Automatic detection of follicle ultrasound images based on improved Faster R-CNN

Tianlong Zeng¹²³, Jun Liu ¹²

¹ School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, P. R. China
² Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan 430081, P. R. China
³ Correspondence: 942636327@qq.com

Abstract. Follicle Ultrasonic images detection technology plays an important role in the monitoring of bull-follicle. Because of follicle ultrasound image containing lots of speckle noise and fuzzy edges, the traditional image detection algorithm is difficult to get better detection results on the ultrasonic image, and the traditional image detection algorithm needs to carry out sample feature extraction for each image, which is time-consuming and labor-intensive. According to the characteristics of the cattle follicle ultrasound image sets, this paper proposes a model of image detection based on improved deep learning Faster R-CNN to automatically detect cattle ovarian follicles, through joint VGG-16 different network layer characteristic figure to replace single deepest characteristic figure, retain the deep semantic characteristics at the same time, also keep the shallow characterization information. The experimental results show that this method has a better effect on the ultrasonic image detection of bovine follicle.

1. Introduction

Nanyang Yellow Cattle is a famous yellow cattle breed in China with good working capacity, meat performance and adaptability. In the field of monitoring follicles of yellow cattle, ultrasound technology has been widely used[1][2]. The traditional yellow cattle breeding industry relies mainly on artificial farming experience to arrange breeding and conception, which leads to high breeding costs and low reproductive rates. Ultrasound imaging technology, as a new technology, is widely used in cattle breeding and embryo production because of its non-invasive, non-radioactive, real-time dynamic display and low price. It is an effective means to improve the breeding of yellow cattle. Ultrasound image detection technology plays an important role in the monitoring of yellow cattle follicles. By monitoring the follicles of yellow cattle, finding the best time and conception for breeding can effectively improve the reproductive capacity of yellow cattle. Follicular detection is one of the important steps in follicular ultrasound imaging. Accurate positioning of the target area using image object detection technology has a decisive influence on the accuracy of subsequent follicle segmentation, quantitative calculation and measurement. Foreign scholars Pierson and Ginthers used ultrasound imaging to test and verify the hypothesis proposed by Rajakoski that the cattle would have two follicular developmental waves in one estrous cycle[3]. In China, Nong et al. used ultrasound to observe the follicular development and ovulation of Leiqiong yellow cattle under different estrus conditions[4]. In the field of ultrasonic image detection, Potocnik summarized three methods for automatic detection of ultrasound images of follicles:
cellular neural network algorithm, region growing algorithm and predictive correction algorithm[5]. In 2010, PS Hirenmath proposed three kinds of follicle detection algorithms: edge detection[6], geometric features[7], and active contourlet waves[8]. Since 2012, the University of Toronto Alex proposed AlexNet[9] and achieved the best results of ImageNet in 2012. After that, various better and deeper networks have been proposed. As a branch of machine learning, deep learning has developed rapidly in recent years and has won all the champions of the years. In 2014, Ross B. Girshick used the region proposal combined with CNN (Convolutional Neural Network) to replace the sliding window and manual design features used in traditional object detection, and designed the R-CNN framework[10], which made a huge breakthrough in object detection and opened a craze of object detection based on deep learning. In the same year, Kaiming He et al. proposed the SPP (Spatial Pyramid Pooling) net[11], which greatly improved the speed of R-CNN. In 2015, Ross B. Girshick improved R-CNN to Fast R-CNN[12] and the final Faster R-CNN[13].

2. Method
Faster R-CNN is Ross Girshick's end-to-end neural network that combines object detection and classification. It has two earlier versions, R-CNN and Fast R-CNN, as well as Ross's early research result.

2.1. R-CNN
R-CNN, the Region with CNN, is a milestone in the application of convolutional neural networks to object detection. CNN has good performance for feature extraction and classification, and the Region Proposal method is used to achieve object detection. At that time, Ross used CNN for feature extraction. From the experience-driven artificial feature paradigm HOG and SIFT to the data-driven representation learning paradigm, the feature representation ability of the sample was improved. The extraction of Region Proposal was based on the traditional Selective Search method. The extracted feature is to train and classify with SVM. In the training process, there are supervised pre-training under large samples and fine-tuning under small samples to solve the problem that small samples are difficult to train or even over-fitting.

2.2. Faster R-CNN
Fast R-CNN has improved R-CNN by first replacing the SVM classifier with Softmax, and replacing the last max pooling layer with the ROI pooling layer to map region proposals of different sizes to the same dimension, and the ROI pooling layer can back propagate, thus there are only once feature extraction of the entire image area avoiding redundant feature extraction operations in R-CNN; the Fast R-CNN network uses parallel different fully connected layers at the end, which can simultaneously output classification results and the box regression results ,and uses SVD to decompose it, which reduces the computational complexity, speeds up the detection, implements end-to-end multitasking training, and does not require additional feature storage space.

