Port Personnel Identification Based on Deep Residual Learning and Multi-Scale Feature Fusion

Zichuang Wang, Jinbiao Zheng, Jun Ling and Xueqian Xu

Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China.

Abstract

Considering the complexity and concealment of the geospatial structure and layout in port, it is a laborious task to observe the activities of personnel, which will also lead to the inability to accurately locate the position of personnel from port surveillance videos and affect the analysis of their behavior. In this paper, we used an improved person detection algorithm framework applied under the working environment of port, which achieves a stronger feature expression ability based on ResNet-101 deep residual network integrated the feature pyramid network for excavating the multi-scale information of port monitoring image. The experimental results show that the algorithm achieves high accuracy and good real-time performance in the actual application environment, and it effectively improves port personnel identification and analysis under different port scenarios.

Keywords

Smart Port; Person Identification; Feature Pyramid Structure; Deep Residual Network.

1. Introduction

In recent years, the combination of modern intelligent technology and various application fields of port promotes the reform and transformation of traditional ports, Smart Port has become the representative of the future port development pattern[1, 2]. In view of the complexity and diversity of the port operation environment where the port personnel in the surveillance video have closely similar with background color and often appear occlusion, irregular activities, poor visibility and other situations, it is a challenging assignment to solve the problem of port personnel target detection and analysis by using computer vision and artificial intelligence technologies.

Port personnel identification based on port surveillance video is an essential part of smart port visual management system solution. Port personnel target detection methods are mainly divided into the traditional approaches via feature selection and the object detection algorithm based on artificial neural network and deep learning. The former method usually first generates candidate regions on the input image, and then infer through feature extraction and pre-trained classifier, which the quality of the extracted features will directly affect the accuracy of classification. For example, Fang et al.[3] design a detector based on DPM (Deformable Part Model) algorithm and the HOG (Histograms of Oriented Gradients) image feature for identification, which used feature pyramid and sliding window to extract image information, and combined with SVM (Support Vector Machine) classifier for model training. On the basis of DPM, D. Dange et al.[4] respectively used SSA (Selective Search Algorithm) and DSD (Density Subgraph Algorithm) to generate and cleanup the region proposals to deal with the poor performance of mutual occlusion of objects. Nguyen et al.[5] improved the person detection framework of HOG-SVM by optimizing the generation of cell histograms and adopting SVM in parallel with block normalization computation to realize the excellent processing ability of image data and network low-power consumption. Das et al.[6] have further improved the speed and accuracy of pedestrian detection based on linear SVM and the cascade of boosted classifier, at the same time HOG and LDB (Local Difference Binary) are used to enrich the expression of image information, and the combination of nonlinear scale-space and image pyramid is used to enhance the sensitivity of pedestrian size. Cheng et al.[7] proposed a method with FFPM (A Fast Fused Part-based Model) to

accurately identify pedestrians even in a highly crowded environment by constructing the Haar-like features of different parts of human body and then integrating to generate deep-space information.

However, port personnel detection is influenced by surrounding buildings, light conditions, and occlusion of target in the image, while the traditional detection algorithm is limited by the poor feature robustness and generalization of manual design, which leads to the unsatisfactory detection accuracy. Recently, object detection algorithm based on deep learning, with its excellent ability of feature learning and transfer learning, has been developed rapidly in the fields of image feature extraction, classification and recognition. The mainstream object detection algorithms based on deep learning are mainly divided into two-stage and one-stage object detection algorithm. The two-stage detection algorithm is based on region proposals represented by R-CNN (Region-based Convolutional Neural Network) series[8-11], and the one-stage object detection algorithm is based on regression analysis represented by Yolo (You Only Look Once)[12, 13] and SSD (Single Shot multi-box Detector)[14]. For example, Peng et al. [15] designed a general detection framework compared with Faster R-CNN algorithm, which applied KD (Knowledge Distillation) to feature extractor based on region proposal network through incremental learning method to improve the performance of the network in learning new categories. To solve the problem of training inefficiency caused by deep convolutional network iteration, Yin et al.[16] optimized the network structure to reduce the number of parameters on the basis of Yolov3 algorithm, so as to achieve the effective feature extraction of input images without fussy training and parameter adjustment.

