Group-housed pigs and their body parts detection with Cascade Faster R-CNN

Deqin Xiao[1,2], Sicong Lin[1,2], Youfu Liu[1,2], Qiumei Yang[1,2], Huilin Wu[3*]

(1. *College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China*;
2. *Guangdong Provincial Key Laboratory of Agricultural Artificial Intelligence (GDKL-AAI), Guangzhou 510642, China*;
3. *Guangzhou National Modern Agricultural Industry Science and Technology Innovation Center, Guangdong, 510642, China*)

Abstract: The detection of individual pigs and their parts is a key step to realizing automatic recognition of group-housed pigs' behavior by video monitoring. However, it is still difficult to accurately locate each individual pig and their body parts from video images of groups-housed pigs. To solve this problem, a Cascade Faster R-CNN Pig Detector (C-FRPD) was designed to detect the individual pigs and different parts of their body. Firstly, the features were extracted by 101-layers Residual Networks (ResNet-101) from video images of group-housed pigs, and the features were input into the region proposal networks (RPN) to obtain the region proposals. Then classification and bounding box regression on region proposals were performed to get the location of each pig. Finally, the body parts of the pig were determined by using the Cascade structure to search on the feature map of the pig body area. These operations completed the detection of the whole body of each pig and its different parts of the body, and established the association between the whole and parts of the body in the end-to-end detection. In this study, 1500 pig pen images were trained and tested. The test results showed that the detection accuracy of C-FRPD reached 98.4%. Compared with the Faster R-CNN without cascade structure, the average detection accuracy was increased by 4.3 percentage points. The average detection time of a single image was 259 ms. The method in this study could accurately detect and correlate the individual pig with its head, back, and tail in the image. This method can provide a technical reference for recognizing the behavior of group-housed pigs.

Keywords: group-housed pigs, body parts detection, Faster R-CNN, Cascade structure
DOI: 10.25165/j.ijabe.20221503.6286

Citation: Xiao D Q, Lin S C, Liu Y F, Yang Q M, Wu H L. Group-housed pigs and their body parts detection with Cascade Faster R-CNN. Int J Agric & Biol Eng, 2022; 15(3): 203–209.

1 Introduction

Under the large-scale intensive pig breeding mode, monitoring pig behavior is important to ensure the health of pigs and improve the quality of pork[1-4]. Video monitoring is a low-cost and easy-to-implement method, combined with computer vision technology that can effectively extract pig information and conduct deeper analysis. To analyze the behavior of group-housed pigs, it is necessary to identify individual pigs first[5-9]. Afterward, research on complex behaviors such as drinking water, eating, and mounting can be carried out. The recognition of these behaviors depends on the analysis of pigs' heads, backs, tails, and other parts[10-14]. Therefore, it is essential to study the detection algorithm of the group-housed pigs' different parts of their bodies.

In recent years, there have been many researches on pig detection algorithms based on video images. Most of the research focuses on the detection of individual pigs, but there are few studies on the detection of pig body parts. The detection of pig body parts is often processed by traditional image processing algorithms in the detection system. Kashiha et al.[10] detected the pig head by analyzing the overall trend of the curve from each point of the pig's body contour to the centroid of the pig's body image. Nasirahmadi et al.[11] used the intersection points between the major axis of the fitting ellipse and the body contour to find the position of head, tail, and sides in pigs, but did not distinguish the head and tail. Guo[13] detected the individual pig by the SVM algorithm by extracting color, texture, shape, and other features, and used the improved generalized Hough transformation and clustering algorithm to distinguish the head and tail of the pig. Such methods rely on artificial features to extract individual contours of the pigs. It is difficult to deal with complicated conditions such as the change of light intensity, body posture, and target occlusion. If the contour of a pig cannot be extracted accurately, the detection of pig body parts will not work correctly.

Different from the above methods, many researchers also applied deep learning for pig behavior recognition[15-17]. Alameer et al.[18] trained a CNN to recognize the feeding and foraging behavior. Yang et al.[19,20] trained a detector based on the Faster R-CNN to detect pigs and their heads, and then recognized the feeding behavior of pigs. This method applied abstract features that have stronger image representation capabilities and achieved high detection accuracy[21-24]. However, the pigs and their heads were detected separately by this method. It needed to find out which pig the head belongs to, which was easy to cause association error. Therefore, further study is needed to achieve an efficient and reliable location of pig body parts.

