



IMAGE RECOGNITION TECHNOLOGY BASED ON DEEP LEARNING IN AUTOMATION CONTROL SYSTEMS

JINGJING WANG*

Abstract. In order to solve the problem of recognizing multiple product images, the author proposes the research of image recognition technology based on deep learning in automated control systems. Firstly, the FasterRCNN method is improved by proposing a non class specific FasterRCNN, which can be used for pre annotation of product images by training only on publicly available datasets. Due to the use of position correction networks, the pre annotation effect is more accurate than that of candidate region networks. Then, combining Grabcut with non class specific FasterRCNN, a sample enhancement method was proposed to synthesize a large number of training images containing multiple products and use them for model training. In addition, based on non class specific FasterRCNN, a re identification layer was proposed to improve detection accuracy. In the end, the recognition and positioning of multiple products achieved a recall rate of 93.8% and an accuracy of 96.3%.

Key words: Supermarket product image recognition, Deep learning, Data annotation algorithm, Non class specific FasterRCNN

1. Introduction. In the design process of a product, in order to stimulate consumer purchasing desire while ensuring product recognition, the product image features are very rich. Common image features can be divided into two categories: One is the low-level visual features, which are the global and local features of the product, another type is the intermediate semantic features of the product, which are mainly applied in the process of product recognition based on the underlying visual features. That is to say, the image features of products can be specifically divided into the following types, including color features, shape features, texture features, point features, semantic features, and other aspects [1]. Among these features, color feature is the most direct way. By using color histogram, the color distribution of the product image can be obtained, while shape regions and contour boundaries can be distinguished. Through specific detection algorithms, the feature points of the product can be effectively extracted. Deep learning technology is a perception technique built on neural networks, commonly including neural models, perceptrons, BP algorithms, convolutional neural networks, etc. Taking convolutional neural networks as an example, they are currently the most widely used type of neural network and have been widely used in facial recognition, speech recognition, license plate recognition, object detection, and other fields [2, 3]. Convolutional neural networks and BP neural networks are similar in that they both consist of an input layer, an intermediate layer, and an output layer. However, compared to the former, the intermediate layer is more complex. In practical applications, this neural network has enormous advantages, especially in network depth and massive image data processing. The most important thing is that using this neural network can better complete learning and training. After determining the specific neural network, deep learning can be carried out, utilizing massive data and network models for learning. Feature learning in the data is a very important content, which can ensure the accuracy of model prediction is improved.

In recent years, due to the rapid development of mobile internet, with the support of cloud computing and big data, online shopping like e-commerce has sparked a new wave of people's lifestyles [4]. However, due to the gradual saturation of online users, transaction profits have also gradually decreased. More and more companies are turning their attention to the offline trading platform. The layout of offline transactions by companies such as Alibaba, Meituan, and Ele.me is evident. The retail industry has also become a focus of offline transaction competition. Enterprises are increasingly focusing on how to use artificial intelligence technology to transform product production, circulation, and sales, reshape the industry ecosystem, and integrate online and offline

*Department of Information Engineering, Yangzhou Polytechnic College, Yangzhou, Jiangsu, 225000, China (yzwj2023@163.com)

experiences. That is to say, future innovative retail may be entirely achieved through artificial intelligence technology. In the current era of booming artificial intelligence algorithms, researching retail product recognition technology to improve productivity has become one of the hotspots in the field of artificial intelligence. In the past two years, some automated retail stores have emerged both domestically and internationally, such as AmazonGo, Taobao Coffee, etc. The use of artificial intelligence technology to achieve automation and unmanned retail scenarios has become a trend and is gaining momentum. Nowadays, artificial and intelligent technologies are using algorithms to process information in order to reach the level of humanity. As is well known, humans are best at using vision and hearing to receive, process, and transmit information. In the process of purchasing goods in supermarkets, people have a clear understanding of the products they want to purchase and can visually locate and identify them. In the settlement process, it is still common to use the cashier to scan the code one by one for settlement. With the development of artificial intelligence algorithms in the field of computer vision in recent years, machines can recognize universal objects like humans, such as roads, pedestrians, vehicles, etc. However, in specific tasks, such as the settlement process of supermarket products, how to use image vision to identify items, in order to liberate productivity, improve automation and intelligence, has become a research hotspot in the field of artificial intelligence. For the recognition and localization of multiple products, the author improved the regression layer of FasterRCNN and fully utilized publicly available datasets to learn object bounding box regression knowledge for pre labeling of product images, which is more accurate than the labeling algorithm of RPN. Then, a combination of pre labeling algorithm and Grabcut algorithm, an image synthesis algorithm, is proposed to solve the recognition and localization problem of multiple products, and its working principle is described in detail [5].

