A method of positioning and recognition of electronic scale characters based on deep learning

Zhengguang Xu¹, Lutian Wang¹, *, Siqiang Niu¹, Guojin Kan²

¹School of Automation and Electrical Engineering, University of Science & Technology Beijing, Beijing, 100083, China
²Science and Technology on Space Physics Laboratory, China Academy of Launch Vehicle Technology, Beijing, 100076, China
35359812@qq.com
4s20180592@xs.ustb.edu.cn

Abstract. Digital characters are widely used in fields such as industry, equipment display and detection. According to the application background of reading recognition of electronic scale in chemical experiment examination of middle school, in order to improve the application of artificial intelligence in the education field, meet the requirements of intelligent examinations, and save manpower, a character positioning and recognition method based on the combination of deep learning and traditional image processing is proposed, which can automatically and quickly identify the reading of digital tube display screen of electronic scale, and according to the test requirements and test sites can automatically determine whether the students' weighing operation is correct. Firstly, the yolov3 target detection algorithm and deeplabv3plus semantic segmentation algorithm are used to locate the character area of the display screen, and then a single digital tube character is obtained by projection segmentation. Finally, the recognition algorithm based on support vector machine is used to recognize characters. The experimental results show that the method has a high accuracy of positioning and recognition, and can adapt to different lighting, digital tube size and character tilt angle, has certain anti-interference ability, meets the real-time requirements, and has high application value.

1. Introduction

With the development of science and technology, the management means of various industries are gradually changing from manual management to automatic or semi-automatic management. In addition to industry and other fields, automation in the education field is also an inevitable trend. Because of its high precision, flexible setting and other advantages, the digital tube is widely used in the display field of all walks of life. In the application environment of the traditional high school chemistry experiment examination to judge whether the students use the electronic scale for weighing and reading, teachers still need to manually monitor the digital tube value display and judge whether the students read correctly. Due to the need to rely on the monitoring of teachers in use, schools need to call a large number of teacher resources for the monitoring of the examination, and a teacher needs to monitor the operation process of multiple students, so that teachers can not accurately obtain each student's experimental operation in the electronic scale reading data, but also can't achieve a comprehensive monitoring, which eventually leads to the misjudgment. With the development of artificial intelligence and digital image processing technology, the requirements of intelligent...
examination become higher. The application of automatic recognition of digital tube reading technology is particularly important. Therefore, the use of digital image processing technology and recognition technology can realize the automatic recognition of digital tube reading, and automatically determine whether students operate correctly according to the examination requirements and examination points. It can not only save human resources in invigilation, but also ensure the accuracy of examination judgment and students’ scores.

In order to solve the above problems, this paper proposes a method to locate the character region by using the depth neural network instead of using the traditional image processing method. In the case of meeting the real-time requirements, the accuracy and adaptability are greatly improved. The method of support vector machine is used for character recognition, and the method of feature extraction and the diversity of samples are improved. Compared with neural network, the algorithm complexity is lower, and the amount of training data is small, and the recognition speed is fast. Compared with other methods such as thread starting and threading method, the recognition accuracy is high.

2. Character Area Positioning of Electronic Scale Display Screen

2.1. YOLOv3 target detection algorithm positioning electronic scale

Because of the high resolution of the collected image, the display area is segmented directly according to the characteristics of the display area, which has a large amount of calculation and a large number of interference areas affect the segmentation. Therefore, the method of locating the character area of the electronic scale display screen in this paper needs to locate the electronic scale area from the whole image, and then segment the display area in this area, which compared with directly locating the character area of the display screen from the entire image, it reduces unnecessary interference areas, reduces related features, and increases the accuracy and integrity of subsequent segmentation of display area.

In this paper, we use the YOLO algorithm based on deep learning, in which YOLO algorithm regards object detection as a regression problem[1], and divides the image into S×S squares according to the characteristics of the whole picture. Each grid is only responsible for detecting the target which the center falls on the grid. Each grid calculates the probability value of the target category, the position of the edge box and the confidence of the edge box respectively, and realizes the prediction of the target category probability, boundary box, and confidence in all areas in one step. YOLO framework network structure is shown in the Fig.1.

