ANOMALY DETECTION FOR HERD PIGS BASED ON YOLOX
/
基于 YOLOX 的群养猪只异常检测

Yanwen LI [1], Juxia LI [*1], Zhenyu LIU [2], Zhifang BI [3], Hui ZHANG [1], Lei DUAN [1]
[1] College of Information Science and Engineering, Shanxi Agricultural University, Taigu/China
[2] Graduate School of Shanxi Agricultural University, Shanxi / China
[3] Basic Department, Shanxi Agricultural University, Shanxi / China
Tel: +86-15803446486; E-mail: lijxsn@126.com
DOI: https://doi.org/10.35633/inmateh-69-08

Keywords: *Abnormal behavior, Target detection, Deep Learning, YOLOX*

ABSTRACT
In order to solve the problem that the complex pig house environment leads to the difficulty and low accuracy of abnormal detection of group pigs, the video of 9 adult fattening pigs were collected, and the video key frames were obtained by the frame differential method as the training set, and the YOLOX model for abnormal detection of group pigs was constructed. The results show that the average accuracy of YOLOX model on the test set is 98.0%. The research results can provide a reference for the detection of pig anomalies in the breeding environment of pig farms.

摘要
为了解决复杂的猪舍环境导致对群养猪只异常检测困难和准确率较低的问题，本研究采集 9 头成年育肥猪视频图像，采用帧间差分法获取视频关键帧作为训练集，构建群养猪只异常检测 YOLOX 模型。结果表明，YOLOX 网络模型在测试集上平均准确率达 98.0%。研究结果可为猪场养殖环境中针对猪只异常检测提供参考。

INTRODUCTION
The monitoring of abnormal behaviors such as attack, ear-biting and other conditions of pigs is mainly conducted in a manual way, which not only increases the cost of pig breeding management, but also leads to stress response of pigs or even results in cross-infection between human and animals *(He et al., 2019)*. With the development of computer vision technology, it becomes possible that this technology can be used for the health monitoring of pigs in a non-contact and low-stress way. Target detection and behavior recognition of pigs are the premises for achieving automatic monitoring and intelligent analysis of pigs *(Hua et al., 2019)*. Rapid and accurate detection of abnormal behavior of pigs is of great significance for achieving accurate, personalized and intelligent health monitoring of pigs.

At present, some progress has been made in the research of abnormal behavior of pigs. In 2014, Viazzi extracted motion features from the historical data of motion images of pigs to classify their abnormal behavior by Linear discriminant analysis (LDA) *(Viazzi et al., 2014)*. Ocazk used BP neural network to classify the degree of aggressive behavior of pigs via having a statistical calculation of their activity indexes *(Oczak et al., 2014)*. In 2016, Lee classified the two abnormal behaviors of chase and attack of pigs by support Vector Machine (SVM) after having the activity features of pigs extracted by Kinect depth sensor *(Lee et al., 2016)*. In 2017, Chen extracted the acceleration features of pigs by analyzing the displacement changes of pig targets between adjacent key frame images, and had the degree of aggressive behavior of pigs classified with the hierarchical clustering method adopted *(Chen et al., 2017)*. All the studies about detection of abnormal behavior of pigs as listed above are based on extracting a feature in pig images Based on image processing technology before machine learning and other means are combined to process the feature. However, in practical application, extra feature extraction is needed in traditional image processing technology, which leads to the problems of low efficiency and large workload. What's more, due to adhesion, occlusion, poor lighting conditions as well as complex abnormal behaviors of pigs in the piggery environment, it is difficult to detect abnormal behavior of pigs in intensive breeding piggeries efficiently and in real time by traditional methods.

In recent years, deep learning has shown its superiority over traditional methods in the fields of image and vision. From the extraction and learning of low-dimensional features to high-dimensional features, deep learning can detect and recognize various tasks in most cases *(Ren et al., 2017, Zeiler et al., 2014, Zhang et al., 2015)*.

Demonstrating strong learning and generalization ability in various fields, deep learning has also been widely used in behavior detection of pigs. In 2018, Yang proposed to use Faster R-CNN to locate and identify individual pigs from a group-housed pen *(Yang et al., 2018)*. Zheng also used Faster R-CNN to identify five postures (standing, sitting, sternal recumbency, ventral recumbency and lateral recumbency) and obtain sows accurate location in loose pens *(Zheng et al., 2018)*.

