Container Number Recognition Method Based on SSD_MobileNet and SVM

Tang Chun-ming[a]*, chen Peng[b]

[a]*School of Artificial Intelligence Institute, Tiangong University, Tianjin, 300387, CHINA*

[b]*School of Electronics and Information Engineering, Tiangong University, Tianjin, 300387, CHINA*

[a]*Email: tangchunminga@hotmail.com*

[b]*Email: 1285495064@qq.com*

Abstract

Aiming at how to realize the recognition of the container number on the container surface at the entrance and exit of the port, a method based on image affine transformation and SVM classifier is proposed. The main process includes truck target detection, box number area detection, text correction stage, image preprocessing stage and segmentation detection and recognition stage. Firstly, a kind of container truck detection program based on frame difference method and decreasing sequence of connected domain is proposed; secondly, a method of container number area detection based on SSD_MobileNet is proposed; in the case number recognition stage, a text correction method based on image affine transformation is proposed, and different processing methods are proposed for vertical sequence box number and horizontal sequence box number in image preprocessing stage In the stage of segmentation detection and recognition, a character segmentation algorithm based on connected domain segmentation and a segmentation detection and recognition algorithm based on SVM classifier are proposed. Through the detection and recognition of container images in the field monitoring video, the accuracy rate of regional detection can reach 97%, and the accuracy rate of character recognition can reach 95%, and it can achieve good real-time performance.

Keywords: SVM; Affine transformation; SSD_MobileNet; Container number recognition; Image processing.

1. Introduction

With the continuous development of science and technology, the global economy and trade show a rising trend.

----

* Corresponding author.

As the transmission point of economic and trade, ports bear the important responsibility of land and water transportation hub and cargo storage and transportation. Port intelligence has also become the inevitable trend of port development, and container business as an important part of port business, how to achieve its intelligence, automation operation has become an important research direction of major port companies and researchers. At present, there has been some research on automatic scheduling of port business, mainly based on different framework of automatic scheduling system, which is based on MAS framework proposed by Thurston [1] and vice [2] and PID based automatic scheduling system proposed by Li [3]. However, there is no good research on the concrete container business to realize automation. In order to better record the relevant information of the container, it is very important to identify the container number with high accuracy. At present, there are two ways to record the container number. One is to identify the container number manually by naked eyes, and then record the container number information. The other is to identify the box number by means of automatic monitoring camera and through image processing technology and machine learning or deep learning for the images in real-time monitoring video. At present, there are still many ports using manual operation for container number identification, which has many shortcomings, such as slow detection speed, high requirements for work experience, dangerous working environment, and the situation of missing and wrong inspection due to long-time work. Automatic identification and recording of container number can greatly reduce the number of workers in the container operation area, and can overcome the shortcomings of manual identification and recording of container number. As a key research subject of intelligent research, image processing is widely used in various industrial scenes due to its strong functionality of algorithm and technology. For example, itsuro [4] uses image processing technology combined with deep learning to segment characters. Machine learning and deep learning are favored by many researchers because of their intelligent and automatic performance, and with the continuous upgrading of hardware. Based on the above analysis, this paper proposes a method of container number recognition based on SSD_MobileNet and SVM classifier.

2. The Proposed Algorithm

2.1. Overall framework of the algorithm

The algorithm mainly includes three modules: Truck detection module, case number area detection module and case number identification module. The box number recognition module is divided into three stages: text correction stage, image preprocessing stage, SVM classifier design and use stage and segmentation recognition stage. The specific implementation method is to correct the text of the input box number picture first. Because of the printing problem of container number, the arrangement order of characters may not be horizontal or vertical, but have angle. Therefore, it is necessary to correct it to make it convenient for further processing. Then the corrected image is preprocessed to facilitate the next step of segmentation and recognition. After that, through collecting the container number pictures in the field monitoring video, the SVM classifier is trained by making data set. Finally, the processed images are sent to the trained SVM classifier to recognize the box number. Edge computing is a computing method that deploys computing resources to the edge of the network to reduce network latency and achieve real-time completion of some computing tasks [5]. For example, Sun [6] put forward a personnel detection scheme that adds image processing and computing power to the camera as an edge device, and has achieved good results. Due to the high real-time requirements of container number identification and recording work, in order to reduce the computing pressure of cloud computing center, this