![YOLO framework network structure](image)

YOLO is connected to the fully connected layer after the convolutional layer, so that only the highest layer feature map is used during detection, which makes the detection effect of small objects poor. The way of grid division makes it a high rate of missed detection in a multi-target environment. However, in this application scenario, only a single target of the electronic scale needs to be detected, and the electronic scale is not a small target, which is obviously different from the background. In this way, the accuracy of the electronic scale target detection using YOLO algorithm is also high when
meeting the real-time requirements. In this paper, we use the improved YOLOv3 algorithm, which improves the detection accuracy as much as possible on the basis of ensuring the detection speed. The following steps are used to carry out the YOLOv3 target detection to detect whether there is an electronic scale on the experimental table in each frame of the camera.

Step 1: Cycle through the original video frame of webcam in a process.
Step 2: Get the original video frame and call YOLOv3 model for detection in another process.
Step 3: Use the current frame image of the detected scale for subsequent processing to obtain the regional image and position coordinates and other information. If the electronic scale is not detected, the next frame will be detected directly. The effect picture of positioning the scale area image is shown in Fig.2.

![Fig.2 Positioning result image of electronic scale](image)

2.2. DeepLabv3plus semantic segmentation algorithm for character segmentation of display screen

The character area of the electronic scale that needs to be located is the display area. Because the area obtained by YOLOv3 algorithm is rectangular area, and in order to segment the image of single character area without interference, the rectangular area obtained by locating the display area directly by using YOLOv3 algorithm can’t accurately obtain the display character area. Therefore, in order to obtain the accurate area of the display screen, this paper adopts a method of image semantic segmentation based on DeepLabv3+ algorithm. This method based on deep convolution neural network can accurately complete the classification and recognition of image pixels, so as to obtain the connected domain of the target, and can accurately determine the target category. DeepLabv3+ algorithm adopts the encoder decoder mode similar to FCN(full convolution neural network), and its schematic diagram is shown in Fig.3.
Fig.3 Deeplabv3+ schematic diagram

In the encoder part, Xception network is taken as the skeleton network[5]. Its network structure is composed of a series of deep separable convolutions, residual connections similar to Resnet and some other conventional operations[6]. ASPP module is introduced into Xception network, which can capture information at multiple scales and realize robust segmentation. The input image uses the feature extraction network to generate a feature map that is 16 times smaller than the original image, and then the feature map is input into a 256 channel 1*1 convolution layer. Finally, the convoluted feature map is input to the decoder. In the decoder part, we use the hop connection method of FCN to connect the low-level features and the high-level features. First, we use 48 channel 1*1 convolution to convolute the low-level feature map to reduce the number of channel in the feature map, and then merge it with the high-level feature map after 4 times bilinear interpolation sampling. Finally, after performing 3*3 convolution operation, it is restored to the original image resolution by 4 times of bilinear interpolation to realize the segmentation of the same size as the original image.

Because the semantic segmentation is aimed at the classification of pixel level, it needs the pixels of the whole image, and the parameters input during training contain more information, so the semantic segmentation directly in the original video frame cannot meet the real-time requirements. Therefore, this paper uses the Deeplabv3+ algorithm to segment the local image after the electronic scale is located by YOLOv3 algorithm, Fewer pixels in the partial image make the algorithm processing faster, and fewer pixels for classification lead to fewer similar interference areas and higher accuracy. Use the following steps to execute the Deeplabv3+ semantic segmentation algorithm to segment the display area of the image with electronic scale in each frame.

Step1: Using Deeplabv3+ semantic segmentation in the electronic scale area located by YOLOv3 to obtain segmentation results such as the display screen.

Step2: Convert the local segmentation result graph into the original video frame resolution size result graph, and judge whether the display area is really detected according to the connected region size.

Step3: If the display area is divided, the position information of the connected area of the display screen is obtained, so as to locate the character area of the display screen of the electronic scale.

The effect pictures of the local image after positioning the electronic scale by YOLOv3 algorithm and the segmented image of the display screen area are shown in Fig.4 and Fig.5.
3. Image Preprocessing

3.1. Get the original image of character area of display screen
Because the image of display screen is binary after semantic segmentation, only the position information of the display screen area, the original color image area corresponding to the connected area of the display screen needs to be taken out to get the image containing character information. For subsequent character recognition.

Semantic segmentation is the classification of pixel level. Because the edge pixels of the display screen are different from the internal pixels of the display screen, the boundary pixels will not be segmented. Therefore, the segmented display area is smaller than the actual display area. As a result, each character image may be incomplete in subsequent segmentation of a single character image, which will affect the recognition. Before obtaining the original color image of the display character area, in this paper, firstly, the rectangle structure element with size of 7 is used to expand the scope of the connected region of the binary image, and then the largest connected region is taken as the real connected region of the display screen, obtain the coordinate information and find the coordinate information of the smallest outer boundary to complete its positioning.

