Pig target detection method based on SSD convolution network

Minghui Deng¹, Ran Yu²*

¹College of Electrical and Information, Northeast Agricultural University, Harbin, Heilongjiang, 150000, China
²College of Electrical and Information, Northeast Agricultural University, Harbin, Heilongjiang, 150000, China

*Corresponding author’s e-mail:yuran1010@outlook.com

Abstract. The behavioral changes of pigs can not only reflect their own health and welfare status, but also have important significance in improving the breeding environment. Rapid and accurate detection of pigs' targets is the basis of behavioral change analysis. In view of this situation, a target detection algorithm based on deep learning is proposed, and its model structure and working principle are expounded. Then, through the test of pig image data set, the results show that this method can accurately detect pig targets under complex background.

1. Introduction

China’s pig breeding quantity and pork consumption rank the first in the world, and pig breeding industry is the pillar industry of China’s agriculture [1-2]. The breeding environment is very important for the growth and development of pigs. Accurate and rapid detection of the target pigs in the video image can help to find abnormal behaviors of pigs and take timely measures to reduce the incidence of disease [3-4]. Accurate and effective pig detection algorithm is the basis of pig behavior analysis and breeding decision [5-7].

Traditional target detection algorithms including two consecutive frames subtraction method and background subtraction for piglet target identification are easy to operate [8-9], but have low accuracy and are sensitive to noise interference. In recent years, deep learn-based convolutional neural network (CNN) has become a research hotspot in the field of computer vision due to its powerful feature expression ability for images [10-11], and achieved fruitful results in image classification and recognition, target detection and other fields.

In order to improve the speed of the pig target detection, considering both the pig target under multiple jamming scenario shows good generalization ability, and can also simplify the complex image denoising feature extraction process, this paper proposes a pig target detection method based on the SSD convolution network, this method supports batch processing image data, and video and video monitoring real-time detection, and test results are stored by diversification.

2. Experimental data set construction

The experimental data were collected from the pig farm in acheng, Harbin city, heilongjiang province, which took real-time photos from multiple angles with the camera, as shown in figure 1 (a) (b). The production steps of the experimental data set are as follows: (1) select 2000 real video images of pig farms. (2) manually annotate the selected images with LabelImg software, and make an XML file to
determine the position of the target pig in the image. (3) Construct the experimental training data set according to VOC2007 data set format. The annotation format is PASCAL VOC data set standard format.

The multi-interference scene labeling scheme for pig target identification is as follows: (1) If both sides of the railing have obvious same pig body parts, all areas within the range can be retained. If the head or tail of the pig is completely covered, only the viewing area is left. Pig out, keep in the video range of the part. (3) Pig adhesion problem marked human eyes can clearly identify the range of the pig part. Figure 1 (c).

3. Target detection algorithm

To integrate the accuracy and speed of the target detection algorithm, the SSD method proposed by Liu [12] et al. was adopted as the framework of target detection in this paper. It combines the regression thought of YOLO [13] and the anchor thought of faster-rcnn [14]. Compared with the Faster-RCNN target detection model, this algorithm has no process of generating region Suggestions and greatly improves the detection speed. Compared with YOLO, it achieved higher accuracy.

3.1 SSD algorithm structure

Figure 2 is SSD300, the input image size is 300×300, the feature extraction part uses the convolutional layer of VGG16, then the two full connection layers of VGG16 are replaced by the convolutional layer, conv6 and conv7 in the figure, then several convolutional layers are added (conv8_2, conv9_2,
conv10_2), and finally conv11_2 is used to become the output of 1×1.SSD conv4_3, conv7, conv8_2, conv9_2, conv10_2, conv11_2 convoluted feature maps obtained from conv4_3, conv7, conv8_2, conv10_2 and conv11_2 were simultaneously classified by softmax and position regression.

3.2 Feature drawing default border size
SSD can obtain multiple feature maps of different sizes by using multi-scale method. If the m-layer feature graph is used in the detection of the model, the default box ratio $s_k$ of the k feature map is represented as follows:

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1}(k-1), k \in [1,m]$$

The aspect ratio of the prior frame is $\{3/1, 2/1, 3, 2, 1\}$. Also set a prior box of $r_a = 1$, which $s_k = \sqrt{s_k s_{k+1}}$. For conv4_3, conv10_2, conv11_2 uses only four prior boxes, not a prior box with a ratio of length to width of 3,1/3, and the featuremap of other layers uses six prior boxes. The width and height of each prior box are calculated as follows:

$$w = s_k \sqrt{r_a}, h = \frac{s_k}{\sqrt{r_a}}$$

A total of 8732 priori boxes are generated for each input picture, which are used to train to get the predicted position and classification.