paper adopts the idea of edge computing, and deploys the identification algorithm to the monitoring camera at the edge of the network to make it have certain computing power, so as to improve the recognition speed and save computing resources. The overall processing block diagram is shown in Figure 1.
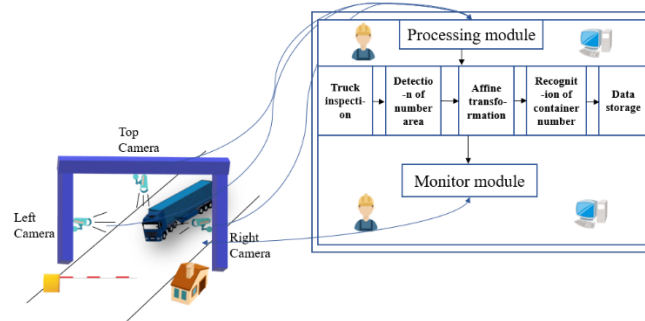

Figure 1: Overall framework of the system.

2.2. Truck detection program based on frame difference method

Containers are transported by truck into and out of the gate, and there is a short pause before monitoring for inspection by staff. As the container number is printed on the front of the container (as shown in the figure), whether the container image can be captured during this period is the key to whether the subsequent container number detection procedure can be carried out smoothly. Due to the problem of the monitoring vision of the camera in this lane, other lanes are also within the monitoring range. In order to eliminate the interference of other lanes, we intercept the scope of this lane as ROI area (as shown in Figure 2) In this region, the method of decreasing connected domain is used to judge whether the vehicle stays or not. The experiment shows that when the decreasing times is 8, the key image can be intercepted.

Figure 2: The processed area.

Set the previous frame as frame_ p. The current frame image is frame_ c. The number of processing times is num, and the statistical list of connected regions is L_ List, the decrement number is t, and the pseudo code is as follows:

Num = 0;

While (Num < 8):

    Read frame_p and frame_c；

    The two images are processed by graying, binarization and morphological processing;

    For the processed frame_ c and frame_ p do the difference operation to get the difference image frame_d；

    Count the number of connected domains in frame_d；

    If the number of connected domains decreases 8 times continuously

    Capture the current frame:

end

2.3. container number area detection based on SSD_MobileNet

When the truck stops, it is necessary to detect the area of container number and extract the part of container number to facilitate the next step of container number detection. In order to find a high-precision and fast detection method, this paper uses SSD_MobileNet model to detect the area of container number. SSD_MobileNet is a combination of lightweight MobileNet[7] and SSD [8] target detection algorithm to perform the detection task. Compared with other networks, MobileNet uses depthwise convolution, which ensures high detection accuracy and realizes low parameter and low computation. The operation process of deep separable convolution is to combine the depth convolution and stagnation point convolution (as shown in Figure 3).
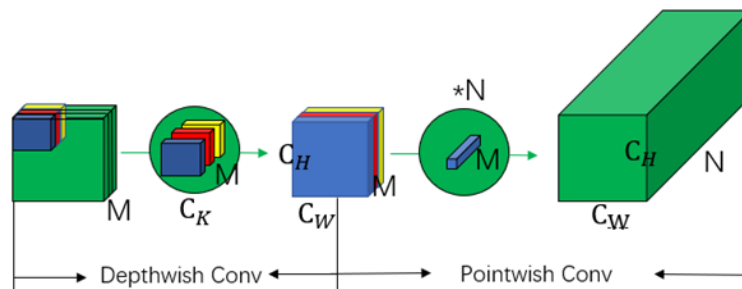

Figure 3: The structure of depthwise Convolution.

Through the calculation, we can compare the parameters and calculation of the network using deep separable convolution and ordinary convolution. The comparison results are shown in Table 1.