The purpose of this study was to establish an association

Received date: 2020-11-22 **Accepted date:** 2021-10-09
Biography: Deqin Xiao, PhD, Professor, computer network, research interest: wireless network technology, wireless networks and mobile computing, Email: deqinx@scau.edu.cn; **Sicong Lin**, MS, research interest: computer vision, Email: 710230897@stu.scau.edu.cn; **Youfu Liu**, PhD candidate, research interest: computer vision and intelligence agriculture, Email: lyf0313@126.com; **Qiumei Yang**, PhD, Lecture, research interest: computer vision and intelligence agriculture, Email: yqmbegonia@163.com.
*Corresponding author: **Huilin Wu**, Engineer, research interest: intelligence agriculture. Room 816, Building 2, No. 135, Xianlie East Road, Tianhe District, Guangzhou 510642, China. Tel: +86-13640889880, Email: 85132973@qq.com.

between the whole body of each pig and its body parts. For this purpose, this study proposed an improved Faster R-CNN algorithm C-FRPD, which is based on the depth abstract feature[25-27] and a Cascade structure. C-FRPD aimed at detecting individual pigs and their body parts by monitoring images of group-housed pigs and proposing an end-to-end detection algorithm for the association between individual pigs and their parts.

2 Data collection and data set production

The experimental data came from the video data collected on a pig farm in Luogang District, Guangzhou Province, China. The total length of the videos was about 2880 h. The videos were shot from above. The monitoring scene of the videos was a pig pen with 4 pigs. The pen size was 4 m×5 m. The pigs in the videos belong to the same species, Landrace. They had similar body sizes and colors. There were letters on the back of pigs to identify each pig. The data contained video images of day and night. The images at night were black-and-white, taken by the infrared night vision function of the camera. In this study, only RGB images taken during the day were selected as the research object. The image resolution was 2560×1440 pixels, and the scene is shown in Figure 1.

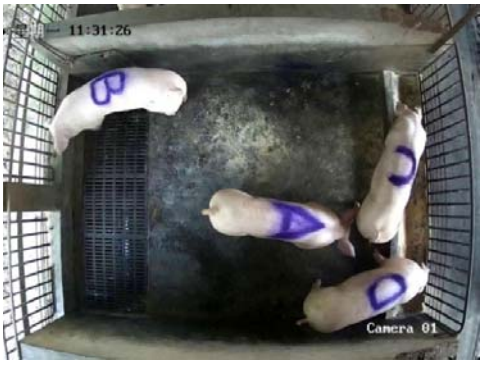


Figure 1 Example of video images of the experimental pig pen

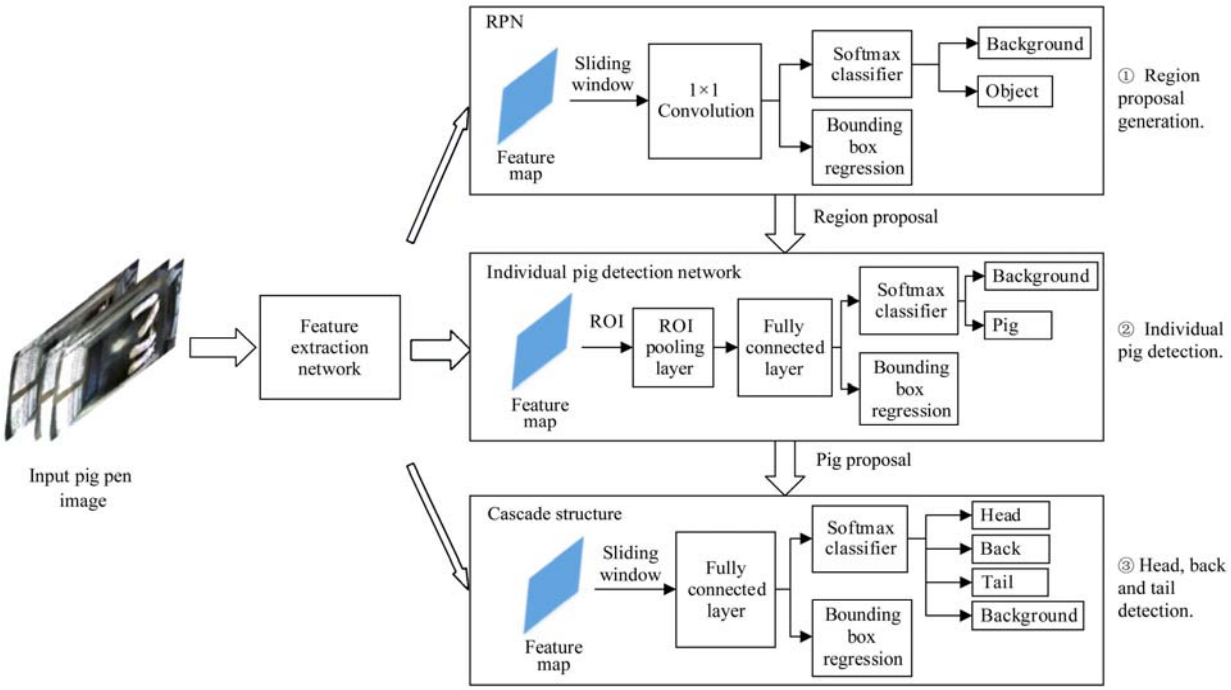
In order to train the C-FRPD and verify the accuracy of the algorithm, sample frames were selected randomly from multiple daytime videos to construct a data set. Each video was extracted with 10-20 images, and the interval between the extracted images was longer than 20 min. In order to make the dataset representative, the diversity of images was taken into account when selecting samples. It contained samples of pigs in different postures, such as standing, lying on their side, and bending over the ground. In addition, the data set contained samples of complex conditions such as pig adhesion. In this study, 1500 pen images were manually labeled with the data labeling tool LabelImg. There were four pigs in the experimental pen, so a total of 6000 pigs were marked. In order to provide a data set for the detection of pig body parts, the head, back, and tail of each pig were labeled. In the experiment, 1500 images were used to construct three data sets for training C-FRPD. The first data set contained 1500 complete images, with only labels of pigs. The second data set contained 6000 images of individual pigs, which were cut from the original images. Each image had three types of labels, head, back, and tail. The third data set contained 1500 complete images having all types of labels. The data set was annotated in PASCAL VOC2007 format[28].