2. Methods.

2.1. Problem Description. In recent years, deep learning has been used in many fields. Such as facial recognition, object detection, etc. Deep learning can capture useful information from a large amount of data, and due to the advent of the big data era and the improvement of computing device performance, the application of deep learning has become a reality. FasterRCNN and its extended versions have become one of the most effective methods in recent years. However, if the above methods are transferred to a new task, a large amount of calibration data needs to be used to readjust the model on the new task. However, in practical scenarios, data calibration is a very difficult task that requires significant financial, material, and human resources.

The author proposes a method to solve the data bottleneck problem, which can detect and locate products without the need for border calibration. The dataset constructed by the author contains only a single product in the training images without border calibration, while the test images contain multiple products [6]. The author first improves FasterRCNN by proposing a non class specific FasterRCNN, and combines transfer learning to pre calibrate the training data; Then, combined with the unsupervised Grabcut2l method, the training data is sample enhanced to generate realistic training images of multiple objects; Then train the non class specific FasterRCNN to detect multiple objects; Finally, the author proposes a re identification method based on FasterRCNN, adding a re identification layer to FasterRCNN to improve the accuracy of multiple object detection.

2.2. System Framework Design. The product recognition and localization method proposed by the author can be divided into three modules: The first part is the bounding box calibration module of non class specific FasterRCNN. The second part is a sample enhancement method that combines non-specific FasterRCNN with Grabcut [7]. The third part is a region based re identification method. The first part, the proposal of non-specific FasterRCNN, is aimed at solving data barriers and fully utilizing prior knowledge in public datasets to solve the calibration work of borderless calibration training samples. On the one hand, it removes the impact of complex redundant backgrounds. On the other hand, it provides border interaction input for Grabcut, and at the same time, it can complete border prediction of product areas without training, effectively solving the problem of product positioning. In the second part, by combining Grabcut and non-specific FasterRCNN methods, accurate region extraction is achieved by effectively utilizing the pre calibration information of borders. Based on the extracted regions, fuse the training images of the products to generate a large number of images of multiple products. By using the generated images for overall model training, it is possible to distinguish between cases of product occlusion or overlapping borders, effectively solving the problem of multiple product