Xue introduced the Center Loss monitoring signal into the Faster R-CNN training to enhance the cohesion of intra-class features, so as to improve the accuracy of identification of postures of lactating sows *(Xue et al., 2018)*. In 2019, Li used Resnet-FPN network to improve the Mask R-CNN deep learning model to identify the mounting behavior of pigs *(Li, et al., 2019)*. Yan proposed a detection model of DAT-YOLO, which is a combination of both attention mechanism and Tiny-YOLO, to detect the facial postures of herd pigs *(Yan et al., 2019)*.

Gao constructed a 3D convolutional neural network model that is used to identify the aggressive behavior of herd pigs *(Gao et al., 2019)*. In 2020, Yan proposed a model of FPA-Tiny-YOLO, which is a combination of feature pyramid attention FPA and Tiny-YOLO, for the shape detection of pigs, so as to achieve multi-target detection of herd pigs *(Yan et al., 2020)*. In 2021, Li used YOLOv4 model to detect the dietary behavior of pigs *(Li et al., 2021)*. Xue proposed a convolutional network model that integrates 2D-3D convolution features together to identify postures of sows *(Xue et al., 2021)*. By introducing lightweight MobileNet and feature pyramid structure FPN in, Fang improved the CenterNet model and conducted a target detection of herd pigs *(Fang et al., 2021)*. Deep learning has shown excellent performance in both target and simple behavior detection of pigs, but few studies on advanced abnormal behavior of herd pigs in interactive states are available at the moment. In this study, inter-frame difference method was adopted to obtain key frames to form a data set for the detection of abnormal behavior of herd pigs, so as to construct a deep learning target detection model YOLOX, which shall be used for the detection of abnormal behavior of herd pigs, so that the problems of both difficulty and low accuracy in detection of abnormal behaviors of pigs that are there because of the complicated postures of pigs and the complex piggery environment may get solved. By training the model, an effective model for detection of abnormal behavior of pigs was constructed, and the generalization and feasibility of the algorithm were verified by having model prediction tests carried out on the test set in different time buckets.

MATERIALS AND METHODS

Data acquisition and pre-processing

The experimental data in this study were collected in July 2020 at Fenxi Pig Breeding Base in Linfen City, Shanxi Province within 24 hours without interruption. The piggery was of closed-type brick-concrete structure. A total of 9 five-month-old finishing pigs were selected and placed in an enclosed piggery of size of 4.5m×4m×2.8m. As for the selection of camera position, the conventional aerial view and flat view acquisition method have the following disadvantages: the images collected from aerial view Angle are mainly the back area of pigs, which cannot be obtained for the face or other key parts, and the specific behavioral characteristics of pigs cannot be obtained. The horizontal Angle of view will show a lot of occlusion in the barn environment. Therefore, this paper innovatively adopted a tilt down 60 degrees Angle of view for data collection, so as to obtain different behavior characteristics of pigs while avoiding large area occlusion of herd pigs.

In this study, a large amount of manual screening was performed on the daily video of pigs, and 185 video segments of pigs with abnormal behaviors were observed and intercepted. Finally, the abnormal behaviors of pigs included four kinds of attack, wall climbing, ear biting and stride climbing. Then, the inter-frame difference method is used to extract key frames, so as to avoid the following problems: Pig daily movement is relatively slow and the pigs rest and are dormant for a long time, if consecutive frames are used as training samples, there may be a large number of repeated samples during training, which can result in an ill-fitted model and decrease the robustness of the test due to a large number of similar samples being present. In order to improve the goodness of fit and robustness of the model, this paper uses the inter-frame difference method to extract key frames as valid images. When the difference between the sum of gray values of pixels between two frames is greater than the set threshold, it is defined as a key frame, and the key frame acquisition formula is shown in Equation (1):

$$D(i,j)=\begin{cases}1, & \sum\left|f_i(x,y)-f_j(x,y)\right|>T_i \quad (j>i)\\ 0, & others\end{cases} \tag{1}$$

wherein: $f_i(x,y)$ represents the gray value at the pixel point (x,y) of frame i, $f_j(x,y)$ represents the gray value at the pixel point (x,y) of frame j, and T_i represents the threshold value. In this study, it is set as 20% of the sum of the gray values of each pixel point of frame i, $D(i,j)=0$ represents that frame j is a key frame relative to frame i, $D(i,j)=0$ indicating that frame j is not a key frame relative to frame i, and frame i is a key frame relative to the previous key frame, where $j>i$.

In this study, the following two steps are used to process the obtained images and obtain the final experimental data set.

(1) In order to improve the computing speed and reduce the amount of calculation, this paper adjusts the image with a resolution of 1920×1080 pixels, and finally gets a resolution of 640×640 pixels.