3 Cascade faster R-CNN pig detector

3.1 C-FRPD framework

The input to C-FPRD is an image of the pig pen. The output of C-FPRD is the individual data of all pigs in the image, including the location information of the whole pig, its head, back, and tail, as well as the classification confidence. The C-FRPD framework is shown in Figure 2. It is composed of four parts, feature extraction network, region proposal networks (RPN), individual pig detection network, and cascade structure.

The detection process of C-FRPD is divided into three stages. In the first stage, the RPN generates region proposals. In the second stage, the individual pig detection network makes classification and bounding box regression on region proposals to get the location of each individual pig. In the third stage, the cascade structure searches the location of the pig body parts on the feature map of the pig body area obtained in the second stage.



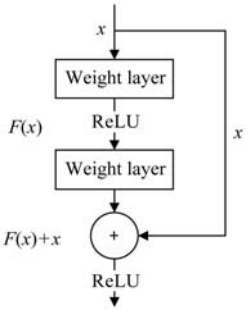
Note: RPN: Region proposal networks; ROI: Region of interest.

Figure 2 Algorithm framework of C-FRPD

3.2 C-FRPD feature extraction network

In C-FRPD, sharing computation was adopted. The feature map of the pen image was calculated by the same feature extraction network and input into three independent networks (RPN, individual pig detection network, and cascade structure). By sharing computation, three networks were cross-trained to get convergent results. Meanwhile, extracting the features of the entire image only once greatly saves the training and detection time.

In this study, ResNet-101 was used as the feature extraction network. ResNet-101 uses the residual structure[29], as shown in Figure 3. This structure can effectively gain accuracy from the considerably increased depth and accelerate the convergence of the network. Shortcut connections were established in ResNet-101 and a residual block was introduced to change the learning target into the residual value. When the network increases in depth, an identity mapping $H(x)=x$ will be learned for the redundant network layer. As shown in Figure 3, the input variable is x, and the residual function $F(x) = H(x) - x$ is learned. When the model accuracy reaches saturation, the training target of the redundant network layer will approach the residual result to 0. It makes $F(x)=0$ to realize identity mapping, so the training accuracy will not decrease when the network increases in depth.



Note: x is the input variable; $F(x)$ is the residual function; ReLU: Rectified linear unit; "+" is an addition operation between tensors.

Figure 3 Structure of the residual network used in this study

3.3 C-FRPD region proposal network

RPN implements the first stage of detection. It extracts object region proposals from the image. The input of RPN is the feature map extracted by ResNet101. Through the sliding window mechanism, the feature vector of each anchor was obtained. Then, a 1×1 convolution layer is used to perform classification and bounding box regression on the feature vector of each anchor. The classifier uses Softmax function to discriminate the object and background in the anchor and output the confidence value. The bounding box regression fine-tunes the anchor position and outputs the result including four parameters, offset of center coordinate, and scale of length and width. In the training process, the target value of the classifier is determined by the intersection over the union (IoU) of the anchors and the ground-truth boxes. When IoU>0.7, the anchor is treated as a positive sample. When IoU<0.3, the anchor is treated as a negative sample. The rest of the anchors were not used. The bounding box regression only uses positive samples for training. It needs to input the center coordinates and the length and width of the ground-truth box, which can be used to train the loss function to fine-tune the position of the bounding box. RPN outputs the region proposals to the next network.