Table 1: The table of data volume comparison

Convolution type	Parameter quantity	Amount of calculation
General convolution	CK*CK*M*N	CK*CK*M*N*CW*CH
Depthwise convolution	CK*CK*M+M*N	CK*CK*M*CW*CH+M*N*CW*CH

The amount of calculation can be reduced to one ninth of the original. SSD algorithm is based on the idea of multi-scale feature map detection and default box, which improves the overall generalization ability of the network, and can achieve high-precision and fast detection. The SSD_MobileNet formed by the combination of the two can handle the target detection task well. Its network model is as follows:
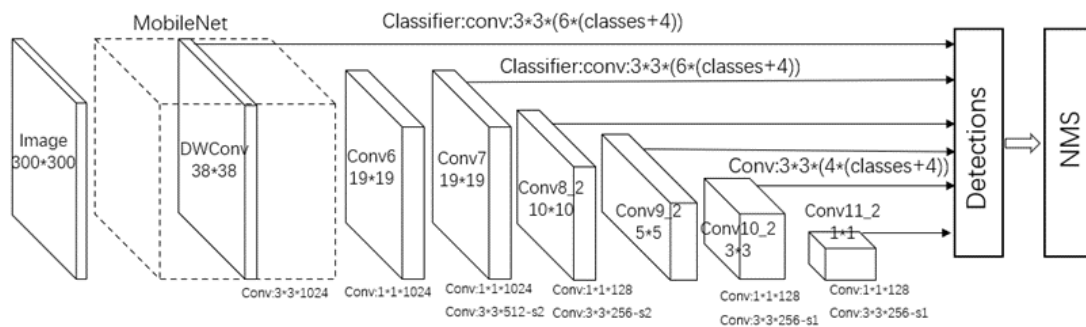

Figure 4: The structure of SSD_MobileNet

In order to detect the box number area rather than classify it, we change the category to 1, only detect the text area, and train the model through the data set. Through 30000 iterations, the loss curve and IOU curve of the model are as follows:
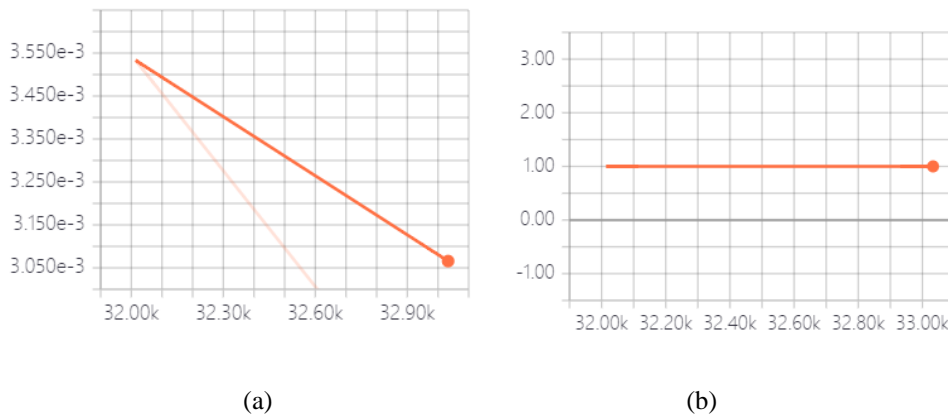

(a)   (b)

Figure 5: The related figs. (a) The loss curve;(b) The IOU curve.