(2) In order to make the data set required by the experiment conform to PASCALVOC format requirements, the data set needs to be labeled. In this study, LabelImg was used to label the pig image data set, and the behavior label of the pig was set as Normal (normal behavior). Green label boxes were used. Abnormal is highlighted in red, as can be seen in Fig. 1.

a. Wall climbing behavior b. Ear-biting behavior

c. Attacking behavior d. Mounting behavior

Fig. 1 - Data annotation

(Note: Abnormal behavior of pigs is shown in Fig. 1a to Fig. 1d, where a shows wall-climbing behavior, b shows ear-biting behavior, c shows attacking behavior, and d shows mounting behavior.)

After the above steps, a total of 1105 valid labels were obtained, including 323 for wall climbing, 356 for aggression, 302 for ear biting and 124 for climbing and crossing.

Deep Learning YOLOX Algorithm

Although currently, YOLOv5 is already performing well, some recent work on object detection has triggered the development of this new YOLOX algorithm *(Ge et al., 2021)*. The most important focus points in object detection are anchor-free detectors, advanced label assignment strategies, and end-to-end detectors. These new focal points are still not integrated into the YOLO algorithm, and YOLOv4 *(Bochkovskiy et al., 2020)* and YOLOv5 are still anchor-based detectors and use hand-crafted assigning rules for training. The backbone network of the YOLOX network structure, or Darknet53, had the Input layer, Neck layer as well as the Prediction layer improved. To address the problem that coupling of detection heads may affect the model's performance, in YOLOX, the decoupled head got improved. By replacing YOLO's detection head with a coupled head, the convergence rate of the model was significantly improved. To better the performance of YOLOX, Mosaic and MixUp strategies were added in data enhancement. As Anchor mechanism solves the problems of both complex detection heads and poor generalization performance in some way, in YOLOX, it was improved to be of Anchor-free mechanism, in which 3 groups of Anchor predictions were reduced to only 1 group of prediction for a feature map, thus reducing the number of parameters and GFLOPs and making the detection much faster.

To reduce training time and avoid extra super-parametric problems, in YOLOX, the OTA tag allocation was improved to get an approximate solution (SimOTA) with a dynamic top-k strategy employed. After undergoing a series of improvements and compared with other advanced target detection models, YOLOX showed significant improvements in both speed and accuracy to some extent. Network structure of YOLOX can be seen in Fig. 2.

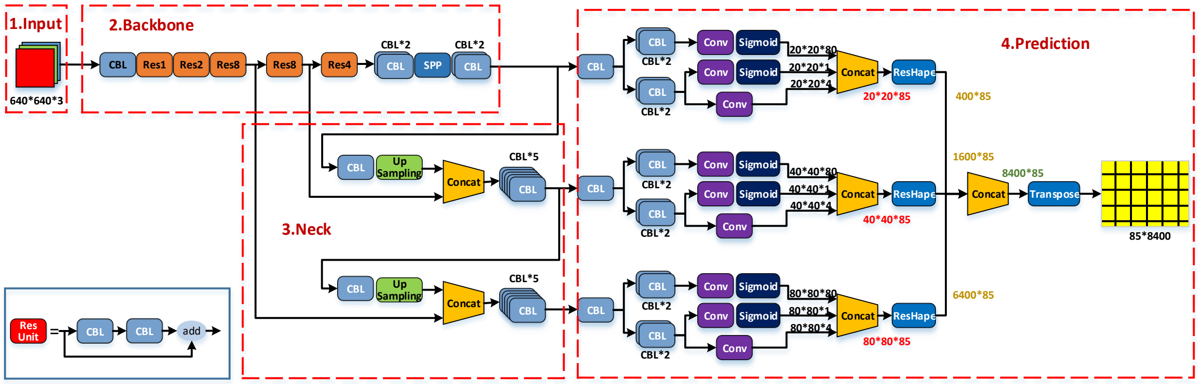


Fig. 2 - Network structure of YOLOX

As shown in Figure 2, CBL is widely used as a component of YOLOX network structure. CBL consists of three network layers (convolution layer, Batch Normalization layer, and Leaky ReLU function). In the convolution layer, multiple different convolution kernels are used to process the input image and different response characteristic graphs are obtained. The BN layer optimizes the network structure by normalizing the scattered data. To address the issue of neuron exhaustion in the ReLU function, the Leaky ReLU function introduces a non-zero slope in the negative half of the function, resulting in improved performance.

$$LeakyReLU(x)=\begin{cases} x, & x>0 \\ ax, & x\le 0 \end{cases} \tag{2}$$

The output of this function has a small slope for negative input. Since the derivative is not zero, this can reduce the presence of silent neurons, allowing gradient-based learning, although sometimes slower, but solves the problem that neurons remain silent after negative ReLU function input. Network structure of CBL can be seen in Fig. 3.