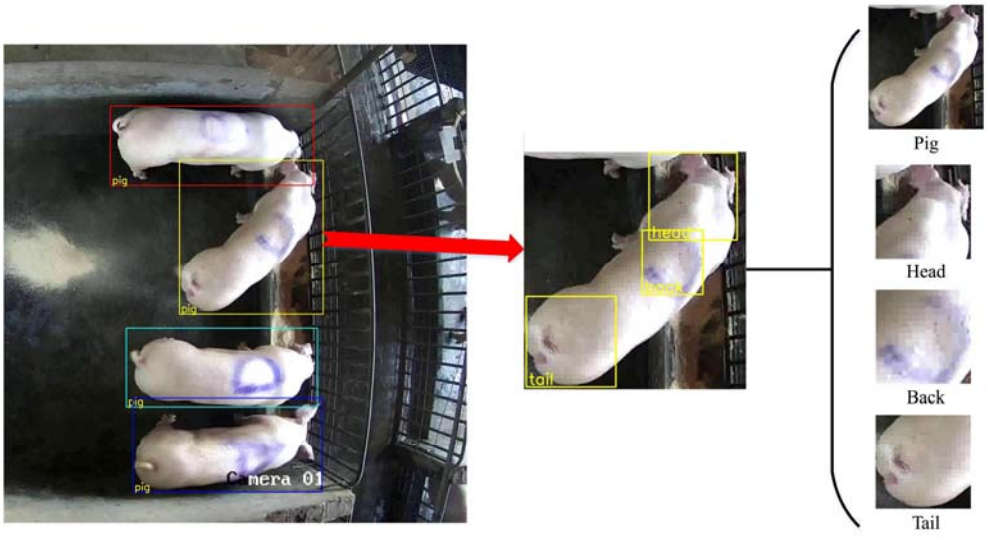
3.4 C-FRPD individual pig detection network

Individual pig detection network implements the second stage detection. It gets the pig proposals from region proposals. The feature map of region proposals generated by RPN is the input of the individual pig detection network. Since the scale of the region proposals is not the same, the scale of the map areas in the feature map is also different. Therefore, the Region of Interest (ROI) pooling layer is needed to pool the regional feature map into a uniform size feature vector and input it to the fully connected layer. Then classification and bounding box regression are performed again. Finally, the network outputs the five-dimensional information of all the individual pigs detected in the image. The five-dimensional information is

$$\mathrm{Pig}_i = [x_i,\ y_i,\ w_i,\ h_i,\ s_i] \quad (1)$$

where, x_i and y_i represent the upper-left coordinates of the detection box for Pig_i, w_i, and h_i represent the size of the detection box for Pig_i and s_i represent the classification confidence of the detection box for Pig_i.

The final result only retains the individual pig detection box with classification confidence $s_i>0.9$ as the input of the cascading structure. After the pig individual detection network, the results are shown in Figure 4.



a. Faster RCNN detection of individual pigs b. Input individual regions into cascade structure to detect the head, back, and tail of the pig c. Results are divided into four regions, the individual, head, back, and tail of each pig

Figure 4 Schematic of detection process of C-FRPD

3.5 C-FRPD cascade structure

Cascade structure implements the third stage detection. It locates the pig body parts according to pig proposals. Figure 5 shows the framework of the cascade structure. The pig proposal is a small region for searching. Cascading structure uses a sliding window to go through the feature map of pig proposal, and outputs a region proposal at each position. Considering that the area of the pigs' head, back, and tail in the image are similar in size, cascade structure selects a fixed-size anchor for detection. The fully connected layer gets the feature vector of region proposal, and then performs classification and bounding box regression. According to the experimental data, the label boxes of pig body parts are almost in a square shape in the image and the size is close to 128×128 pixels. Therefore, anchors with a size of 128 and a 1:1 length-width ratio were selected. According to the mapping relationship between the input image and the feature image, the size of the sliding window is calculated as 5×5.

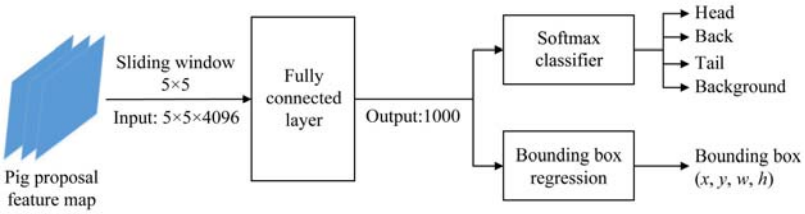


Figure 5 The cascade structure of C-FRPD used in this study

The input of the fully connected layer is 102 400 (5×5×4096) parameters and the output is 1000 parameters for the classifier and bounding box regression. The classifier uses Softmax function to classify the anchors into four parts: head, back, tail, and background. Bounding box regression uses the following equations to fine-tune the bounding box.