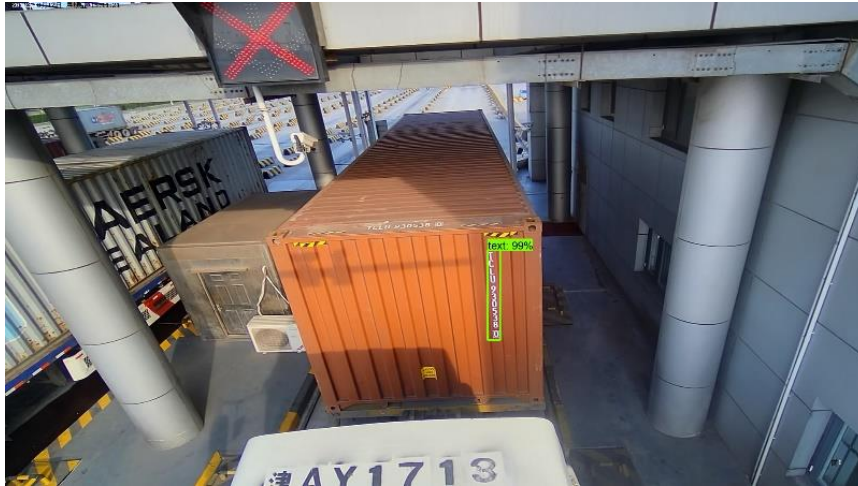So we can get the container area as follows:

Figure 6: The result

After getting the container number area, we can see that due to printing problems, not all characters are on the same horizontal or vertical line, and many of them often have deviation, so it is necessary to correct them to facilitate subsequent processing. At present, there are mainly image correction methods based on Radon transform proposed by Wu [9] and Hough transform based on Hough transform proposed by Zhou [10]. These two correction methods have certain effect on slanted text with only a certain angle and internal parallel relationship. This paper proposes an image text correction method based on affine transformation.

Figure 7: The inclined container number

2.4. Image text rectification based on affine transformation

The affine transformation of image is a spatial linear transformation or a combination of several different spatial linear transformations. The main affine transformation forms and corresponding transformation matrices can be divided into four types: translation transformation, rotation transformation, shearing transformation and zooming transformation. Let the point (x, y) be the coordinates of a point on the original image, and the point obtained after rotation is (x1, y1), and the transformation matrix M is:

$$M = \begin{bmatrix} \cos\theta, -\sin\theta \\ \sin\theta, \cos\theta \end{bmatrix} \quad (1)$$

From the transformation matrix, the relationship between the two points can be obtained as follows:

$$\begin{cases} x1 = x\cos\theta - y\sin\theta \\ y1 = x\sin\theta + y\cos\theta \end{cases} \quad (2)$$

However, the traditional affine transformation can only be transformed at the origin. In order to transform at any position, we can set the rotation angle as Ang and the scaling scale as Sca:

$$\begin{cases} \alpha = Sca * \cos Ang \\ \beta = Sca * \sin Ang \end{cases} \quad (3)$$

Furthermore, we can improve the transformation matrix to M1:

$$M1 = \begin{bmatrix} \alpha, & -\beta, & (1-\alpha)C_x - \beta C_y \\ -\beta, & \alpha, & \beta C_x - (1-\alpha)C_y \end{bmatrix} \quad (4)$$

Where C_x and C_y are the horizontal and vertical coordinates of the center point of the image, and the correction results are shown in the figure below.

Figure 8: The corrected images

2.5. Image preprocess

Image preprocessing container number can be divided into two forms: one is to print character by character vertically, the other is to print character by character horizontally. We determine the arrangement of characters according to the aspect ratio of the character area. The container number is mainly divided into three parts and 11 characters. Since the size and thickness of each character are consistent and separated from each other, we can regard these 11 characters as 11 independent units. We use the method of detecting the number of connected domains to check the noise part. Although the processing methods of horizontal box number and vertical box number are slightly different, but the ideas are mostly the same, we take the vertical box number as an example for analysis. Firstly, the number of all connected regions of the whole image is calculated as sum, and the height and width of the image are set as follow:

if sum >= 12:

The area, width and height of each connected region are calculated;

if area < 30 or w > 0.75*width or h >= (height/11):

This area is interference area, and it is filtered out;

The number of remaining areas is num1;

if num1 < 11:

The area, width and height of each connected region are calculated;

if area <= 15 or w >= 0.75*width or h >= 0.33*height:

This area is interference area, and it is filtered out；

By calculating the area, width, and height of each connected region, we can basically remove the noise and retain the character area, as shown in the Figure 9.

