

Character Detection Algorithm Based on Yolov5

Changhao Lao, Weiping Hu*, Yuge Liu, Xiuqing Fan

College of Electronic Engineering, Guangxi Normal University, Guilin Guangxi, China

Abstract

Dot jet printing code on product packaging has been a difficult problem in industrial inspection due to its complex background, diverse characters and changeable fonts. An improved YOLOV5 algorithm is proposed to detect the complex background of goods. By adding segmentation and decoding tasks to the output layer, the precise location of the inkjet region is achieved, and the inkjet code is effectively separated from the complex background to obtain a pure inkjet region, which provides a simple task for subsequent recognition. At the same time, SE attention mechanism is used to make the network pay more attention to the extraction of point character features, so as to improve the accuracy of ink-jet code location and segmentation. Finally, the improved algorithm is combined with character recognition CRNN. Through the actual measurement on the production line of the food packaging factory, the character positioning accuracy is 98.7%, and the recognition rate is 98.5%, with good robustness.

Keywords

Yolov5 algorithm Target detection Segmentation decoder Attention mechanism

1. Introduction

As food safety issues emerge one after another, production date, as an important indicator to measure food safety, has become a necessary inspection object. In food production, food packaging enterprises generally use inkjet printing as the marking method of production date. In order to avoid errors or missing codes in the inkjet printing process, food suppliers must spend a lot of manpower and material resources to check. In actual production, due to the complex background color of the ink-jet code, the ink-jet code is sometimes covered by areas with similar or darker colors, which makes it difficult to identify the characters of the ink-jet code. Therefore, the task of ink-jet detection is a very challenging task. [1]

At present, character recognition algorithm based on deep learning can be divided into two steps: text detection and text recognition [2]. First, the network locates the inkjet text instance by regressing the corresponding bounding box, and then recognizes the inkjet characters in the located inkjet code area by character recognition algorithm. Among them, the v5 version of YOLO [3] series target detection not only has obvious advantages in speed, but also meets the positioning requirements in accuracy. However, in the subsequent character recognition, due to the interference of the background, it can not fully meet the requirements of character recognition.

Based on the analysis and research of existing algorithms, this paper proposes a text detection technology based on improved Yolov5, which focuses on object detection, and uses the idea of image semantic segmentation to realize the simultaneous location of the inkjet code area and the separation of its background from the inkjet code, so as to improve the accuracy of the entire system. The character detection network adopts the improved Yolov5 network. Its main innovations and contributions include: adding Yolov5 character detection network to split branch tasks, increasing attention mechanism [4], and sending it to CRNN character recognition network after location and background elimination [5] to improve character location and recognition accuracy. The experiment shows that the method has the characteristics of high precision, fast speed and good robustness. It has important significance and

ICCEIC2022@3rd International Conference on Computer Engineering and Intelligent Control
EMAIL: 583128928@qq.com (Weiping Hu)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

application value for the detection of ink-jet characters, and can be practically applied to the deployment of industrial production lines.

2.Improved Yolov5 Network

Yolov5 has excellent performance in terms of precision and speed on open source datasets. However, when applied to the commodity inkjet code inspection task, due to the diversity of commodities and the complexity and diversity of background colors, the inkjet code is submerged in the dark background, making it difficult to identify the text. To solve the problem that text cannot be recognized due to complex background interference, this paper realizes the location of ink-jet region based on Yolov5, and separates the ink-jet characters from the complex background to obtain pure ink-jet characters.

2.1. output layer is improved to two specific decoders

2.1.1. Character positioning decoder

The character positioning decoder continues the anchor based multi-scale detection scheme adopted by Yolov5. First, we use the path structure aggregation network (PAN), which is a bottom-up feature pyramid network. FPN transfers semantic features from top to bottom, combines them to obtain better feature fusion effect, and then directly uses multi-scale fusion feature map in PAN for detection. Finally, three previous anchors with different aspect ratios will be assigned to each grid of the multi-scale feature map. The detector head will predict the position offset, height and width scaling, as well as the corresponding probability and prediction confidence of each category.

2.1.2. Character Splitter Decoder

The character segmentation decoder classifies the picture pixel by pixel and judges that the pixel belongs to an inkjet character or background ^[6]. The backbone network is used to extract features together with the character location task, and the bottom layer of the FPN is fed to the segmentation branch with a size of (W/8, H/8256). After three times of up sampling, the segmentation branch restores the output feature mapping to the size of (W, H, 2), which represents the probability of each pixel in the input image to the inkjet character and background. Since the down sampling will lead to the loss of information in the feature map, the feature map and the shallow dimensional feature map will be used for channel splicing before three times of up sampling to recover the lost features.