Fig. 3 - CBL structure

In YOLOX, ResNet *(He et al., 2016)* is a residual block, which contains two convolution blocks and an Add layer. The Add layer is the addition of tensors of the same dimension. The residual structure can ensure that the network structure can still converge at a very deep level, so that the model can continue to train.

ResX in YOLOX is composed of a CBL and X residual components. The CBL in front of each Res module plays a role of down-sampling, and each ResX contains 1+2*X convolution layers. Figure 4 shows the ResX structure.

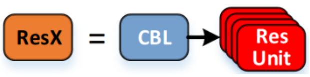


Fig. 4 – ResX structure

The SPP component consists of four parallel branches, which are the maximum pooling of kernel size 5×5, 9×9, 13×13 and a jump connection respectively. The main function of the SPP component is to effectively avoid image region clipping and zoom operations caused by image distortion. Additionally, it solves the problem of repeated feature extraction from images by the convolutional neural network, greatly improving the speed of candidate frame generation and saving computational cost.

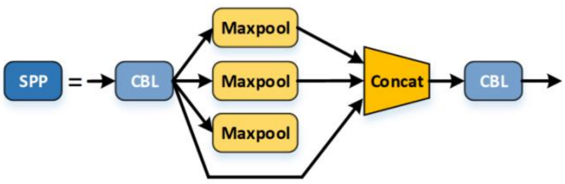


Fig. 5 - SPP structure

Model Evaluation Metrics

In this study, the commonly used evaluation indexes in target detection field, including IoU, Precision, Recall, P-R curve, AP value and mAP value, were used. IoU refers to the ratio of intersection and union of two detection frames, as shown in Equation (3):

$$IoU = \frac{S_A \cap S_B}{S_A \cup S_B} \tag{3}$$

where:

S_A represents the collection of pixels in the Prediction frame area, S_B represents the collection of pixels in the True frame area.

Precision represents detection precision, and Recall represents detection recall rate, for which the *equation* is as follows:

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

where:

TP (True Positive) refers to number of samples predicted by the model as individual frames of pigs (whose IOU is greater than the threshold) with a category tag that is consistent with the actual tag. FP (False Positive) refers to number of samples predicted by the model as individual frames of pigs with a category tag that is inconsistent with the actual tag. FN (False Negative) refers to the number of samples in which no individual pig is detected.

In the graph of P-R curve, Recall was taken as the abscissa and Precision was taken as the ordinate. Precision was negatively correlated with Recall. The larger the area (AP value) surrounded by P-R curve was, the better the model effect was. The mAP value represents the average accuracy of the model, as defined in equation (6) and (7):

$$AP = \int_0^1 p(r)dr \tag{6}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{7}$$

Considering that different IoU thresholds may have a direct impact on TP and FP values and thus cause fluctuations in mAP values, in this study, ten IoU thresholds of 0.5:0.05:0.95 (where 0.05 represents the grown step length) were used to calculate the mAP value. Finally, $mAP_{0.5}$, $mAP_{0.75}$, $mAP_{0.5:0.95}$, $mAP_{0.5:0.95\text{-}medium}$ and $MAP_{0.5:0.95\text{-}large}$ were used to measure the detection performance of the model, where $mAP_{0.5}$ represents the mAP value when the IoU threshold is 0.5, $mAP_{0.75}$ represents the mAP value when the IoU threshold is 0.75, $mAP_{0.5:0.95}$ represents the average value of mAP under the 10 thresholds of IoU at 0.5:0.05:0.95 respectively, while $mAP_{0.5:0.95\text{-}medium}$ indicates the median mAP value under the 10 thresholds of IoU at 0.5:0.05:0.95 respectively, and $mAP_{0.5:0.95\text{-}large}$ indicates the maximum mAP value under the 10 thresholds of IoU at 0.5:0.05:0.95 respectively.