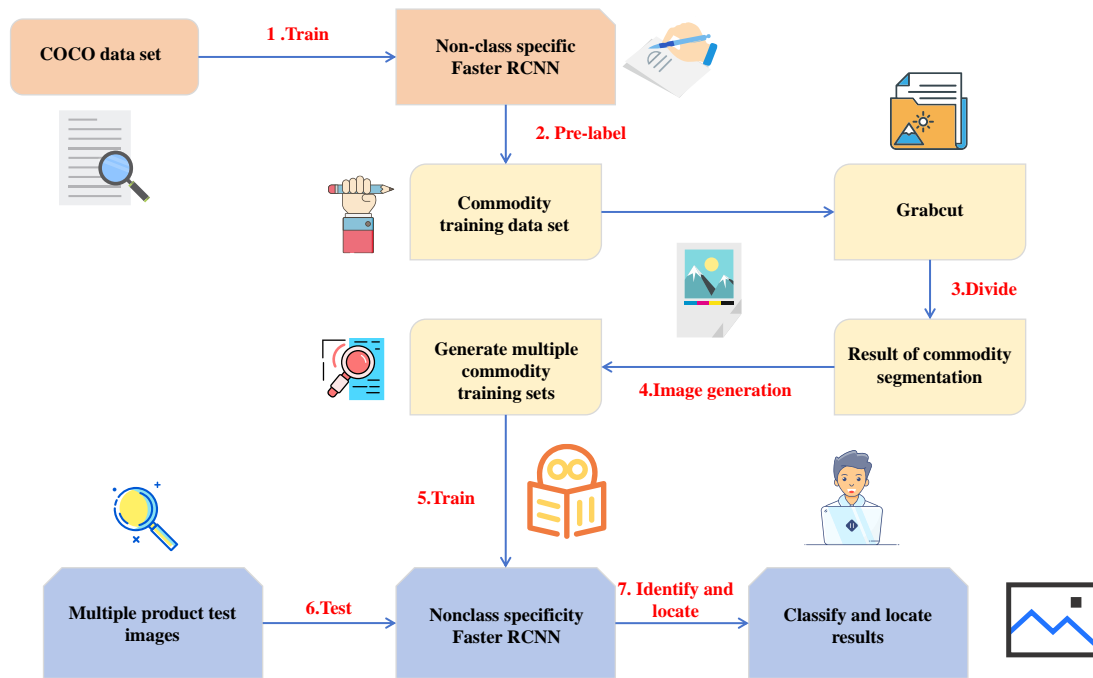


Fig. 2.1: Block diagram of multiple product identification and positioning systems

neighbors or occlusion. Realized the recognition and localization of multiple product images based solely on a single product training image calibrated without borders. In the third part, in order to improve the accuracy of system recognition, a region based re identification method is proposed to solve the problem of misidentifying objects and improve the prediction accuracy of the system. As shown in Figure 2.1.

2.3. Specific System Implementation. Combining Grabcut's sample enhancement method: The non class specific FasterRCNN solves the problem of bounding box regression, making the pre calibration of product positions in training images more accurate. When there is only a single product in the training data, and there are multiple products in the test images, the classification problem still has significant difficulty. Even if a single product can be used to train a classification model, multiple products in the test image may have overlapping or even occluded borders. Due to the lack of product overlap in the training images, training the model will make its classification ability less robust and insufficient to recognize products with overlapping or even occluded borders, increasing the difficulty of classification. Therefore, the author proposes a sample enhancement method to generate training images with multiple products by processing a single product training image. Because using category non-specific FasterRCNN, the borders of individual product images can be obtained. A direct idea is to extract the border part of the product in the image, rotate or translate it, and combine it with other products. However, this method will result in the background area of the product border covering other product areas, and the edge area of the border will differ significantly from the actual image. Therefore, using only the borders of product images is not enough to achieve realistic sample generation. If precise area information of the product can be obtained, such as the product object mask, the background area can be separated to solve the problem of the product being obscured by the background in the generated sample. Therefore, the author uses the Gabcut method to segment the products in the training images.

Both Grabcut and Graphcutsl methods are interactive image segmentation methods. Graphcut needs to provide precise foreground and background pixel seed regions during interaction, calculate the similarity between other pixels and foreground and background, and use graph theory algorithms to calculate the optimal segmentation. The Grabcut algorithm has less user interaction and only needs to provide a rectangular border containing the foreground. Then, the pixels inside the border are used as the foreground, and the pixels outside

the border are used as the background. Gaussian Mixture Model (GMM) is used to model the background, and graph theory algorithm is used for segmentation. The GMM background modeling and graph theory algorithm are repeatedly used until the iteration converges, completing the segmentation.

After utilizing the non class specific FasterRCNN proposed by the author, the rectangular border of a single product in the training image can be obtained. Therefore, it is only necessary to combine the Grabcut algorithm to segment the precise region of the product. Then, the individual product areas in the training set are randomly rotated and translated, and randomly combined to generate training images for multiple products. It is worth noting that, considering the accuracy of the data, products cannot be completely covered between each other, because if the products are excessively covered, it will lead to almost all the real products in the area being covered, and the products occupying a large area will not match the actual labels, which will mislead the training of the recognition model. Therefore, when combining randomly, it is necessary to constrain the overlapping area of the products, assuming that the upper limit of the overlapping area is Sup . Consider three fusion strategies: (1) Perform random rotation and translation, only constraining the upper limit of the overlapping area, that is $sc \leq sup$, which means that during fusion, the product may be far away, at which point $sc=0$. (2) While limiting the upper limit of overlapping area, constrain the lower limit of overlapping area, that is $sc>0$. This approach requires overlap between products, ensuring a close distance between them without extensive coverage. (3) Increase the constraint on the lower limit of overlap area, that is $sc \geq sm$, in order to increase the possibility of overlap between products and differentiate the overlapping situation through model training.