2.2. Attention mechanism

Because the task of this paper is to detect and identify the characters that occupy several regions in the image, while the background information occupies a large part of the image, when the image is convoluted many times, the extracted information is mostly useless interfering information, and the noise information formed will cover the characteristic information of the spray, resulting in a poor segmentation effect. To this end, we added the se attention mechanism to the network, which enables the network to improve its sensitivity to inkjet informative features by significantly modeling the interdependencies between the evolving feature channels of the network, and ultimately enables the network to better segment the inkjet.

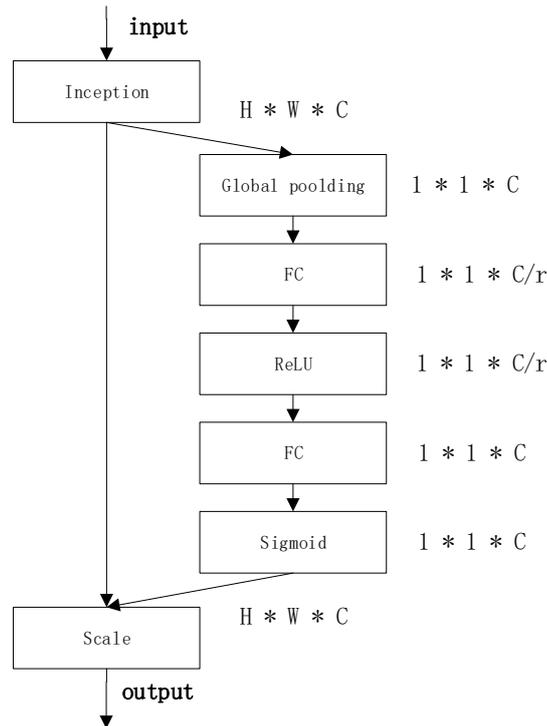


Figure 1 SE Structure

As shown in Figure 1, the SE module is mainly divided into three steps: the first step is to collect the global average value based on the width and height of the input feature map to reduce the dimension of the spatial feature to $1 * 1$. The second part uses the full connection layer to establish the connection between channels. Finally, the sigmoid activation function obtains the normalized weight. The normalized weight is weighted to each channel of the original feature map channel by multiplication to complete the re calibration of the original features concerned by the channel. The improved network structure is shown in Figure 2.

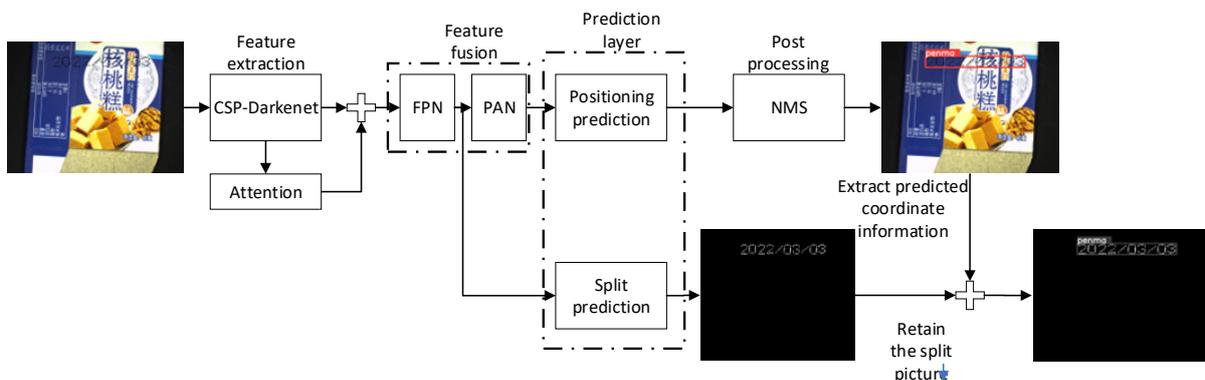


Figure 2 Improved Yolov5 network structure

3. Experiment and result analysis

3.1. Data Set Introduction

The data set used in the experiment is the food packaging box provided by Dejunwang Food Co., Ltd. in Jinjiling, Guilin. There are 12 categories, 200 pictures in each category, and 2400 pictures in total. This paper selects 2000 pieces as the training set, and expands the text detection data set to 4000 pieces through data enhancement.

The processor used in the experiment is AMD Ryzen 5 3500X, the graphics card is NVIDIA GTX 2070s, and the operating system is Windows10, 64 bit. The whole experiment uses python 1.8+python 3.7 to build the network model.

3.2. model training

Train and test the improved Yolov5 text detection network and CRNN text recognition network. The text detection network experiment optimization algorithm uses the Adam optimizer. The initial learning rate is set to 1E-4, the batch size is 8, and the epoch is 100.