For the traditional solution, in the face of the port environment, the personnel and the surrounding environment may form similar brightness, which leads to the failure to accurately locate the specific position of the port personnel, observe and analyze the problems confronted by their movement and posture. Combined with the current mainstream method of feature extraction and object detection algorithm, we selected several typical application scenarios with demonstration significance which trained and identified a few classic scenarios of port personnel target detection in the port environment of quayside container crane, and it provided support for the follow-up research of port scene.

1.1 Port Personnel Identification Framework in Port Environment

1.1.1 Improved Faster R-CNN Port Personnel Identification Framework Based on Deep Residual Learning and Feature Pyramid Network

The port personnel identification method proposed in this paper contains the task of bounding-box regression and target classification with confidence, which is conducive to deal with the obstacles, lighting conditions and other adverse conditions in the port environment. Firstly, we combined the ResNet-101 deep residual network with the Faster R-CNN object detection algorithm where the replaced backbone network integrated deep residual learning has stronger image feature expression ability. Secondly, the feature pyramid network (FPN) was applied to the region proposal network (RPN) layer to solve the loss of semantic information of the image when the convolutional neural network (CNN) was pooled to the last layer, and the multi-scale information of the image was mining by using the pyramid form of CNN hierarchy features, which could achieve accurate estimation for the port personnel.

In the port scenarios, due to the high similarity caused by the small chromatic aberration between the port personnel and background buildings, or the problems such as too small target of the personnel, mutual occlusion and confusion with the background may occur in the port surveillance video, thus it affects the misjudgment of identification and position prediction for port personnel. In this module, the residual learning is introduced to avoid the phenomenon of gradient disappearance or gradient explosion when the depth increasing of CNN network, so as to ensure the prevention of network degradation, reduce the network scale and computing costs at the same time. Feature pyramid network is a representative model of image pyramid feature representation generated by object detection. By combining FPN with deep learning unit, construction of cross-scale image feature and multi-scale object detection are realized.

ResNet-101 is used as the backbone network to improve the Faster R-CNN in this module. While deep learning is introduced into the object detection algorithm, the accuracy of data sets decreases with the steeply increase of convolutional neural network layers, which is obviously not caused by overfitting. For a neural network, the parameter matrix is repeatedly adjusted to make the output result closer to the expected value through continuous iteration. Because the neural network needs to continuously propagate the gradient in the process of back propagation, the gradient will gradually disappear when the number of network layers deepens, so that the weight of the previous network layer cannot be effectively adjusted. ResNet-101 deep residual network introduces the residual network structure, as shown in Figure 1. Compared with the usual CNN, the most obvious difference lies in the skip connections, also called shortcut connections, that is, the branch of the bypass connects the input to the following layer, so that it can learn residuals directly.