The Figure 2 shows the schematic diagram of Fast R-CNN.
The RoI pooling layer is the main reason why the Fast R-CNN is significantly faster than the R-CNN. It is simply the special-case of the spatial pyramid pooling layer used in SPP nets in which there is only one pyramid level. Each RoI is defined by a four-tuple \((r, c, h, w)\) that specifies its top-left corner \((r, c)\) and its height and width \((h, w)\). There are two main functions, one is to locate the RoI in the image to the corresponding patch in the feature map, and the other is to down sample (max pooling) the feature map patch into a fixed-size feature with a single-layer SPP layer and pass it to the fully-connected layer FC.

![Figure 2 Schematic diagram of Fast R-CNN](image)

2.3. Faster R-CNN

Following Fast R-CNN, the region proposal algorithm Selective Search (2s/image) and EdgeBoxes (0.2s/image) implemented on the CPU become the biggest bottleneck in object detection speed improvement. The Faster R-CNN replaces the Selective Search with the RPN layer, thus achieving a complete end-to-end of the network.

RPN, Region Proposal Networks, draws on the ideas of SPPnet and RoI, first maps the feature map obtained by CNN to the original image, and designs boxes of different sizes on the original image. The box is called an anchor. Then calculate the IOU between these anchors and ground truth, and divide the anchor into positive and negative according to the calculated IOU value. The positive class indicates that the value of IOU is greater than 0.7, and the negative class indicates that the value of IOU is less than 0.3. Finally, a certain amount of positive and negative samples are selected according to the numerical value for training, so that the boundary of the boxes is regressed and corrected.

The RPN structure is shown as Figure 3:

![Figure 3 Schematic diagram of RPN](image)

In a certain amount of anchors as the positive and negative samples obtained by the RPN, each anchor will output two scores first, one indicating the probability of the object, another indicating the probability of not being the object, and uniformly represented by \(p_i\); then, outputting 4 coordinate values are used to represent the position coordinates of the anchor. Therefore, the loss function of the RPN layer is a multi-task loss function, which consists of two parts: one is the SoftmaxLoss of the object probability, and the other is the smooth L1 Loss of the coordinate position between the anchor and the ground truth. The specific definition of the Loss function is as follows:
Here, $i$ represents the $i$-th anchor in a mini-batch and $p_i$ is the predicted probability of anchor $i$ being an object. The ground-truth label $p_i^\star$ is 1 if the anchor is positive, and is 0 if the anchor is negative. When the label of the anchor is positive, the label is negative when it is negative. $t_i$ is a vector representing the 4 parameterized coordinates of the predicted bounding box, and $t_i^\star$ is that of the ground-truth box associated with a positive anchor. Assume $(x, y, w, h)$ denote the box’s center coordinates and its width and height, then variables $(x_a, y_a, w_a, h_a)$ are for the predicted box, anchor box, and ground-truth box respectively (likewise for $y, w, h$). So the task of learning is to make the values of the two similar. as the figure 4 shows:

![Figure 4](image)

The predicted box, anchor box, and ground-truth box

Including:

$$t_x = (x - x_a) / w_a, t_y = (y - y_a) / h_a,$$

$$t_x = \log(w / w_a), t_y = \log(h / h_a),$$

$$t_x^\star = (x' - x_a) / w_a, t_y^\star = (y' - y_a) / h_a,$$

$$t_x^\star = \log(w' / w_a), t_y^\star = \log(h' / h_a)$$

In formula (2), $L_{cls}$ is the SoftmaxLoss for two categories, and the formula is:

$$L_{cls}(p_i, p_i^\star) = -\log(p_i^\star p_i + (1 - p_i^\star)(1 - p_i))$$

$L_{reg}$ is the smooth L1 Loss for two offsets, the formula is:

$$L_{reg}(t_i, t_i^\star) = R(t_i - t_i^\star)$$

in which the function $R$ is defined as follows:

$$R(x) = \text{smooth}_{\lambda}(x) = \begin{cases} 0.5x^2 & |x| \leq 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

The hyper-parameter $\lambda$ in formula (2) controls the balance between the classification loss and the regression loss. $N_{cls}$ and $N_{reg}$ are used to normalize the classification loss $L_{cls}$ and the regression loss $L_{reg}$, respectively.

For the entire network, nearly half of the time is in the full connection layer operation. Therefore, the truncated SVD is used to accelerate the FC calculation.

The Figure 5 shows the schematic diagram of Faster R-CNN.
2.4. Improved Faster R-CNN

The improved faster R-CNN borrowed the thought of FCN which connected the conv3, conv4 and conv5 layers of vgg_16 network together to form a Hyper Feature map instead of the original conv5_3 for RPN and Fast R-CNN. Because of the small target of follicular ultrasound image, in order to maintain high accuracy, the deep semantic features and the shallow characterization information are retained at the same time.