Finally, the pixels of the original color image corresponding to the connected region of the expanded binary graph are taken out to form a color image, which is used for subsequent character segmentation. The color image is shown in Fig.6.

3.2. Use affine transformation for tilt correction
The electronic scale placed by the student during the experiment is not completely facing the camera, and it will appear to be placed on the desktop obliquely, which will cause the character area of the display screen in the image to be tilted, and the subsequent standard single character image cannot be
segmented for character recognition. Therefore, it is necessary to perform tilt correction on the positioned image of the character area of the display screen before character segmentation.

Due to the distortion of the camera, the actual rectangular display area, the two sides of the image may be parallelograms, and the direct rotation method will still cause the characters to tilt. Therefore, the affine transformation is used to correct the image in this article. Affine transformation maintains the "straightness" (straight line is still a straight line after affine transformation) and "parallelism" (the relative positional relationship between the straight lines remains unchanged, the parallel lines are still parallel lines after affine transformation, and straight lines The position sequence of the upper point will not change) of Two-dimensional graphics, so the display area with parallel opposite sides is still parallel to the opposite side after being transformed into a straight rectangle, so the tilt correction can be completed.

According to the principle of affine transformation, 6 equations are needed to solve the 6 variables of the affine transformation matrix \( T \), that is, 3 sets of points, so select 3 points in the original image, and 3 points in the target image after transformation. The affine transformation matrix \( T \) can be obtained by the equation. The original image can be transformed to the target position according to the mapping method of the affine transformation to obtain the transformed Images. Namely:

\[
\begin{bmatrix}
    x'_1, x'_2, \ldots, x'_n \\
    y'_1, y'_2, \ldots, y'_n
\end{bmatrix} =
\begin{bmatrix}
    a_1 & a_2 & t_x \\
    b_1 & b_2 & t_y
\end{bmatrix}
\begin{bmatrix}
    x_1, x_2, \ldots, x_n \\
    y_1, y_2, \ldots, y_n \\
    1, 1, \ldots, 1
\end{bmatrix}
\]

Therefore, according to the four vertices of the minimum circumscribed boundary rectangle calculated by the connected domain, three points are selected as the three origin points on the color image of the original pixels in the character area of the display screen, that is, the three points before correction, and then three points parallel to the upper and left boundaries of the image are selected as the target points, so that the character image of the display screen after correction is no longer inclined, The horizontal distance is the long side distance of the smallest bounding rectangle, and the vertical distance is the short side distance of the smallest bounding rectangle. In this way, the corrected image will not be deformed. The correction results are shown in Fig.7.

Since the three original points before correction are the three vertices of the minimum external boundary rectangle of the character area of the display screen, the rectangular area formed by the three target points after transformation also corresponds to the minimum external boundary rectangle of the character area of the display screen. Therefore, the rectangular area composed of the three target points after transformation is the corrected character area of the display screen. At this time, the precise positioning of the character area of the display screen is completed. The image after positioning is shown in Fig.8

4. Character Segmentation Based on Projection Method

4.1. Binary image

In order to recognize characters, the image of a single character must be extracted from the image of the character group. The character recognition method used in this article needs to input a single binary image obtained after segmentation into the recognition engine for recognition, so the character group image needs to be binarized before character segmentation.

Since the digital tube characters in the display area are black, the background gray value of the character area is higher, which is quite different from the digital tube character gray value. You can directly use the OTSU maximum inter-class variance method to obtain a better binarized image. Based on the prior knowledge of the size of the display area, the height and width of the display area are...
generally 32, 200, and the electronic scale operated by students is opposite to that of the camera. Therefore, before binarization, the image needs to be normalized to (200, 32) and rotated 180 ° to facilitate subsequent character segmentation. The final binarized character area image is shown in Fig.9.

### 4.2. Character segmentation based on projection method and connected domain analysis

This paper presents a method of character segmentation based on projection method and connected domain analysis, which can effectively solve the influence of non-character noise and adhesion on character segmentation.