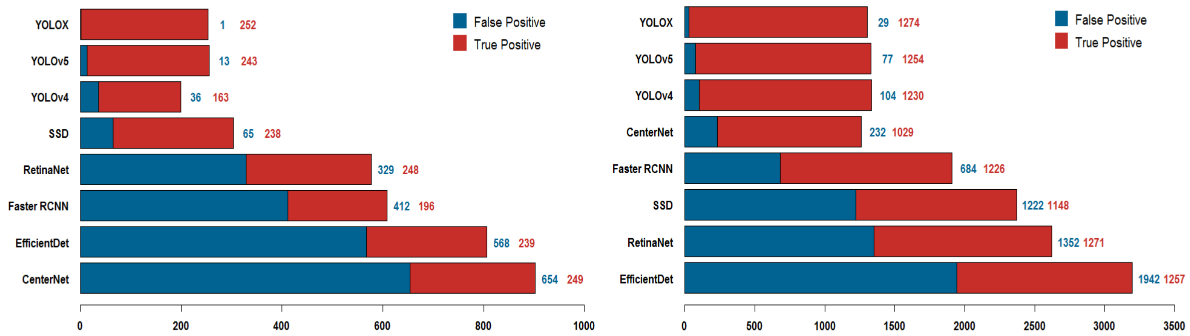
RESULTS

In this experiment, with PyTorch deep learning framework adopted, the operating system was Windows10, and GPU was RTX2080Ti. There were 1105 labels of abnormal behaviors in pigs, including 323 wall-climbing behaviors, 356 aggression behaviors, 302 ear-biting behaviors and 124 crawling behaviors. The 1105 collected image data sets were divided into training sets and test sets according to the ratio of 8:2, including 884 training sets and 221 test sets.

In the experiment, eight target detection models were used for comparison, respectively YOLOX *(Ge et al., 2021),* YOLOv5, YOLOv4 *(Bochkovskiy et al., 2020)*, RetinaNet *(Lin et al., 2017)*, EfficientDet *(Tan et al., 2019)*, Faster R-CNN *(Ren et al., 2017)*, CenterNet *(Duan et al., 2019)* and SSD *(Liu et al., 2016)*. In this study, the size of input image was set as 640*640 pixels, and epoch was set as 200, momentum 0.9, regular coefficient for weight decay 0.005. With learning rate decay mechanism adopted, the initial learning rate was set as 0.0001. Once the learning got stagnating, the learning rate of the model would attenuate at 2~10 times the rate, and the attenuation frequency was set as 80% and 90% of the maximum number of iterations.

TP/FP and P-R curve

In this study, the training set with the same experimental data was used to train eight models before tests were conducted on the same test set. With results compared and analyzed, the respective performance of the eight models was evaluated. In order to refine the results in the detection of abnormal behavior of pigs by models of YOLOX, YOLOv5, YOLOv4, RetinaNet, EfficientDet, Faster R-CNN, CenterNet and SSD, the TP and FP values predicted by the eight models were respectively calculated, as shown in Fig. 6. In the 221 images in the test data set of this paper, 260 frames of actual abnormal behavior and 1,326 frames of actual normal behavior were there.



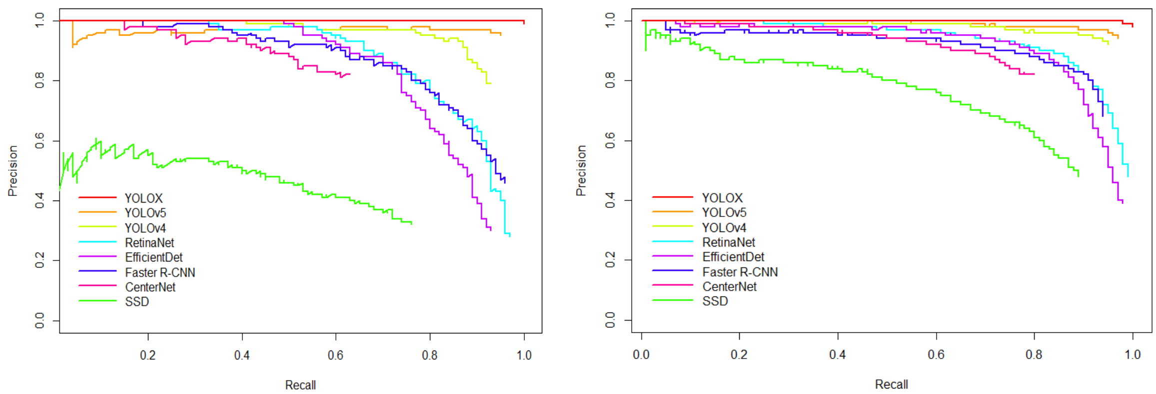
(a) Abnormal behavior *(b) Normal behavior*

Fig. 6 - TP and FP values predicted in detection by eight models

(Note: In case of the number of TP, the higher the value is, the better the model is; in case of the number of FP, the lower the value is, the better the model is.)