Figure 9: The images after pretreatment

2.6. SVM alphanumeric classifier based on hog

SVM is a model used to solve the problem of data classification, which belongs to supervised learning algorithm. Its key idea is to find the hyperplane with the max security interval on the basis of completing the binary classification. Therefore, the support vector machine (SVM), also known as the maximum margin classifier, can deal with linear separable problems well, and can be extended to nonlinear separable applications through kernel techniques. For example, the classification method of complex transformer based on SVM proposed by Dong [11]. Because the container number is composed of letters and numbers, in order to improve the recognition accuracy, this paper proposes a method of cooperation between number classifier and letter classifier. According to the location characteristics of characters, specific classifiers are used to train them. *Directional gradient histogram (HOG) is widely used in the field of image recognition by calculating the gradient direction histogram of the contract and local region of the image. Hu [12] extracted a new human*

detection feature by concatenating the histograms of all local regions of the image local Gabor amplitude binary pattern map as the feature set. In order to better train the data set, we extract the hog in the character image as a descriptor to describe its characteristics. According to the calculation method in reference [13], the point (x, y) is a point on the image, G_h (x, y) and G_v (x, y) are the horizontal and vertical gradients of the point, respectively. Then the gradient intensity Q (x, y) of this point can be written as follows:

$$Q = \sqrt{G_h(\text{x,y})^2 + Gv(\text{x,y})^2} \quad (5)$$

The gradient direction Angle (x, y) of this point can be written as follows:

$$Angle(\text{x, y}) = \arctan \frac{G_h(\text{x,y})}{Gv(\text{x,y})} \quad (6)$$

Then, the image is divided into specific regions, and the gradient histogram is calculated and L2 regularization is performed to extract the features of the segmentation region:

$$k = \frac{k}{\sqrt{\|k\|_2^2 + \varepsilon^2}} \quad (7)$$

According to the different features of the SVM, we can train them according to the different features of the SVM. Since character recognition is a non-linear classification problem, we add Gaussian kernel function to realize the function:

$$K(\text{x, y}) = e^{-\gamma\|x-y\|^2} \quad (8)$$

Where (x, y) is the sample and γ is the hyper parameter.

After setting the training batch, iteration times and related parameters, we construct a SVM number or letter classifier for model training. After the model training is completed, in order to better achieve the effect of character recognition, this paper makes corresponding treatment according to the various problems of box number printing, so as to separate all the characters and transmit them to SVM classification for recognition. After the single character recognition is finished, the recognition results will be spliced and displayed, as shown in Figure 10.

['T', 'C', 'L', 'U', '9', '3', '0', '5', '3', '8', '0']

(a)

['T', 'T', 'N', 'U', '8', '6', '4', '3', '0', '4', '3']
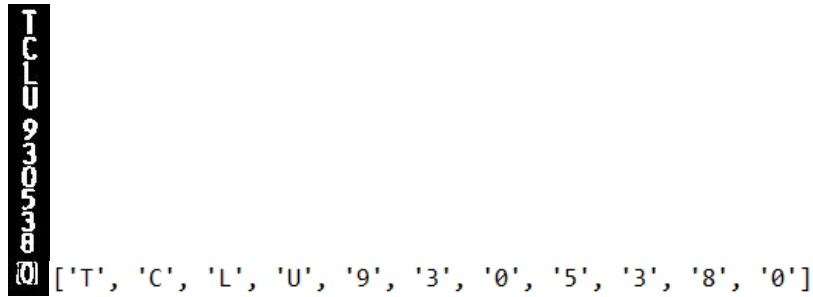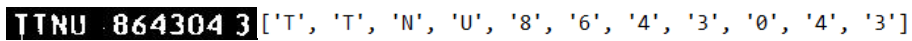
(b)

Figure 10: Recognition results of container numbers. (a) Recognition results of vertical numbers; (b) Recognition results of horizontal numbers.

3. Experimental Results and Discussion

3.1. Experimental environment

As there is no data set about container number at present, we have to intercept it from the monitoring video to get the relevant data of container number. We use Python language in programming, CPU is core i5-8300h four core processor developed by Intel company, 4GB video memory and GTX 1050ti independent graphics card.