$$\begin{aligned}\hat{G}_x &= P_w d_x(P) + P_x \\ \hat{G}_y &= P_h d_y(P) + P_y \\ \hat{G}_w &= P_w e^{d_w(P)} \\ \hat{G}_h &= P_h e^{d_h(P)}\end{aligned} \tag{2}$$

where, $\hat{G}_x$ and $\hat{G}_y$ represent the center point coordinates of the prediction box; $\hat{G}_w$ and $\hat{G}_h$ represent the size of the prediction box; P represents the feature vector of the anchor; P_x and P_y represent the center point coordinates of the anchor; P_w and P_h represent the size of the anchor, and d_x, d_y, d_w, and d_h represent the function that needs training in the bounding box regression which calculate the offset of the center point coordinates and the scale of the size according to P.

After classification and bounding box regression, the cascade structure outputs the coordinates of multiple prediction boxes and their classification confidence. Only the detection box with the highest confidence is reserved for each category, because the head, back, and tail of the pig are unique. The final outputs of the cascade structure are the five-dimensional information of the pig head, back, and tail.

The pig pen image is input into C-FRPD, passing through feature extraction network, individual pig detection network, and cascade structure. Finally, C-FRPD outputs the location information of each pig and its head, back, and tail, as shown below.

$$X_i=[\text{Pig}_i, \text{Head}_i, \text{Back}_i, \text{Tail}_i] \tag{3}$$

Because the design of C-FRPD cascade structure and RPN are similar, the loss function of C-FRPD is consistent with Faster R-CNN. C-FRPD loss function is shown as follows:

$$L(p_i, t_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{4}$$

where, L_{cls} represents classification loss; L_{reg} represents regression loss; i represents the index of an anchor; p_i represents predicted category score; p_i^* represents truth category; t_i represents the coordinate and length and width of prediction box; t_i^* represents the coordinate and length and width of truth box; N_{cls} represents the number of anchor in a batch; N_{reg} represents the number of non-background prediction boxes.

4 Experiment and discussion

4.1 Experimental methods and environmental parameters

In order to evaluate the performance of C-FRPD, 1500 images were divided into the training set, verification set, and test set. 1000 images were randomly selected from the data set as the training set and 400 as the verification set. The remaining 100 images were used as the test set to evaluate the generalization of the model over unknown data. The test evaluated the accuracy of C-FRPD and analyzes the accuracy of the detection of each category. In addition, the experiment compared the effect of C-FRPD and Faster R-CNN algorithm on the detection of individual pigs and their parts, in order to verify the effectiveness of the cascade structure.

The experiment was done under a TensorFlow deep learning framework, with NVIDIA GeForce RTX2080ti for GPU acceleration. The parameter setting of C-FRPD is listed in Table 1.

Table 1 Experimental parameter setting of C-FRPD

Hyperparameters	Value
Initial learning rate	0.001
Decay of Initial Learning Rate	0.1
Momentum	0.9
Batch	128
RPN Scale of Anchor	[128,256,512]
RPN Rate of Anchor	[2:1,1:1,1:2]
RPN Batch	256
Cascade Structure Scale of Anchor	[128]
Cascade Structure Rate of Anchor	[1:1]
Cascade Structure Batch	256

The initial learning rate was 0.001 and its decay was 0.1. Momentum was 0.9. During the training process, Adam optimization algorithm was used to optimize and update the learning rate. For each iteration, 128 images were input as training samples, and 256 samples were taken for training RPN and cascading structure. RPN had 9 sizes of anchors, while the cascade structure had only 1 size.

4.2 Model training

The training process adopted the method of transfer learning. The feature extraction network was initialized with parameters pre-trained on PASCAL VOC 2012. Then the data set was used

to train the model and fine-tuned parameters until the ideal effect was obtained. The model training was divided into eight steps.

Step 1 Input 1000 images of pig pen from the training set into C-FRPD;

Step 2 Loaded the pre-training model parameters;

Step 3 Trained RPN separately;

Step 4 Input region proposal to train the individual pig detection network separately;

Step 5 Input 4000 images of individual pigs from the training set to train the cascade structure;

Step 6 Changed the training set to the images of pig pen, and train RPN again. At this time, fixed the parameters of ResNet-101 and only update the parameters of RPN;

Step 7 Used the region proposal to fine-tune the individual pig detection network. At this time, fix the parameters of ResNet-101 and only updated the parameters of the pig detection network;

Step 8 Used the pig proposal to fine-tune the individual pig detection network. At this time, fix the parameters of ResNet-101 and only updated the parameters of the cascade structure.