For the detection of abnormal and normal pigs, the TP value represents the number of pigs that correctly detected normal or abnormal behavior, and the FP value represents the number of pigs that incorrectly detected normal or abnormal behavior. As shown in Fig. 6, the TP numbers of the YOLOX model were 252 and 1274, respectively, which were 3-245 more than the other seven models. In addition, the FP numbers of the YOLOX model are 1 and 29, which are 12~1913 less than other models. The FP value of the YOLOX model is much lower than the other seven models. Although the difference in TP values of the eight models is not obvious, the huge difference in FP values finally makes the advantage of the YOLOX model more obvious.

In object detection, the P-R curve reflects the trade-off between the recognition accuracy and coverage ability of the model's correct object detection. The P-R curve for abnormal and normal behaviors of pigs detected by eight models, respectively YOLOX, YOLOv5, YOLOv4, RetinaNet, EfficientDet, Faster R-CNN, CenterNet and SSD, were as shown in Fig. 7.



(a) Abnormal behavior *(b) Normal behavior*

Fig. 7 - P-R curve

(1) As can be seen in Fig. 7, in the P-R curve for abnormal and normal behaviors of pigs, the curves for models YOLOX, YOLOv5 and YOLOv4 in YOLO series were all located at the upper right of the figure in comparison with that for other mainstream models, indicating that the models in YOLO series were better than other mainstream models in the detection of abnormal and normal behaviors of pigs. The curve for YOLOX model was above that for models YOLOv5 and YOLOv4, indicating that YOLOX model was superior to other models in the same series.

(2) As shown in Fig. 7(a), for abnormal behaviors of pigs, when the Recall value was less than 0.15, the Precision value for models YOLOX, YOLOv4, RetinaNet, EfficientDet, Faster R-CNN and CenterNet maintained around 1.0, and the difference was not significant, while the Precision value of SSD model remained around 0.6. As the Recall value increased, when the Recall value was greater than 0.6, the Precision value for models of YOLO series changed slightly, while the Precision value for other models decreased significantly, indicating that the models of YOLO series kept stable in the detection of abnormal behavior of pigs, and the YOLOX model owned the optimum detection performance.

(3) As can be seen in Fig. 7(b), for the normal behavior of pigs, as the Recall value increased, the Precision value for the models in YOLO series decreased slightly, while the Precision value for the YOLOX model kept staying at the top of the same series all the time, indicating that the YOLOX model had the optimum performance in the detection of normal behavior of pigs.

Detection accuracy of the model

In order to better evaluate the performance of the model, eight target detection models, respectively YOLOX, YOLOv5, YOLOv4, RetinaNet, EfficientDet, Faster R-CNN, CenterNet and SSD were used to predict the abnormal behavior of pigs on the same test set. The detection effects of the eight models were compared through different values of the mAP index, as can be seen in Table 1.

Table 1

Different values of mAP index for eight models

Model name	$mAP_{0.5}$	$mAP_{0.75}$	$mAP_{0.5:0.95}$	$mAP_{0.5:0.95\text{-}medium}$	$mAP_{0.5:0.95\text{-}large}$
YOLOX	98.0%	92.6%	81.2%	96.7%	98.0%
YOLOv5	94.5%	90.9%	86.3%	91.8%	94.5%
YOLOv4	92.3%	47.5%	58.6%	76.3%	92.3%
RetinaNet	90.1%	34.3%	51.7%	62.9%	90.1%
EfficientDet	86.7%	24.2%	46.3%	53.0%	86.7%
Faster R-CNN	86.6%	22.5%	45.6%	51.9%	86.6%
CenterNet	67.6%	15.2%	35.4%	39.6%	67.6%
SSD	54.3%	4.20%	21.6%	54.3%	54.3%

The experiment result showed that:

(1) In the detection of abnormal behavior of pigs, models of YOLO series behaved obviously much better than other mainstream target detection models. In $mAP_{0.5}$ index, the index value for YOLO series was 43.7 percentage points higher than that for other models; in $mAP_{0.75}$ index, the index for YOLO series was 88.4 percentage points higher than that for other models; in $mAP_{0.5:0.95\text{-}medium}$ index, the index for YOLO series was 59.6 percentage points higher than that for other models; in $mAP_{0.5:0.95\text{-}large}$ index, the index for YOLO series was 43.7 percentage points higher than that for other models. It can be seen that models of YOLO series performed significantly better than other target detection models. The pig detection process in the YOLO series models was completed in a neural network which can optimize the pig detection performance in an end-to-end way, while in other detection models, with sliding window or region proposal adopted, the detector could only get partial information of the image, though YOLO series models could obtain the target feature information of a whole image during training, that is why models in YOLO series could make good use of context information to not to make wrong predictions in the background, thus the detection accuracy was assured.