3.3 Loss function
The training of SSD regresses the target type and the target position at the same time. The target loss function is the sum of confidence loss $L_{conf}(x,c)$ and position loss $L_{loc}(x,l,g)$ . The total loss function $L(x,c,l,g)$ is expressed as follows:

$$L(x,c,l,g) = \frac{1}{N} \left( L_{conf}(x,c) + \alpha L_{loc}(x,l,g) \right)$$

In the formula, $N$ is the number of default boxes matching the object box; $x$ is the result of matching the default box with different categories of object boxes; $c$ is the confidence degree of the predicted object box; $l$ is the position information of the predicted object box; $g$ is the position information of the object box; $\alpha$ is the parameter of weighing confidence loss and position loss, which is generally set to 1. In the course of training, by reducing the value of the loss function, we can ensure that the reliability of the prediction box category can be improved, and at the same time, the position credibility of the prediction box can be improved.

4. Experiment and analysis
Adopting Ubuntu16.04 operating system, based on Python, OpenCV3, Caffe to build convolution neural network model.
4.1 Training efficiency analysis

The loss value (loss) change of the pig identification training process is shown in Figure 3. An analysis of FIG.3 shows that the model loss value drops sharply in the 0-1000 iterations, and the entire process converges to near 2 in 1000 iterations. The model is close to the local optimal value at the time of 30000 iterations, and the training effect is ideal.

Fig. 3 Training error curve of pig detection.

4.2 Analysis of detection effect

In order to test the effectiveness of the proposed method, the video image (which does not overlap with the training sample) is input into the trained CNN model for classification. After non-maximum suppression, the verification effect is shown in figure 4. It can be seen from figure 4 that the proposed algorithm can detect the target pig accurately and distinguish the background from the target accurately, which shows that the proposed method can eliminate the interference of the complex background and has strong robustness to the complex background.

Fig. 4 Effect of the method in this paper.

5. Conclusion

In this paper, the SSD target detection network in deep learning is used to construct the pig image data set for network training, and the effectiveness of the constructed data set is verified by experiments. For different sizes of targets, SSD algorithm uses the feature maps of different convolution layers to achieve accurate and target measurement. This method supports image batch processing, and can also detect video and surveillance video in real time, and the recognition results can be saved in the form of image and video. The state and behavior of live pigs can be accurately determined by the combination of diverse recognition results.
Acknowledgments
This paper was supported by Harbin Science and Technology Innovation Talent Project (NO.2017RQXJ079).

References
[1] Peng, Y.K, Yang, Q.H, Wang, W.X. (2018) On-line detection and classification of pork moisture based on near-infrared spectra [J]. Transactions of the Chinese Society for Agricultural Machinery, 49 (3): 347 - 353. (in Chinese)
[2] Guo, H.W. (2018) Analysis of current situation and trend of pork consumption in China [J]. Today’s Pig Industry, (4): 60 -67. (in Chinese)
[3] Duan, Y.Y., Ma, L., Liu, G. (2015) Remote monitoring system of pig motion behavior and piggery environment based on Internet of Things[J]. Transactions of the CSAE, 31( Supp. 2) : 216 - 221. ( in Chinese)
[4] Li, Y.Y, Sun, L.Q., Sun, X.X. (2017) Automatic tracking of pig feeding behavior based on particle filter with multi-feature fusion[J]. Transactions of the CSAE, 33(Supp.1): 246 - 252. (in Chinese)
[5] Sun, L.Q., Li, Y., Zou, Y.b., et al. (2017) Pig image segmentation method based on improved Graph Cut algorithm[J]. Transactions of the CSAE, 33(16): 196 - 202. (in Chinese)
[6] Zou, Y.B, Sun, L.Q., Li, Y.et al.(2017) Video monitoring and analysis system for pig breeding based on distributed flow computing[J].Transactions of the Chinese Society for Agricultural Machinery, 48(Supp.): 365 - 373.
[7] LI, Y.Y, Sun, L.Q., Zou, Y.B. et al. (2017) Individual pig object detection algorithm based on Gaussian mixture model[J]. International Journal of Agricultural & Biological Engineering, 2017, 10(5): 186 - 193.
[8] Zhu, M.H, Luo, D.Y. (2006) Moving objects detection and tracking based on two consecutive frames subtraction background model[J]. Computer Measurement & Control,14(8):1004 - 1006. (in Chinese)
[9] Wang,G.Q., Gai,Q.L., Yu,H.Y., et al. (2014) Video target detection algorithms based on background subtraction[J].Journal of Engineering of Heilongjiang University,5(4):64 - 68. (in Chinese)
[10] KRIZHEVSKY, A., SUTSKEVE, R.I., HINTON, G. E. (2017) Image Net classification with deep convolutional neural networks[J]. Communications of the ACM, 60(6): 84 - 90.
[11] He,K.M, Zhang, X.Y., Ren, S.Q., et al. (2016) Deep residual learning for image recognition[C] Computer Vision and Pattern Recognition. IEEE, 770 – 778.
[12] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]// European conference on computer vision. Springer, Cham, 21-37
[13] Redmon J, Divvala S, Girshick R, et al. (2016) You only look once: Unified, real -time object detection [C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 779-788
[14] Ren S, He K,Girshick R, et al. (2017) Faster R-CNN:Towards real-timeobject detection with region proposal networks[J].IEEE Trans-actions on Ratern Analysis & Machine Intelligence,39(6):1137-1149