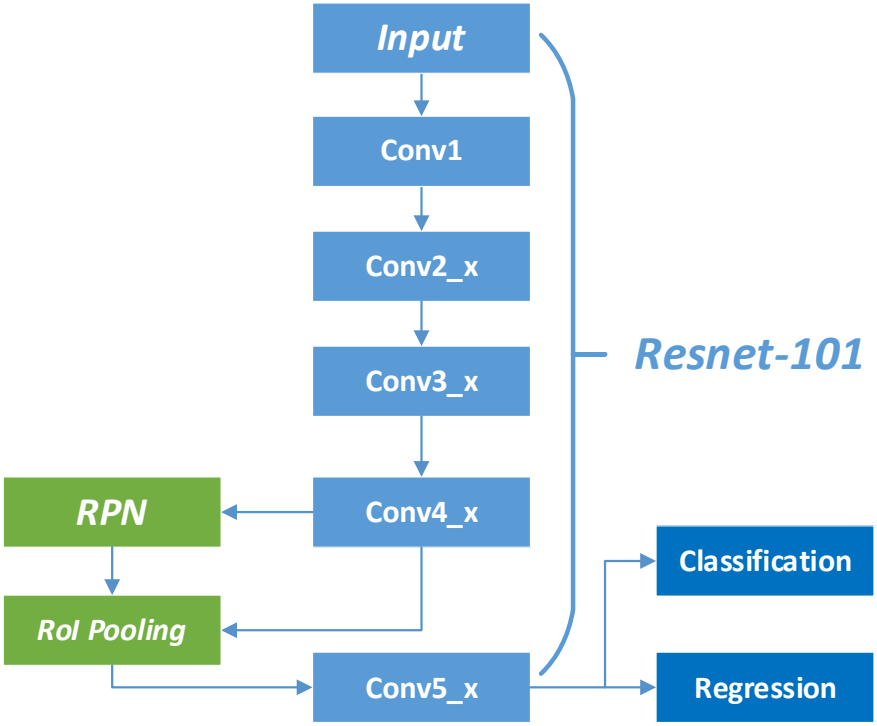


Figure 1. ResNet-101 Deep Residual Network

Thus, the introduction of deep residual neural network can significantly improve the depth of Faster R-CNN object detection network, and avoid the network degradation caused by depth increase. The Visual Geometry Group (VGG) network[10] adopted by the original Faster R-CNN has only 16 layers as the last three Fully Connection layers (FC) contain a large number of parameters, the depth can reach 101 layers when it's replaced by ResNet network adding direct connections in networks to compress computation. The ResNet network passes the input information to the output by a detour, which the problem of information loss, to a certain extent, is solved and the integrity of information is protected.

1.2 ResNet-101 Deep Residual Network for Port Personnel Identification

To cope with the change of the detection scale of personnel in port surveillance videos, in this paper we combine the original Faster R-CNN object detection algorithm with the feature pyramid network in RPN layer to achieve the purpose of multi-scale detection.

RPN can directly generate region proposals and greatly improves the speed of the detection frame generation, which shares the image convolutional features with the object detection network, thus producing almost costless region proposals. However, the original Faster R-CNN network only uses a single high-level feature map for object classification and bounding box regression, which will lead

to the loss of most pixel information in the downsampling process, especially the limited itself that of small target object. Combining FPN with RPN, the Region of Interest (RoI) of each level is predicted on the feature pyramid structure where the feature map with low-resolution and strong semantic information and the rich spatial information one with high-resolution and weak semantic information are fused on the premise of increasing less computation. It makes the possibility to quickly build the feature pyramid of strong semantic information with all scales from a single input image on a single scale without too much cost. As shown in Figure 2, C5 layer is first convoluted by 1×1 kernels to change the channel number of the feature map (the same as the dimension of RPN layer in Faster R-CNN, which is convenient for classification and regression). M5 through upsampling, coupled with the feature map of C4 after 1×1 convolution (each element of the same position in the feature map is directly added), then got M4. By twice again, M2 and M3 are obtained respectively. The M-layer feature map is then convoluted by 3×3 to reduce the aliasing effect caused by nearest neighbor interpolation and generate the final P2, P3, P4, P5 layer features.