In Step 8, the training set needs to be adjusted in order to match the pig proposal with the labeled data. All pig proposals with ≥0.9 IoU overlap were treated with a ground-truth box as a positive sample. The label data of this pig and its parts were used as the ground truth during the training cascade structure.

4.3 C-FRPD algorithm performance analysis

4.3.1 C-FRPD detection accuracy analysis

The final detection effect of C-FRPD is shown in Figure 7a. The bounding box of the same color represents the same pig. The locating information of a pig generally consists of four parts, the whole pig, the head, the back, and the tail.

The trained C-FPRD was used to detect the accuracy of 100 images in the test set. Table 2 lists the test results.

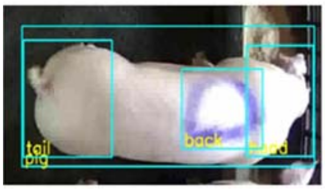
Table 2 Detection results of trained C-FPRD

Item	Pig	Head	Back	Tail
Correct detection	396	392	396	390
False detection	0	3	0	2
Missing detection	4	5	4	8

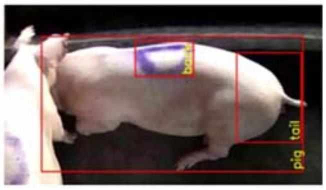
The results showed that the recognition accuracy of the individual pig was 99%, the head was 98%, the back was 99%, and the tail was 97.5%.

In the process of individual pig detection, several cases were missing. The reason is that when pig adhesion occurs, the two detection boxes are so close to each other that Non-Maximum Suppression (NMS)[30] only keeps one detection box. When the detection of the individual pig is missing, the detection of its body parts will also be missing.

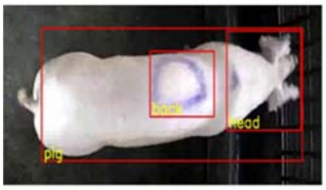
Figure 6 shows the detection results of C-FRPD. By analyzing the missed detection about the detection of the head and tail, two problems were found. First, the head and tail are easily occluded in the image, resulting in detection loss, as shown in Figure 6b. Second, some error is caused by phased detection. In the stage of individual pig detection, the detection box is slightly smaller than the ground-truth box of individual pigs due to the error of bounding box regression. Therefore, in the third stage, when the feature map is input to the cascade structure for pig body parts detection, the details of the head or tail are lost, resulting in detection missing, as shown in Figure 6c. By analyzing the false detection, it was found that it was caused by the adhesion of pigs. Multiple pig heads or tails appeared in the individual pig area, leading to false detection, as shown in Figure 6c. Although C-FRPD still has defects, the actual detection accuracy has reached the requirements of production and application.



a. Correct detection



b. Head missing due to occlusion



c. Tail missing due to border regression error

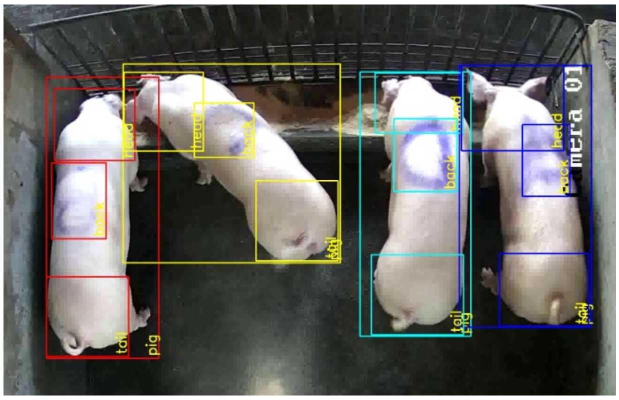
d. False detection due to multiple heads in the detection box

Figure 6 C-FRPD results of pigs and their body parts detection

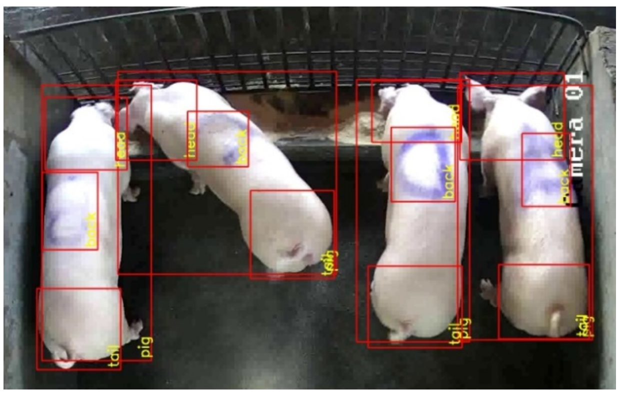
4.3.2 Comparison and analysis of detection results with Faster R-CNN

In order to show the improvement of the cascade structure on the Faster R-CNN, the same data set was used to train the Faster R-CNN algorithm and used the same test set to detect and compare it with the C-FRPD.