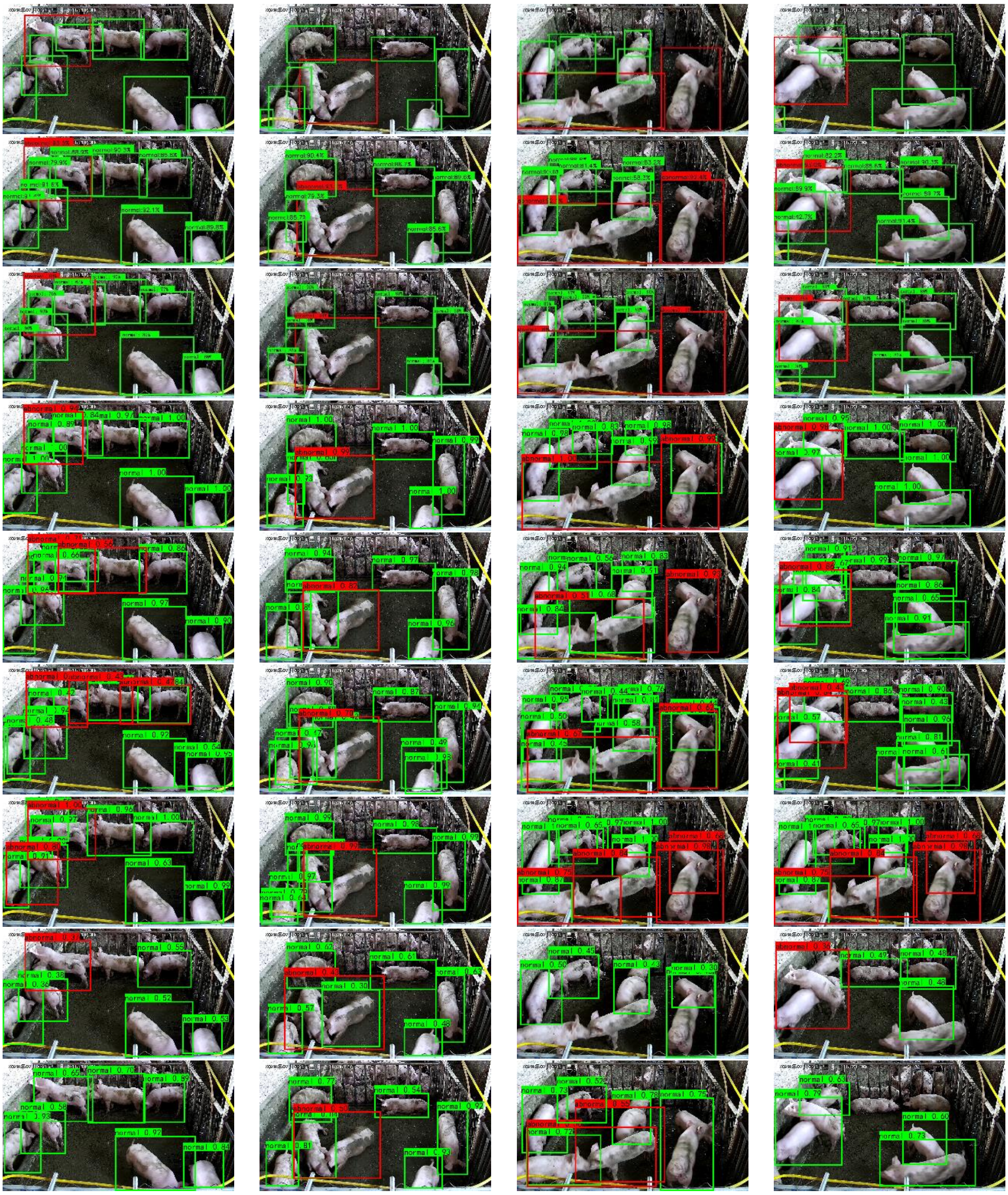
(2) Compared with the same series of YOLOv5 and YOLOv4 models, YOLOX model had the highest index value of 98.0 percentage points in $mAP_{0.5}$ index, 3.5 percentage points higher than that of YOLOv5 and 5.7 percentage points higher than that of YOLOv4 model, respectively; in $mAP_{0.75}$ index, YOLOX model owned the highest index value of 92.6 percentage points, 1.7 percentage points higher than that of YOLOv5 and 45.1 percentage points higher than that of YOLOv4. The $mAP_{0.5:0.95\text{-}large}$ index value was completely consistent with the $mAP_{0.5}$ index value, though the maximum value was in the YOLOX model. It can be seen that after the improved decoupled head and label allocation strategy were used in YOLOX model, the performance of YOLOX model got improved.