The steps of character segmentation algorithm based on projection method and connected domain analysis are as follows:

**Step1:** Analyze the connected components of the binary image to obtain the number, number of pixels, and position information of each connected component. Assuming that the height and width of the image are h and w, remove the connected domain with too small width, the connected region close to the upper and lower boundary but less than 4 in height, the connected region close to the left and right boundary but less than 15 in width, the connected domain with width greater than 30 but less than h/4 in height, and the connected domain with the leftmost point on the left edge of the image and the connected domain with the rightmost point on the right edge of the image, which can effectively remove non-character noise.

**Step2:** Calculate the horizontal projection value of each line from the top to the bottom of the binary image processed in the previous step. For the projection value greater than the threshold value 5, the line is considered to contain characters. Then, the starting position and ending position of the longest continuous line whose horizontal projection value meets the threshold conditions are calculated, and the image is intercepted according to this position, that is, the upper and lower boundaries of the character area are accurately taken out in the character area of the display screen, and the redundant background is effectively removed.

**Step 3:** Calculate the height of the intercepted image to judge whether the characters are adhered to the upper and lower edges of the display screen. If adhesion occurs, the image height after the second step of interception will still be very high due to adhesion, which is greater than a certain threshold value. According to the difference between the start position and the end position of the run and the upper and lower boundaries calculated in step2, the upper or lower adhesion can be judged, and then the empirical value can be taken to intercept the image according to the test results for many times. For the upper adhesion, the image in the [h-1-23, H-1] line interval is intercepted, For the bottom adhesion, the image in the [h-1-23, H-1] line interval is intercepted, which can solve the problem of the characters adhering to the upper and lower edges of the display screen.

**Step4:** Since the electronic scale characters are digital tube characters and standard digital characters, there is no adhesion between characters. There must be a background gap between characters. Therefore, it is enough to use the projection of binary image in the vertical direction to find the vertical tangent line. However, due to the large tilt angle of the electronic scale or the camera distortion, the character pixels of the original image after the initial affine transformation are deformed. although there is a gap between characters, the character itself is inclined so that the vertical projection can not find the tangent line. Therefore, this paper uses the second vertical projection to segment the character. The first vertical projection result from left to right is used to judge whether the character is slanted. If it is slanted, the second skew correction is needed, and then the vertical projection is used to segment the character. If the character is not slanted, the first vertical projection is used to split the character directly.

The method of vertical projection to find the dividing line is to traverse the vertical projection value from left to right. If the projection value of the column is 0 and the projection value of the next column is greater than 2, then the column is considered to be the left boundary of the character. The projection value of is greater than 2, and the projection value of the next column is equal to 0, then the next column is considered to be the right boundary of the character, and the left and right boundary of the
character can be retained as the coordinate of the character's dividing line, and one character will correspond to two divisions line, and the final remaining split lines will form a pair, and the x coordinate difference between the two dividing lines of each pair is the width of a single character.

If the calculated width of a single character is not greater than the empirical value of 25, the second skew correction is not needed, and each pair of segmentation lines is used to intercept the current character region image to form a single character image.

If there is a pair of cutting lines with a width greater than the empirical value of 25 or less than two pairs of cutting lines, it is considered that there is no background gap between characters, and if no tangent line is found, it is considered that the character is deformed and needs to be corrected again. The correction method is also affine transformation, looking for three points in the current binary graph as the origin. According to the characteristics of character skew, when the character tilts from left to right, find the leftmost point of the upper boundary of the character, the leftmost point of the lower boundary of the character and the rightmost point of the lower boundary of the character as the three origin points. When the character tilts from right to left, find the leftmost point of the upper boundary of the character, the leftmost point of the lower boundary and the rightmost point of the upper boundary of the character are used as the three origin points, and the three target points are the upper left corner, lower left corner and lower right corner (slanting from left to right) or upper right corner (tilting from right to left) of the image processed in step3. According to the method mentioned in this paper, the secondary correction can be completed. Then the vertical projection is used to find the tangent line again to form a single character image.

Finally, a single character image is obtained by removing the upper and lower background regions by horizontal projection, so as to obtain a character image that does not contain the background up, down, left, and right. The character image after the above processing is shown in Fig. 10.