Figure 7b shows the detection result of the Faster R-CNN. The detection results of each category are independent, so it cannot use the same color to mark a pig and its body parts. C-FRPD conducts detection in stages, to ensure that the detection boxes of pig body parts are contained within the detection box of the whole pig. Therefore C-FRPD can directly express the relationship between the whole pig and its body parts. These four parts can be represented by the same color. C-FRPD describes an individual pig more accurately.



a. Detection result of C-FRPD



b. Detection result of Faster R-CNN

Figure 7 Result comparison of different pig detection methods

Table 3 lists the comparison results of various tests of Faster R-CNN and C-FRPD. The average detection accuracy of C-FRPD was 98.4%, and that of Faster R-CNN was 94.1%. The cascade structure increases the average detection accuracy by 4.3 percentage points. The detection accuracy of C-FRPD was higher than that of Faster R-CNN. It is thought that C-FRPD separates the whole pig and its parts in two detection stages so that the size of the bounding box is more uniform than in the same detection stage. It is beneficial to the training of bounding box regression, thus improving accuracy. By analyzing the wrong data, when two pigs get too close in the image, the detection of the same body part in different pigs will be lost due to the NMS keeps only one result without cascade structure. But with the cascade structure, an individual pig area retains one bounding box with maximum confidence for each part, thus avoiding missed detection.

Table 3 Result comparison of different pig detection methods

Item	C-FRPD	Faster R-CNN
Average detection accuracy/%	98.4	94.1
Detection accuracy of pig/%	99.0	97.0
Detection accuracy of head/ %	98.0	92.5
Detection accuracy of back/%	99.0	96.5
Detection accuracy of tail/%	97.5	90.5
Average detection time/ms	259	148

The average detection time of C-FRPD for one image was 259 ms, while the Faster R-CNN was 148 ms. Faster R-CNN is a two-stage detection algorithm. After adding on the cascade structure, C-FRPD becomes a three-stage detection algorithm, so the detection time is longer than the Faster R-CNN.

5 Conclusions

In this study, an algorithm names C-FRPD was proposed, which had a cascade structure based on the Faster R-CNN, for the detection of individual pigs and their body parts. C-FRPD could effectively establish an association between the whole pig and its body parts to get the location information of three body parts including its head, back, and tail.

The detection effects of C-FRPD and Faster R-CNN were compared. With the use of the cascade structure, the output of C-FRPD could accurately describe the inclusion relationship between the whole pig and its body parts with higher detection accuracy. Compared with Faster R-CNN, the average detection accuracy of C-FRPD reached 98.4%, with an increase of 4.3 percentage points.

C-FRPD can be effectively applied to the detection of individual pigs through video images and provide the location information of the whole pig and its body parts. This algorithm can provide technical reference for further research on individual behavior recognition of pigs.

Acknowledgements

This work was financially supported by the Key Research and Development Program of Guangdong Province (Grant No. 2019B020215004; 2019B090922002) and the Agricultural Research Project and Agricultural Technology Promotion Project of Guangdong (Grant No. 2021KJ383).

[References]

[1] Guo Y, Corke P, Poulton G, Wark T, Bishop-Hurley G, Swain D. Animal behaviour understanding using wireless sensor networks. In: Proceedings of 2006 31st IEEE Conference on Local Computer Networks, Tampa, FL, USA: IEEE, 2006; pp.607–614. doi: 10.1109/LCN.2006/322023.

[2] Shen M X, Liu L S, Yan L, Lu M Z, Yao W, Yang X J. Review of monitoring technology for animal individual in animal husbandry. Transactions of the CSAM, 2014; 45(10): 245–251. (in Chinese)

[3] Andrea S, Paolo B, Riccardo T, Torben G, Kristof M, Jarissa M, et al. The PigWise project: A novel approach in livestock farming through synergistic performances monitoring at individual level. In: Proceedings of the EFITA 2013 Conference, 2013; pp.1–8.

[4] Yu H A, Gao Y, Li X, Tong Y, Lei M G, Wang J J. Research review of animal behavior monitoring technologies: Commercial pigs as realistic example. Chinese Journal of Animal Science, 2015; 51(20): 66–70, 75. (in Chinese)

[5] Nasirahmadi A, Edwards S A, Sturm B. Implementation of machine vision for detecting behaviour of cattle and pigs. Livestock Science, 2017; 202: 25–38.

[6] Ott S, Moons C P H, Kashiha M A, Bahr C, Tuyttens F A M, Berckmans D, et al. Automated video analysis of pig activity at pen level highly correlates to human observations of behavioural activities. Livestock Science, 2014; 160: 132–137.