The re identification layer found that the model trained by the above method will have a small number of adjacent bounding boxes predicting two different results, one of which is correct and the other is incorrect. Because FasterRCNN is a two-level (two-stage) method, in the first stage, the RPN (candidate region network) first filters out candidate regions and filters out a portion of the background region; The second level uses the head network to finely classify candidate regions, while correcting the borders of each candidate region, known as border regression. Obviously, the candidate regions extracted by RPN are imprecise, which can affect the accuracy of head network recognition. Because some candidate regions only cover local areas of the object, although bounding box regression can correct them to the correct position, directly using candidate regions for prediction may inevitably lead to misidentification. Therefore, the author proposes a re identification layer to improve the accuracy of FasterRCNN recognition. Because the bounding box position is more accurate after passing through the bounding box regression layer of the head network, the bounding box regression layer here is a non class specific regression method proposed by the author. Moreover, the classification layer of the head network filters out a large portion of the background area. Therefore, the results of bounding box regression can be classified first, and then more accurate bounding box regression results can be used as candidate regions, followed by some re identification. The author will use the precise regions obtained from the head network regression, combined with the ROIAlign method, as inputs to pass through the classification layer of the head network again. Traditional FasterRCNN can be defined as:

$$\begin{cases} c = f_{cls}(roi) \\ reg = f_{reg}(roi) \end{cases} \quad (2.1)$$

Among them, roi represents the candidate regions generated by RPN, f_{cls} represents the classification layer, f_{reg} represents the regression layer, c represents the classification result corresponding to the candidate region, and reg represents the bounding box regression result of the candidate region. The added re identification layer selects the regions classified as non background in the candidate regions, with the background category represented by 0. Then, its regression border is used as the new candidate region for classification and regression, represented as:

$$\begin{cases} c' = f_{cls}(reg_{c \geq 0}) \\ reg' = f_{reg}(reg_{c \geq 0}) \end{cases} \quad (2.2)$$

Among them, c' represents the final classification result after re identification, and reg' represents the final bounding box regression result.

Table 2.1: Basic parameter settings for non class specific FasterRCNN

base_lr	0.001
momentum	0.9
weight_decay	0.0001
ROI number	2000
ROI positive and negative sample ratio	1:3
mini_batch	1
Step size	1000
max_iter	300

Network Model and Training. The author proposes a non-specific regression layer to improve the class specific regression layer of the original FasterRCNN, forming a non-specific FasterRCNN model. I hope to learn border regression knowledge from public datasets and directly apply it to pre annotation of individual product training images.

Firstly, the original FasterRCNN is trained on COCO, and the Resnet model in its backbone network is pre trained using ImageNet. Then, the classification layer and regression layer are trained, and finally, the overall network model is jointly trained [8]. This is done to enable the model to learn effective feature generalization ability from the COCO dataset. Change the bounding box regression layer in the trained FasterRCNN model to the non class specific regression layer proposed by the author, while keeping the parameters of other parts unchanged, and only train the non class specific regression layer on the COCO dataset. For the new non class specific FasterRCNN, tune the entire network using the COCO dataset. This is to enable the features obtained by ROIAlign to balance the ability of classification and bounding box regression.

The model trained through the above steps can be directly used for annotating product training images. RoiPooling/ROIAlign performs feature extraction in the corresponding area of FeatureMap, and normalized to 7×7 in size, with 256 channels. Then followed by an 7×7 convolutional layer, where an 7×7 convolutional kernel of the same size as the input is used, with an output dimension of 1×1 and a channel count of 1024, the convolutional layer here can be considered as a fully connected layer. Then use the convolutional kernel of 1×1 , with an output dimension of 1×1 and a channel count of 1024. The convolutional layer of this layer is also an alternative to the fully connected layer. Moreover, the output dimension is the same as the output dimension of the previous layer. This is generally done to enhance the non-linear ability of the network model, as each convolutional layer is followed by an activation function, where Relu is used as the activation function. The last layer is the softmax layer, which outputs 4 channels, which is the position parameter for bounding box regression. In addition, the hyperparameters in the model are shown in Table 2.1.

By combining non class specific FasterRCNN with Grabcut, a large number of product image samples can be generated. And used for training the overall model. The training steps are as follows:

- (1) Firstly, based on the parameters of the non class specific FasterRCNN used for pre calibration of training samples, training is carried out while keeping the backbone network and non class specific regression layer parameters unchanged, and only training the classification layer model.
- (2) Then, while keeping the parameters of the non class specific regression layer unchanged, train the classification and regression layers of the RPN network, as well as the classification layer in step (1).
- (3) Train the entire network, including the Resnet parameters in the backbone network, with only fixed non class specific regression layers. This is because the features in the backbone network are trained by COCO, and in order to better extract features from product data, it is necessary to train the backbone network parameters.