5. Character Recognition Based on Support Vector Machine

5.1. Principles of Support Vector Machine
Support vector machine (SVM) is a supervised learning model [7], which is usually used for pattern recognition, classification and regression analysis. The basic idea of SVM is to map the input data to a high dimensional inner product space through a nonlinear mapping, in which the generalized optimal classification surface can be obtained. At the same time, by using the kernel function, all necessary calculations are carried out in the input space, in which the generalized optimal classification surface is obtained. By training the training set to find such a hyperplane, the samples are classified to ensure that some points in the classified samples are close to the hyperplane at the maximum distance (geometric interval) from the hyperplane. Finally, the hyperplane trained can be used to classify the samples. The classification decision function under the condition that the classifier is linearly indecomposable is as follows:

\[ F(x) = \text{sgn} \left( \sum_{i=1}^{l} a^*_i y_i K(x_i, x) + b^* \right) \]  

Currently, commonly used kernel functions include: Gaussian Radial Basis Function (RBF); polynomial kernel function; Sigmoid kernel function, etc.
5.2. Character recognition based on LIBSVM

In this paper, the nonlinear classification algorithm of SVM is applied to character recognition. Because the character recognition of electronic scales in this article is to classify ten numbers from 0 to 9, the number of classes is small, so the One-against-One multi-class support vector machine method is selected for character recognition. In this process, I used libsvm-3.23 support vector machine library[8]. The algorithm includes two aspects: training and recognition. The following two aspects will be introduced respectively.

Training process:
(1) A large number of single character images can be obtained through the character positioning and segmentation algorithm in this paper. Because the size of the single character image is different, it needs to be normalized to the same size. According to the experience, the single binary character image is normalized to $24 \times 32$.

(2) The single character image data is extracted to form a feature vector. This paper proposes a feature extraction method that combines gray-scale features and internal pixel statistical features, the method of feature extraction is to first select the gray value of each pixel position in the image as the feature value, that is, the pixel values of each binary image are arranged in order from the beginning and the end of the line to obtain a $24 \times 32 = 768$-dimensional feature vector. Then, two coarse grids are divided into the character image, and the number of character pixels contained in the two grids is calculated to obtain two eigenvalues, which are combined to form a 770 dimensional line vector. This combination method is not only simple and fast, but also does not require the image quality as strictly as other related systems, but also takes into account the information of each position of the characters, so as to avoid over reliance on the extracted structural features of characters. It can distinguish some characters with similar peripheral features but big differences in internal features.

(3) Obtain feature vectors of all binarized character images of the same character according to the above method, and classify them into a feature vector set; give each feature vector set A unique label (all feature vectors in the feature vector set have the same label as the feature vector set), so that it corresponds to the character one to one;

(4) The feature vectors formed by all samples in step3 are transformed into text data of specified format and input into support vector machine for training. This paper uses C-SVC support vector machine model and RBF function as the kernel function of support vector machine to train the input training data, cross check these data, find a set of good penalty factor C and kernel function parameter $\gamma$[9], and use these parameters to build a SVM model.

Recognition process:
(1) Perform the same normalization operation on the character image to be recognized.
(2) A 770 dimensional feature vector is formed according to the same feature extraction method.
(3) The 770 dimensional feature vector of the input data is predicted by using the model obtained in the training process, and a label is given to the group of data. Through the one-to-one correspondence model between the label and the character category built in the training process, the character image classification can be judged, that is to say, to recognize the character image of a single electronic scale.

6. Test Results and Discussions

In this paper, we have carried out experiments on the mentioned algorithms of character location, segmentation and recognition.

A total of 4770 electronic scale samples were labeled in YOLOv3 model training, including samples under different illumination conditions and tilt angles. The momentum constant was 0.9, the learning rate was dynamic attenuation, the initial value was 0.0001, the attenuation step size was 40000, the decay rate was 0.0005, the batch size was 64, the maximum iteration was 80000 times, and the framework was Darknet. Finally, the loss value of the model was stable below 0.02, and the mAP (average accuracy of multiple verification sets) reaches more than 98%. In the actual 10 test videos, 7700 electronic scale images and 2120 images without electronic scale, the positioning accuracy rate is more than 98%, and there is no false detection.
In the training of Deeplabv3+ model, a total of 1000 display samples were labeled, including samples under different illumination conditions and samples with different digital tube values. Because the actual segmentation is carried out in the target frame image of the electronic scale, the resolution of the training samples is not fixed. In terms of data enhancement, the input image is randomly scaled by a ratio of 0.5 to 2 and the image is randomly flipped during the training process. The training category has only 1 display foreground category and 1 background category, and the initial learning rate is 0.002. The mIoU of the final model (calculating the ratio of the intersection and union of the two sets of true and predicted values) reached about 0.9. Take the image of the electronic scale located by Yolo for actual test. All the targets on the display screen are segmented, and the segmentation accuracy is about 80%. For the samples without accurate segmentation, only the boundary area of the display screen is not accurately covered, but the characters on the display screen are still completely included, which does not affect the subsequent character segmentation.