3.3. Experimental results and analysis

In order to verify the improved YOLOV5 inkjet text detection method in this paper, experiments were conducted on 500 test datasets to evaluate the overall performance of combining CRNN with improved YOLOV5 network text detection. The evaluation index of text recognition rate is the accuracy rate by comparing the predicted text with the real text. If four characters correctly predict three characters, the accuracy of text prediction sequence is 75%.

Table 1 shows the ablation experiment results of the improved Yolov5 text detection algorithm used to separate the position and background of the inkjet code. In this paper, a segmented decoder and an improved loss function are added to YOLOV5 network to enable it to locate the ink-jet characters and separate the ink-jet characters from the background. The SE attention module is added, and the recognition accuracy is improved by 4.9 percentage points.

Table 1 Ablation experiment of improved YOLOV5 algorithm

Network	check accuracy	Division precision	Recogni accuracy
Yolov5+ decoder +CRNN	97.2%	86.2%	93.6%
Improve Yolov5+CRNN	98.7%	90.3%	98.5%

The left figure of Figure 3 shows the image to be detected of the input network, and the right figure of Figure 3 shows the output result of the improved Yolov5 algorithm network.



Figure 3 Improved Yolov5 network output

In order to verify the effectiveness of the improved Yolov5 algorithm in inkjet detection of complex backgrounds, several common character detection schemes are compared, and the results are shown in Table 2. The common methods do not remove the background interference, resulting in insufficient final recognition accuracy. In this paper, the improved Yolov5 algorithm improves the positioning accuracy by 2.8%, and the improved algorithm realizes the segmentation of inkjet code and background at the same time. The algorithm obtains pure inkjet code area, and the recognition rate is significantly improved.

Table 2 Comparison of Common Character Algorithms

Network	check accuracy	Division precision	speed/ (fps)
ABCNet	— —	96.0%	5.0
FOTS	— —	93.5%	14.0
CTPN+CRNN	95.8%	92.1%	4.5
Yolov5+CRNN	95.9%	87.6%	32.0
Improvem Network	98.7%	98.5%	16.0

After comprehensive consideration, the improved Yolov5 algorithm effectively combines the tasks of semantic segmentation and object detection to obtain an ink-jet region without background. Compared with the code spraying area obtained by unmodified YOLOV5, the character recognition accuracy using CRNN network is significantly improved. The final test results show that the character location accuracy and recognition rate are 98.7% and 98.5% respectively.

4. Conclusions

In order to solve the problem of ink-jet detection of packaging boxes in food production, this paper proposes an improved algorithm based on Yolov5 ink-jet character segmentation detection, adds semantic segmentation decoder to the output layer of the network, and realizes the combination of semantic segmentation and target detection tasks. At the same time, combined with the attention mechanism, enhance the extraction of dot characters and ink dot features, effectively improve the segmentation of ink-jet characters and background, and locate the ink-jet region. The clean character area obtained by the improved Yolov5 algorithm can greatly improve the character recognition rate. This algorithm can effectively detect the ink-jet area on the food packaging box. Next, we can try to carry out more effective feature extraction for small targets, so as to optimize the segmentation between inkjet characters and background, obtain more pure character regions, and further improve the recognition rate of characters.

5. ACKNOWLEDGMENTS

Fund project: Guangxi Postgraduate Education Innovation Project (XYCSZ2021002); Chinese National Natural Science Foundation (61966004)

6. References

- [1] Wu Huiying, Chen Ming, Fan Yanjun, Huang Shuai. Research on lattice character recognition in complex background [J]. *Computer Applications and Software*, 2021, 38(09):146-152.
- [2] Wang JX, Wang ZY, Tian X. Review of natural scene text detection and recognition based on deep learning. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(5):1465–1496.
- [3] Qian X Q, Liu W F, Zhang J and Cao Y. Underwater-relevant image object detection based feature-degraded enhancement method. *Journal of Image and Graphics*, 27(11) : 3185-3198.
- [4] Zhou B, Li C H and Chen W. 2021. Region-level channel attention for single image super-resolution combining high frequency loss. *Journal of Image and Graphics*, 26(12):2836-2847.
- [5] Duan B, Fu X, Jiang Y and Zeng J X . 2020 . Lightweight blurred car plate recognition method combined with generated images . *Journal of Image and Graphics*, 25(09) : 1813-1824.
- [6] Bi XL, Lu M, Xiao B, Li WS. Pancreas Segmentation Based on Dual-decoding U-Net. *Ruan Jian Xue Bao/Journal of Software*, 2022, 33(5): 1947–1958.