Through the above steps, the model trained using the generated samples can be used for the detection task of real product images. Moreover, the non class specific regression layer does not need to be retrained on the target dataset, which further confirms its knowledge transfer ability.

In the experiment, the MaskRCNN method from the FasterRCNN algorithm set was used, which, in addition to the FasterRCNN method, utilized the Feature Pyramid Network (FPN) and Region of Interest Alignment (ROIAlign) methods. During the training process, there was no need for segmentation prediction,

Table 3.1: Comparison of Properties of Different Datasets

Data set	Number of categories	Average training samples for each class	Does the training sample contain multiple targets	Is there a location label
COCO	80	25000	Yes	Yes
VOC	20	1350	Yes	Yes
data set	40	8	No	No

so the segmentation branch in MaskRCNN was removed and only its classification and regression branches were used. The experiment was conducted on two NVIDIA TITAN X GPUs. The initial learning rate is 0.001 and is manually adjusted during training. The momentum parameter Momentum is 0.9. Among them, conv1, conv2, conv3, conv4, and conv5 are components of the Resnet network structure.

3. Experiments and Analysis.

3.1. Dataset Introduction. The author validated the proposed method on the constructed product dataset. The training set consists of training images for a single product. Using the author's method, there is no need to train bounding box regression on the product dataset, so the constructed product dataset training images only contain category information [9]. The training images were captured from four different perspectives using two cameras, with each image containing only one product object. There are a total of 40 product categories, each of which includes 8 training images. The expanded single product image training set contains a total of 3200 training images. The test set consists of multiple product images captured using another camera, and the product positions and angles in the images are diverse. The test set consists of 400 test images, totaling 40 product categories. The proposed non class specific FasterRCNN was trained on the COCO dataset and directly applied to pre labeling of product training image data. The COCO dataset consists of 80 categories, including a large number of images, borders, and category annotations. The main difference between the product dataset constructed by the author and the COCO dataset is that the objects in the product dataset are rotated, and the training data is much less than the COCO dataset. In addition, the dataset constructed by the author only contains a single product in the training images and does not require border calibration.

As shown in Table 3.1, the differences between this dataset and the publicly available dataset were compared. Among them, there are as many as 40 categories of the dataset, and the objects can be rotated, which adds some difficulty to the recognition task. More importantly, the dataset has very few category samples and is a single object sample, without position border labels, greatly increasing the difficulty of localization and recognition tasks.

3.2. Experimental Results.

(1) *Sample Enhancement Strategy Analysis.* After extracting the product border from the training image, the Grabcut algorithm is combined to segment the product area. Because the training image contains a large area of background, if the Grabcut algorithm is directly used to segment the original training image, the segmentation effect is very unsatisfactory. Because there is no border to calibrate the background area of the image, the outermost pixel of the image is generally taken as the background. However, its area is very small, making it difficult to model the entire background. Combining the author's proposed non class specific FasterRCNN pre labeling algorithm with Grabcut algorithm for product image segmentation in the training set. Then use simple image processing methods to generate training images of multiple products for the FasterRNN model training.

When using the Grabcut algorithm for training image data generation, the upper limit of object overlap area sup is set to 10000. The author compared the performance of the trained model on the testing machine for training data generated by different overlapping areas sc. As shown in Figure 3.1, when generating training images, when the overlap area is 0, that is, when the product distance is far, the effect is not good. Because the product distance is far, it is difficult to see folding in the training data, making it difficult for the network to receive training on folding, so the testing effect is relatively poor. When the overlapping area is 6000, the recall and precision of the model reach 93.8% and 96.3%, respectively, and the testing effect is the best. When