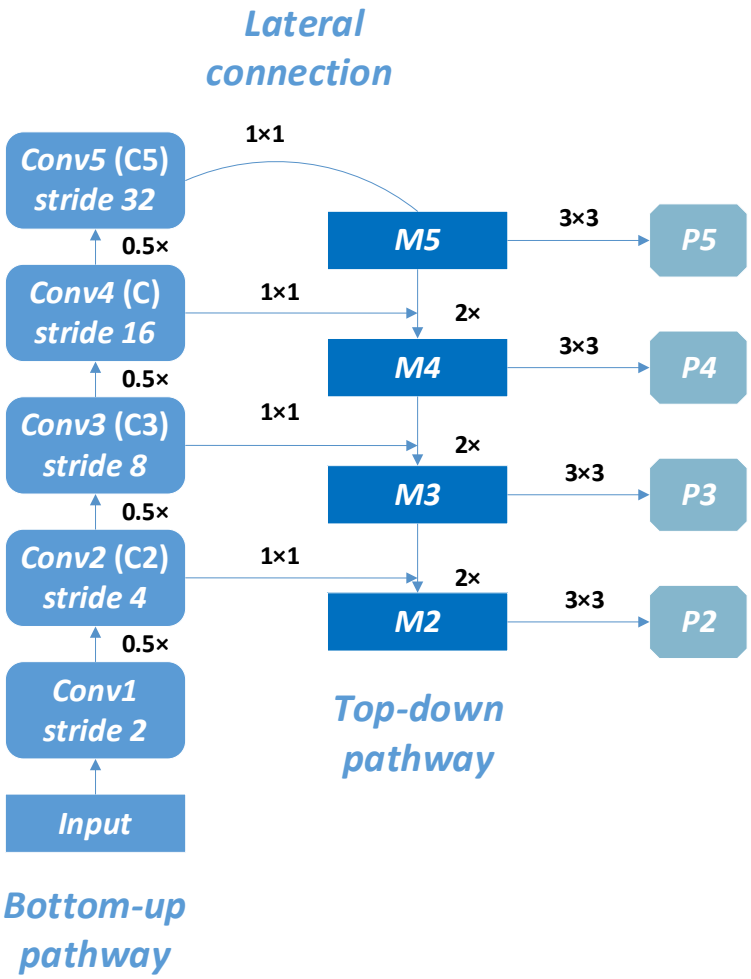
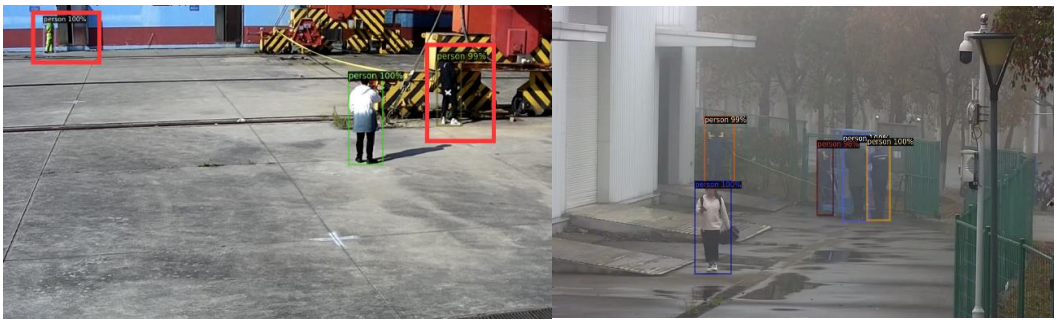


Figure 2. Feature pyramid network for the multi-scale detection

2. Literature References

In this part, we selected several specific application scenarios with typical demonstration significance in the port surveillance video as shown, and compared the detection performance of multi-target identification for port personnel in different port operating environments. The identification of port personnel uses the label of person in the object detection of Microsoft COCO dataset[17] as the training to verify the performance of the proposed method in the actual environment. In the upper left corner of the bounding box is the category label and the corresponding confidence.



(a) (b)

Figure 3. Port personnel identification in actual port scenarios

The results under different scenarios are shown in Figure 3. The identification performance of port personnel in a typical port scene is shown in Figure 3(a). In this scene, the optical chromatic aberration between the color of the buildings and the clothing of personnel near the container yard gantry crane or the container ships is so similar that it's particularly prone to neglect or even impossible to observe with naked eye, while the improved Faster R-CNN object detection algorithm can accurately identify inconspicuous targets. Simultaneously, through the multi-scale feature extraction of image pyramid, the algorithm proposed in this paper can improve the detection effect of small targets to a certain extent. Figure 3(b) shows the disturbance of severe weather (e.g. foggy weather). The image captured by surveillance video in severe weather contains a mass of irregular noise which leads to image quality degradation. Because of the influence of visibility, many features of port personnel in the image are covered up, which resulted in poor visibility of image details and greatly reduced identification of target personnel. In view of the above cases, the algorithm proposed in this paper is nicely adequate for achieving the detection results for port personnel recognition under the conditions of restricted visibility such as weak illumination condition or fog weather.

In the part of port personnel identification, the detection accuracy Pa, false alarm probability Pf and sensitivity Ps are selected as the evaluation parameters of the verification algorithm, and their definitions are shown in Equation(1)-(3):

$$P_a = N_{TurePositive} / N_{GroundTruth} \quad (1)$$

$$P_f = N_{FalsePositive} / N_{total} \quad (2)$$

$$P_s = N_{FalseNegative} / N_{GroundTruth} \quad (3)$$

Where, $N_{TurePositive}$ is the number of port personnel correctly identified by the algorithm, $N_{GroundTruth}$ is the number of all port personnel in the actual port scene, $N_{FalsePositive}$ is the number of other targets mistakenly identified as port personnel, $N_{FalseNegative}$ is the number of port personnel missed by the algorithm, and N_{total} is the number of port personnel identified by the algorithm.