In this paper, the character segmentation method proposed in this paper can effectively segment a single character when there is no hand covering the display area and the non-standard value caused by the change of characters on the display screen. The accuracy of correct processing reaches 96%, and the segmentation result is stable. The processing results of character positioning and character segmentation are shown in Table 1.

| Test image          | Processing correctly | Processing correctly% | Precise processing |
|---------------------|----------------------|-----------------------|--------------------|
| YOLOv3 positioning  | 9820                 | 9525                  | 98%                |
| Deeplabv3+ segmentation | 9525               | 9240                  | 98%                |
| Character segmentation | 9429               | 9051                  | 96%                |

Table 1. Statistics of character location and segmentation results

In the character recognition using libsvm, in the image processed by the positioning and segmentation algorithm in this article, for each type of character, 200 character sample images of varying quality are selected for training, and the parameters c and γ are tested with different values. Finally, it is found that the effect is better when C = 1000, γ = 0.0013. Finally, from more than 7000 correctly located and segmented character sample images, the character images with different quality are selected for testing, and the character recognition rate of electronic scale using SVM method is obtained. In addition, this paper also compares the two algorithms of BP neural network and template matching to compare the effect. The experimental results are shown in Table 2. It can be seen from table 2 that the recognition rate of the SVM based character recognition method is significantly higher than that of the template matching and BP neural network. Template matching requires the corresponding relationship between points, and the anti-interference ability is not strong. When the character segmentation is not accurate enough, it is difficult to ensure the corresponding relationship, and the recognition rate is not high. In the learning process of BP neural network, it is easy to produce "over learning" problems, which will affect the recognition rate. The SVM classification algorithm can effectively avoid the problems caused by the above methods. In a word, the character recognition of electronic scale based on support vector machine classification algorithm is more effective, the recognition rate is higher, and has high robustness.

| Test image                | Template matching | BP neural network | SVM       |
|---------------------------|-------------------|------------------|-----------|
| High quality character segmentation | 88.33%            | 93.34%           | 96.23%    |
| Low quality character segmentation | 81.24%            | 91.17%           | 92.33%    |

Table 2. Recognition rate of character reading of electronic scale
7. Conclusion
According to the practical needs of intelligent examination, in order to automatically judge whether the reading of electronic scale is correct in the middle school chemistry experiment examination, this paper proposes a method of electronic scale character reading location and recognition based on the combination of deep learning and traditional image.

Firstly, the method uses the YOLOv3 target detection algorithm and the Deeplabv3+ semantic segmentation algorithm to realize the character area positioning of the electronic scale, which has strong accuracy and adaptability, and meets the real-time requirements.

Character segmentation based on projection method and connected domain analysis can effectively eliminate noise and adhesion problems.

In character recognition, this paper uses SVM method, which is different from the previous classification methods. It focuses on minimizing the structural error of the training set, that is, through a quadratic programming problem, it gives an upper bound of the error recognition rate on the basis of minimizing the error rate, so as to obtain a good adaptability and promotion ability. The result of SVM training is a set of support vectors. SVM obtains the classification results determined by the samples related to the support vector. The other samples are useless in classification, but in the neural network algorithm, the statistical information of all samples will be used. It is a feasible way to introduce SVM into character recognition of electronic scale. Especially in the case of small sample learning, SVM has good performance. The experimental results also show that the proposed method has better recognition rate, stability and real-time performance, which verifies the feasibility and accuracy of electronic scale reading recognition.

References
[1] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. CVPR 2016, 2016.
[2] Redmon J, Farhadi A. YOLOv3: an incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.
[3] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018:801-818.
[4] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J]. Computer Science, 2014(4):357-361.
[5] Chollet F. Xception:Deep learning with depthwise separable convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii:IEEE, 2017: 1251-1258.
[6] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.
[7] Burges C J C. A Tutorial on Support Vector Machines for Pattern Recognition. Knowledge Discovery and Data Mining, 1998,2(2).
[8] Chang Chih-Chung, Lin Chih-Jen, LIBSVM: a Library for Support Vector Machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[9] Huang Zhibin, Chen Duansheng. Application of Support Vector Machine in License Plate Character Recognition[J]. Computer Engineering, 2003, 29(5):192-194.