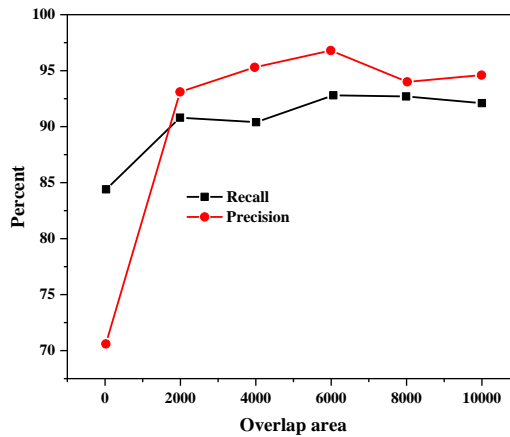


Fig. 3.1: Detection results of different fusion strategies

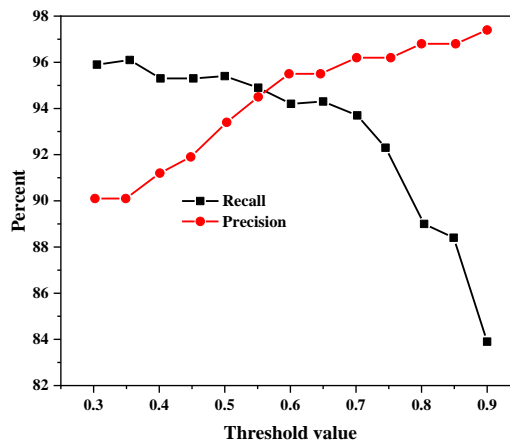


Fig. 3.2: Detection performance of different probability thresholds

the overlap area is too large, the large coverage between products in the training image tends to mislead the network into misidentification and reduce the testing effect.

(2) *Analysis of the effects of each part of the model.* When the model identifies and locates, it will output the probability of its corresponding category for each region. When arranging the model, it is usually necessary to normalize the probability, filter out predictions with low probability, and retain results with high probability. Therefore, the impact of different thresholds on the recall and accuracy of the model was analyzed, as shown in Figure 3.2. In general, the higher the probability threshold, the higher the accuracy, and the lower the recall. The lower the probability threshold, the lower the accuracy, and the higher the recall rate. In Figure 3.2, when the probability threshold is 0.3, the model can achieve high accuracy and recall at the same time. This is because the model has a high probability of predicting categories, and a low threshold has little effect on it. The model has strong predictive ability. In order to balance accuracy and recall, the author determined a probability threshold of 0.7, which is a recall rate of 93.8% and an accuracy of 96.3%.

As shown in Table 3.2, by analyzing various parts of the model, the proposed sample enhancement method combined with Grabcut improved the detection recall by over 40% and accuracy by 30%. In order to improve the accuracy of multiple product detection, the author proposes a re identification layer, which corrects the candidate regions after classification and regression through the bounding box regression layer and inputs them

Table 3.2: Analysis of the effectiveness of the methods proposed by the authors

Method	Recall (%)	Precision (%)
Non class specificity FasterRCNN	58.14	50.96
Non class specificity FasterRCNN +Grabcut	90.43	92.76
Non class specificity FasterRCNN++Grabcut + Re identification	93.81	96.3

Table 3.3: Performance Comparison of Multiple Product Identification Methods

Method	Recall (%)	Precision (%)
SIFT	33.47	20.30
VGG16	41.21	29.38
VGG19	36.50	26.25
Xception	58.50	42.50
Resnet	58.92	43.75
Author's method	93.80	96.29

into the classification layer again. After correction by non-specific bounding box regression layers, classification errors caused by imprecise candidate regions can be effectively avoided. When using the re identification layer, the recall rate is increased by 3% and the precision rate is increased by 4% compared to when not using the re identification layer.

(3) *Comparison of detection models.* Because the proposed non class specific FasterRCNN can detect a single product border, the main problem when applied to the detection of multiple products is that it can interfere with recognition when there are areas of other products within the product border. The bounding box regression of FasterRCNN, which is not specific to category, is not affected by multiple products. Therefore, when using training data from multiple products generated by the author for training, only the parameters of the classification layer are trained while keeping the parameters of the non class specific regression layer unchanged.

Through the proposed image enhancement technology, the detection of multiple products has been achieved, and the non class specific regression layer is only trained on public datasets, and regression knowledge has been learned. Moreover, there is no need for further training when transferring to product image detection. Multiple products can also be well positioned in situations with local occlusion, different perspectives, or across scenes. However, there is no perfect system in the world, and when there is severe occlusion, missed detections can also occur. Among them, the product "Hanshan Pepper and Pickled Vegetable" was not detected due to being partially occluded. The author quantitatively validated the proposed method in the constructed product dataset. Due to the author's intention to address data bottlenecks. The constructed training dataset only has category labels and no border calibration. In this case, traditional image detection methods generally use unsupervised features to calculate local features of the retrieved image and perform similarity matching with the features of the image in the training set. The currently most effective deep learning methods, such as VGG16, VGG9, Xception, and Resnet, are generally regarded as multi label classification tasks for recognition.