[7] Nasirahmadi A, Edwards S A, Matheson S M, Sturm B. Using automated image analysis in pig behavioural research: assessment of the influence of enrichment substrate provision on lying behaviour. Applied Animal Behaviour Science, 2017; 196: 30–35.

[8] Zhu W X, Guo Y Z, Jiao P P, Ma C H, Chen C. Recognition and drinking behaviour analysis of individual pigs based on machine vision. Livestock Science, 2017; 205: 129–136.

[9] Nasirahmadi A, Sturm B, Olsson A C, Jeppsson K H, Müller S, Edwards S, et al. Automatic scoring of lateral and sternal lying posture in grouped pigs using image processing and Support Vector Machine. Computers and Electronics in Agriculture, 2019; 156: 475–481.

[10] Kashiha M, Bahr C, Haredasht S A, Ott S, Moons C P H, Niewold T A, et al. The automatic monitoring of pigs water use by cameras. Computers and Electronics in Agriculture, 2013; 90:164–169.

[11] Nasirahmadi A, Hensel O, Edwards S A, Sturm B. Automatic detection of mounting behaviours among pigs using image analysis. Computers and Electronics in Agriculture, 2016; 124: 295–302.

[12] Yang X, Zhu W X. A recognition method for head and tail of adhesive pigs based on generalized Hough clustering. Jiangsu Agricultural Sciences, 2018; 46(9): 230–235.

[13] Guo Y Z. Study on individual recognition and drinking behaviour analysis of topview group-housed pigs based on machine vision. Doctoral dissertation. Zhenjiang, China: Jiangsu University, 2018; 124p. (in Chinese)

[14] Gao Y, Yu H A, Lei M G, Li X, Guo X, Diao Y P. Trajectory tracking for group housed pigs based on locations of head/tail. Transactions of the CSAE, 2017; 33(2): 220–226.

[15] Chen C, Zhu W X, Norton T. Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning. Computers and Electronics in Agriculture, 2021; 187: 106255. doi: 10.1016/j.compag. 2021.106255.

[16] Chen C, Zhu W X, Steibel J, Siegford J, Han J J. Recognition of feeding behaviour of pigs and determination of feeding time of each pig by a video-based deep learning method. Computers and Electronics in Agriculture, 2020; 176:105642. doi: 10.1016/j.compag.2020.105642.

[17] Yang Q M, Xiao D Q, Cai J H. Pig mounting behaviour recognition based on video spatial-temporal features. Biosystems Engineering, 2021; 206: 55–66.

[18] Alameer A, Kyriazakis I, Dalton H A, Miller A L, Bacardit J. Automatic recognition of feeding and foraging behaviour in pigs using deep learning. Biosystems Engineering, 2020; 197: 91–104.

[19] Yang Q M, Xiao D Q, Lin S C. Feeding behavior recognition for group-housed pigs with the Faster R-CNN. Computers and Electronics in Agriculture, 2018; 155: 453–460.

[20] Yang Q M, Xiao D Q, Zhang G X. Automatic pig drinking behavior recognition with machine vision. Transactions of the CSAM, 2018; 49(6): 232–238. (in Chinese)

[21] Lu H T, Zhang Q C. Applications of deep convolutional neural networks in computer vision. Journal of Data Acquisition and Processing, 2016; 31(1): 1–17. (in Chinese)

[22] Liu W B, Wang Z D, Liu X H, Zeng N Y, Liu Y R, Alsaadi F E. A survey of deep neural network architectures and their applications. Neurocomputing, 2016; 234: 11–26.

[23] Zhang S, Gong Y H, Wang J J. The development of deep convolution

neural network and its applications on computer vision. Chinese Journal of Computers, 2019; 42(3): 453–482. (in Chinese)

[24] Yao Z Y. Research on the application of object detection technology based on deep learning algorithm. Master dissertation. Beijing: Beijing University of Posts and Telecommunications, 2019; 58p. (in Chinese)

[25] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA: IEEE, 2014; pp.580–587. doi: 10.1109/CVPR.2014.81.

[26] Girshick R. Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile: IEEE, 2015; pp.1440–1448. doi: 10.1109/ICCV.2015.169.

[27] Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015; 39(6): 1137–1149.

[28] Everingham M, Eslami S M, Gol L V, Williams C K I, Winn J, Zisserman A. The Pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, 2015; 111(1): 98–136.

[29] He K M, Zhang X Y, Ren S Q, Sun J. Identity mappings in deep residual networks. In: Computer Vision - ECCV 2016, Springer, 2016; pp.630–645. doi: 10.1007/978-3-319-46493-0_38.

[30] Neubeck A, Van Gool L. Efficient non-maximum suppression. In: 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China: IEEE, 2006; pp.850–855. doi: 10.1109/ICPR.2006.479.