3.2. Experimental results and analysis of the algorithm

We collected container images and container number characters from surveillance video of a port as SSD_MobileNet data set and SVM classifier data set to train them, including 912 data sets for SSD_MobileNet training. For SVM classifier, 3527 images of character and 1833 digital images are trained. When testing the model, 459 box number pictures were taken. In order to better show the accuracy of the design method, we compared with other methods in accuracy. As shown in Table 2:

Table 2: Comparison of recognition rate and recognition time of several algorithms

Algorithm name	Recognition accuracy	Time
General template matching algorithm	75.2%	0.58s
Feature weighted template matching algorithm	82.3%	0.30s
Proposed	94.6%	0.63s

In this experiment, the accuracy of box number area detection can reach 97%, while in character recognition, the recognition rate of letters is 95.6%, the recognition rate of numbers is 94.4%, and the recognition accuracy of characters as a whole is 94.6%. Through the experimental results, we can find that the recognition algorithm

proposed in this paper has greater advantages in accuracy than other algorithms, and according to the actual situation, the proposed algorithm can also meet the requirements of real-time.

4. Conclusions

In order to promote the intelligent development of port, this paper proposes a container number recognition method based on SSD combined with SVM classifier, and designs a container truck detection program based on frame difference method and decreasing sequence of connected domain, which is used to capture the key frame image of container truck parking; for container number recognition, SSD_MobileNet is used to detect container number area, and the HOG feature of image is used to train the SVM classifier. The model is tested by the pictures extracted from the port monitoring. The accuracy rate of box number area detection and container number recognition can reach 97% and 94.6% respectively. Compared with other algorithms, it has great advantages and can be processed in real time

5. Recommendations

This paper analyzes the key technology of container number identification process, and divides it into three modules: truck detection, container number area detection and container number recognition. Based on the deep learning, the algorithms of the three modules are designed respectively, which can achieve high accuracy and real-time performance in the experimental test. In the further work, we will be committed to improve the system in terms of integration to achieve better results.

References

[1]. T. Thurston and H.S Hu. "Distributed agent architecture for port automation," Proc of the 26th Annual International Computer Software and Applications Conference, pp. 81-87, 2002.

[2]. B.J. VICENT. "Multi-agent system technology in a port container terminal automation," ERCIM News, pp. 37-39, 2004.

[3]. B. Li and J.Q Yang. "A Scheduling Algorithm for Container Terminals within PID Control Framework," Journal of Transportation Systems Engineering and Information Technology, pp.124-130, 2014.

[4]. K. Itsuro, M. Mike and I. Takanobu. "SC-1 Introduction to practical AI image processing and analysis without programming," Microscopy, vol. 68, pp. i11, Nov. 2019.

[5]. S.W. Shi, X.Z. Zhang and Y.F. Wang. "Edge Computing: State-of-the-Art and Future Directions," Journal of Computer Research and Development, pp. 73-93, 2019.

[6]. J. Sun and L. Qian. "Research on Intelligent Video Surveillance Based on the Edge Computing Model, " Computer & Digital Engineering, 2019.

[7]. W. Liu, D. Anguelov and D. Erhan. "SSD: Single Shot MultiBox Detector," European Conference on Computer Vision. Springer International Publishing, 2016.

[8]. A.G. Howard, M. Zhu and Chen B. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.

[9]. K. Wu, D.P. Huang and S.J. Liu. "Skew correction of text image based on Radon transform," Automation and information engineering, pp. 17-21, 2013.

[10]. G.W. Zhou, X.J. Ping and J. Cheng. "Skew correction method for text image based on improved Hough transform," Computer applications, vol. 27, pp. 1813-1816 , 2007.

[11]. Z.L. Dong. "Fault diagnosis of power transformer based on multi-layer SVM classifier," Electric Power Systems Research, 2005.

[12]. B. Hu and C.X. Zhao. "An HOG-LGBPHS human detector with dimensionality reduction by PLS," Pattern Recognition and Image Analysis, vol. 24, pp. 36-40, 2014.

[13]. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886-893, 2005.