The performance of the improved algorithm proposed in this paper is compared with the original Faster R-CNN object detection algorithm under the background of actual port environment. Then we ran it through repeated experiments in the same type of different port scenarios and took the average. The probabilities under different circumstances are shown in Table 1. It can be seen that the port personnel identification framework based on ResNet-101 deep residual neural network and feature pyramid network can almost detect the port personnel in each set of scenes, and the accuracy is improved compared with the original Faster R-CNN algorithm. Among them, the detection precision can reach 99.0% and 94.1% in the typical port environment and the occasions with poor visibility, and the accuracy for small targets can also be improved to more than 90%. With the optimizing of residual unit and pyramid structure, the improved algorithm reduces the Pf and Ps index by about 2% compared with the original version, and the average false-alarm probability and the average miss-detection probability are 4.58% and 6.16% respectively. All of the above show that the model has good accuracy and effectiveness, which can be applied to personnel identification in actual port application scenarios.

Table 1. Identification performance of port personnel in different scenarios

Scheme	Improved port personnel identification algorithm			Faster R-CNN detection algorithm		
Index	P_a / %	P_f / %	P_s / %	P_a / %	P_f / %	P_s / %
Typical Port Scenarios	99.0	3.26	5.34	96.2	4.11	7.67
Poor Visibility	94.1	5.89	6.98	92.0	7.37	8.51

3. Conclusion

In this paper, a method of port personnel recognition is proposed which can quickly and accurately locate the position of personnel and then analyze in the input image or port surveillance video. The improved Faster R-CNN detection algorithm based on ResNet-101 deep residual network and feature pyramid network is designed to adapt to the actual port application scenarios, aiming at factors such as high degree of color fusion between people and background, significant difference in illumination conditions and other conditions. Finally, the experimental results show that the framework could generally achieve satisfied performance with good real-time and accuracy in the actual port scene, which can provide certain research support for the future construction of smart port and the security application of port container terminal. In the future work, we will further improve the level of detection speed and accuracy of port personnel, especially the small target objects which are difficult to identify.

References

[1] Yau, K.A., et al., Towards Smart Port Infrastructures: Enhancing Port Activities Using Information and Communications Technology. IEEE Access, 2020. 8: p. 83387-83404.

[2] Rajabi, A., et al., Towards Smart Port: An Application of AIS Data. 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2018: p. 1414-1421.

[3] Fang, L., et al., A DPM Based Approach to Joint Object Detection and Sub-category Recognition. 2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), 2017: p. 164-168.

[4] Dange, A.D. and B.F. Momin, The CNN and DPM based approach for multiple object detection in images. 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019: p. 1106-1109.

[5] Nguyen, N., D. Bui and X. Tran, A Novel Hardware Architecture for Human Detection using HOG-SVM Co-Optimization. 2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), 2019: p. 33-36.

[6] Das, A.J. and N. Saikia, Design of pedestrian detectors using combinations of scale spaces and classifiers. Journal of King Saud University - Computer and Information Sciences, 2019.

[7] Cheng, E.J., et al., A fast fused part-based model with new deep feature for pedestrian detection and security monitoring. Measurement, 2020. 151: p. 107081.

[8] Girshick, R., et al., Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014: p. 580-587.

[9] Girshick, R., Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), 2015: p. 1440-1448.

[10] Ren, S., et al., Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017. 39(6): p. 1137-1149.

[11] He, K., et al., Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020. 42(2): p. 386-397.

[12] Redmon, J., et al., You Only Look Once: Unified, Real-Time Object Detection, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. p. 779-788.

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

[14] Liu, W., et al., SSD: Single Shot MultiBox Detector. Springer, Cham, 2016.

[15] Peng, C., K. Zhao and B.C. Lovell, Faster ILOD: Incremental learning for object detectors based on faster RCNN. Pattern Recognition Letters, 2020. 140: p. 109-115.

[16] Yin, Y., H. Li and W. Fu, Faster-YOLO: An accurate and faster object detection method. Digital Signal Processing, 2020. 102: p. 102756.

[17] Lin, T.Y., et al., Microsoft COCO: Common Objects in Context, in European Conference on Computer Vision. 2014.