The author compared these methods and as shown in Table 3.3, the performance of SIFT and other currently optimal deep learning methods is significantly lower than the method proposed by the author [10]. On the one hand, SIFT does not distinguish background features, which leads to background features affecting the matching effect; On the other hand, it is an unsupervised artificial feature that is not as effective as supervised methods in recognition, and the product packaging will have severe reflection, which also results in lower feature performance. Other deep learning methods can be extended from a single product training image to multiple product training images. Not learning the distinguishing information when multiple products are similar, and also not distinguishing background features, resulting in low recognition rate. Some deep learning models, such as VGG16 and VGG19, have similar performance to SIFT, this is because the cross task recognition task of training images from a single product to identifying and locating multiple products results in low performance

of deep learning models. And this method proposes a sample annotation and sample enhancement method that does not require training on the target dataset. It can use the training image of a single product to learn the distinguishing information of multiple products, serving as a bridge across tasks and greatly improving performance.

4. Conclusion. The author designed and implemented multiple product recognition and localization methods based on improved FasterRCNN and GrabCut. Regarding the above issues. Firstly, the author cleverly improves FasterRCNN by proposing a non class specific bounding box regression layer that can learn prior knowledge of object positions using only public datasets for training, without the need for retraining on the target dataset, and uses it for pre calibration of object positions in commodity training data. Secondly, based on the pre labeled position information of non class specific FasterRCNN products, combined with the GrabCut interactive segmentation method, accurate segmentation can be performed on the training images of individual products. By fusing individual product images from multiple categories, a large number of training images containing multiple products and labels for product positions are generated for model training. At the same time, it solves the problems of limited data volume, no object position annotation, and the expansion of single product image recognition tasks to multiple product recognition tasks.

REFERENCES

- [1] B. Wu and C. Zheng, Pattern recognition of holographic image library based on deep learning, *Journal of Healthcare Engineering*, 2022 (2022), p. 1–9.
- [2] J. Gao, An image recognition method for speed limit plate based on deep learning algorithm, *International Journal of Information and Communication Technology*, 20 (2022), p. 216.
- [3] T. Takayama, T. Yashiro, S. Sanada, T. Katsuragi, and R. Sugiura, Evaluation of detection accuracy of image recognition for automatic counting of rice planthoppers captured on sticky boards, *Agricultural Information Research*, 30 (2022), p. 174–184.
- [4] J. Xu, S. Zhou, F. Xia, A. Xu, and J. Ye, Research on the lying pattern of grouped pigs using unsupervised clustering and deep learning, *Livestock Science*, 260 (2022), p. 104946.
- [5] S. Liu, R. Hu, J. Wu, X. Zhang, J. He, H. Zhao, H. Wang, and X. Li, Research on data classification and feature fusion method of cancer nuclei image based on deep learning, *International Journal of Imaging Systems and Technology*, 32 (2021), p. 969–981.
- [6] L. Wang, Q. Kou, Q. Zeng, Z. Ji, L. Zhou, and S. Zhou, Substation switching device identification method based on deep learning, in *2022 4th International Conference on Data-driven Optimization of Complex Systems (DOCS)*, 2022, pp. 1–6.
- [7] W.-x. WANG, J. JIA, K. GUI, and J. GAO, Research on automatic compliance checking method based on bim and ontology technology for architectural drawing, *Journal of Northeastern University (Natural Science)*, 43 (2022), pp. 1346–1353.
- [8] P. Rani, S. Kotwal, J. Manhas, V. Sharma, and S. Sharma, Machine learning and deep learning based computational approaches in automatic microorganisms image recognition: Methodologies, challenges, and developments, *Archives of Computational Methods in Engineering*, 29 (2021), p. 1801–1837.
- [9] Z. Wang, H. Chen, Q. Zhong, S. Lin, J. Wu, M. Xu, and Q. Zhang, Recognition of penetration state in gtw based on vision transformer using weld pool image, *The International Journal of Advanced Manufacturing Technology*, 119 (2022), p. 5439–5452.
- [10] J. Sun and X. Yuan, Application of artificial intelligence nuclear medicine automated images based on deep learning in tumor diagnosis, *Journal of Healthcare Engineering*, 2022 (2022), p. 1–10.

Edited by: Zhigao Zheng

Special issue on: Graph Powered Big Aerospace Data Processing

Received: Jan 4, 2024

Accepted: Feb 26, 2024