paper method. At the same time, we realize tracking algorithm, which avoids counting repeatedly to some extent due to data relevance. The passenger flow statistics in the railway scene do not require the number of pedestrians to be counted in every frame. Because for real-time video of about 25 frames per second, the pedestrian's moving speed is not too fast, so combining the methods in this paper with the operation of frame skipping, the passenger flow statistics will get fast and accurate statistical results. What’s more, the work of this paper can provide a certain feasibility reference for the further realization of personID, trajectory synthesis, and target tracking.

Based on Mask-Flow, passenger flow statistics of railway scenes are expected to be widely used in the field of intelligent monitoring, using machine monitoring to replace human labor, and truly realize the intelligentization of safety supervision.

5. Acknowledgments

This work is by the Natural Science Foundation of China under Grant Nos. 61773290 and 71571107, and the Opening Foundation of The Center of National Railway Intelligent Transportation System Engineering and Technology under Grant No. RITS2018KF04.

References

[1] Francisco Javier Cantu-Ortiz. Advancing artificial intelligence research and dissemination through conference series: Benchmark, scientific impact and the MICAI experience, J. Expert Systems With Applications, 41(2014)3: 781-785.
[2] Masoud O, Papanikolopoulos N P. A novel method for tracking and counting pedestrians in real-time using a single camera, J. IEEE Transactions on Vehicular Technology, 50 (2001) 5: 1267-1278.
[3] Lin T Y, Goyal P, Girshick R. Focal loss for dense object detection, J. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, pp. 2999-3007.
[4] Joseph Lin Chu, Adam Krzyżak. The Recognition Of Partially Occluded Objects with Support Vector Machines, Convolutional Neural Networks and Deep Belief Networks, J. Journal of Artificial Intelligence and Soft Computing Research, 4(2015) 1: 5-19.
[5] Ruimin Ge, Yu Zhu, Yuling Xiao, Zhihua Chen. The Subway Pantograph Detection Using Modified Faster R-CNN, M. Springer Singapore, 2017.
[6] Riverabautista J A, MarínHernández, Antonio, Marinurias L F. Using color histograms and range data to track trajectories of moving people from a mobile robot platform, J. Time of Flight, 3(2012) 288 - 293.
[7] Beltramello O , Benincasa G , Berckhart H J , et al. Implementation of a complex monitoring system to track people in the ATLAS detector, C. Conference of the IEEE Industrial Electronics Society. IEEE, 2010.
[8] Shafiee M J, Chyw1 B, Li F, et al. Fast YOLO: A Fast You Only Look Once System for Real-time Embedded Object Detection in Video, J. 2017.
[9] Nikhil Ketkar. Deep Learning with Python: A Hands-on Introduction, M. California: Apress, 2017.