The Figure 6 shows the schematic diagram of Hyper Feature map.

3. Implementation Detail

The ultrasound yellow cattle follicle image set of this experiment is derived from the National Natural Science Foundation of China, “Key Techniques for Monitoring Dynamic Changes of Follicles in Nanyang Yellow Cattle Based on 3D Ultrasound Imaging”. The data set contained 6000 section images from 115 yellow cattle follicles three-dimensional ultrasound images.

3.1. Data process

The raw data of this experiment is a three-dimensional ultrasound image of the mvl file. Before training, the 3D ultrasound images in the file are saved as bmp files in frames by View3DXI software. Then, through the MATLAB software, the obtained bmp file is image data enhanced by using image flipping, rotation, random cutting, etc., and is saved as a jpg file and numbered. Finally, the image is labeled by ImageLabel in github, and finally the data set is saved in the form of a VOC 2007 data set.

3.2. Training

4-Step Alternating Training [13] is used during training. During training, the data set is randomly divided into two sets of Trainval and test, each accounting for 50%, 3000 images, and Train_val is divided into two parts of train and val, which still account for 50%, that is, 1500 image. Pre-training was performed using VGG-16 network during training.
Table 1 The results of the experiment

| Model                  | Proposals | mAP(%) |
|------------------------|-----------|--------|
| Faster R_CNN 300       | 300       | 70.1   |
| Faster R_CNN 2000      | 2000      | 75.4   |
| Improved Faster R_CNN 300 | 300       | 73.9   |
| Improved Faster R_CNN 2000 | 2000      | 78.3   |

Figure 7 The results of the experiment

3.3. Main result

The improved Faster R_CNN mAP experimental results are shown in the table 1.

The results of the experiment on the ultrasound images of various types of follicles are shown in the figure 7.

Experiments show that the deep learning method Faster R-CNN has higher recognition accuracy and recognition precision and faster recognition speed for follicular ultrasound images, compared with the traditional image processing method. Moreover, the multi-follicle ultrasound image and the multi-image combined follicle ultrasound image have good recognition ability, which is not possible by the traditional image processing method.

4. Conclusion

This paper introduces the research background, significance and research status of follicle ultrasound images and neural networks. Then, using the improved Faster R-CNN model to detect and locate the ultrasound image of follicle, the speed and precision are greatly improved, especially for multi-follicle images. However, there are still many shortcomings in this experiment, and there is a lot of room for improvement in the recognition accuracy and recognition speed of the follicle super-god image. And generally a follicle will have 30-60 slices. Therefore, it is possible to consider the continuous relationship between the upper and lower layers when detecting, which greatly reduces the calculation amount of the network and greatly improves the recognition speed of the ultrasound image of the follicle.

References
[1] G. Cortela, L. Leija. Elastograms of the diabetic foot by ultrasonic impulse elastography [J]. IEEE Transactions on Image Processing, 2016.
[2] Filteau V, DesCôteaux L. Predictive values of early pregnancy diagnosis by ultrasonography in dairy cattle[C]//American Association of Bovine Practitioners. Conference (USA). 1998.
[3] Fricke P M. Scanning the future—Ultrasonography as a reproductive management tool for dairy cattle[J]. Journal of Dairy Science, 2002, 85(8): 1918-1926.
[4] Nong H J. Study on Dynamic Changes of Follicle Development in Leiqiong Yellow Cattle based on B-ultrasound [D]. Guangxi University, 2007.
[5] Potocnik B, Cigale B, Zazula D. The XUltra project - Automated Analysis of Ovarian Ultrasound Images[C]// Computer-Based Medical Systems. IEEE, 2002:262-267.
[6] Hiremath P S, Tegnoor J R. 2010 “Automatic Detection of Follicles in Ultrasound Images of
Ovaries using Edge Based Method[J]. International Journal of Computer Applications Special Issue on Recent Trends in Image Processing & Pattern Recognition, 2011, RTIPPR(3):120-125.

[7] Hiremath P S, Tegnoor J R. Recognition of follicles in ultrasound images of ovaries using geometric features[C]// International Conference on Biomedical and Pharmaceutical Engineering. IEEE, 2010:1-8.

[8] Hiremath P S, Tegnoor J R. Follicle detection in ultrasound images of ovaries using active contours method[C]// International Conference on Signal and Image Processing. IEEE, 2010:286-291.

[9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc, 2012:1097-1105.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.

[11] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Trans Pattern Anal Mach Intell, 2015, 37(9):1904-1916.

[12] Girshick R. Fast R-CNN[J]. Computer Science, 2015.

[13] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(6):1137-114