In $mAP_{0.5:0.95\text{-}medium}$ index, the index value for YOLOX model was 5.1 percentage points lower than that for YOLOv5, which may be influenced by the fact that a full consideration of reducing the complexity of the model is given in YOLOX and thus the detection accuracy was affected in some way.

Analysis of detection results of models

To further show the effects of eight kinds of models on prediction of abnormal behavior of pigs, different abnormal behaviors of pigs were predicted on the test set, among which four visualized results are as shown in Fig. 8. In the first row was the TP labeling frame, while all the other rows showed the prediction map for different models. In each prediction map, the displayed value at the top left corner of the prediction frame showed the confidence coefficient. The red frame in the map indicates abnormal behavior, while the green frame indicates normal behavior.

Real Box
YOLOX
YOLOv5
YOLOv4
RetinaNet
EfficientDet
Faster R-CNN
CenterNet
SSD



a. Wall climbing behavior *b. Ear-biting behavior* *c. Attacking behavior* *d. Mounting behavior*

Fig. 8 - Prediction results of eight models

1) In case of wall climbing, models YOLOX, YOLOv5, YOLOv4, RetinaNet, EfficientDet, Faster R-CNN and CenterNet could all detect the wall climbing behavior of pigs, while SSD model could not. Among all the models, false detection occurred in RetinaNet model for it detected the normal behavior of pigs as abnormal, also false detection and over-detection occurred in the EfficientDet model. In addition, false detection and missed detection of the normal behavior of occluded pigs occurred in the Faster R-CNN and CenterNet models.

2) In case of ear biting, all the eight models could detect abnormal behavior of pigs, among which missed detection occurred in SSD, RetinaNet and CenterNet models for normal behavior of occluded pigs could not be predicted, also false detection occurred in the Faster R-CNN and EfficientDet models.

3) In the case of attacking, as shown in Fig. 8(c), two groups of pigs displayed attacking behavior at the same time. YOLOX, YOLOv5, YOLOv4, RetinaNet, and EfficientDet models were all able to detect the attacking behavior of the two groups of pigs. False detections occurred in the EfficientDet and RetinaNet models as they detected abnormal behavior of pigs as normal behavior. Additionally, false detection occurred in the Faster R-CNN model because the pigs displaying attacking behavior were not in the same frame at all. Furthermore, missed detection occurred in CenterNet and SSD models as they failed to detect the normal behavior of occluded pigs.

4) In case of mounting, all the models could detect abnormal behavior of pigs, while SSD could not. Among the models that could detect normal behavior of pigs, false detection occurred in the EfficientDet model, and missed detection occurred in YOLOv4, RetinaNet, Faster R-CNN as well as CenterNet models for they could not detect the normal behavior of occluded pigs.

In conclusion, different degrees of missed detection and false detection occurred in the prediction of different abnormal behaviors of pigs in models YOLOv4, RetinaNet, EfficientDet, Faster R-CNN, CenterNet and SSD. Both YOLOX and YOLOv5 models could detect behaviors of pigs in a correct way, though between the two, the YOLOX model had the maximum confidence in the prediction results, so the YOLOX model owned a better detection effect, being able to have the abnormal behavior of pigs detected effectively under different background conditions, thus the occurrence of missed detection and false detection was lowered, also the detection confidence coefficient for it was higher than that for the other seven models.

CONCLUSIONS

1) The image obtained by the frame differential method is used as the research material for pig anomaly detection, which overcomes the problems of adhesion and occlusion that are easy to occur in the pig house environment, and makes the 24-hour anomaly detection application possible.

2) Compared with other target detection models, the YOLO series models have stronger detection performance, and the detection accuracy of the YOLOX model in the YOLO series models reaches 98.0%.

3) The P-R curves of the YOLOX model are all located at the upper right of the figure, and the precision value changes very little, while other models have a significant decrease, indicating that the YOLOX model has better stability for the detection of abnormal behavior of pigs.

4) A comparative analysis is undertaken of different model prediction results in the detection of abnormal pig behaviour under conditions of adhesion and obstruction. The YOLOX model has the lowest false detection and missed detection rates, and can correctly detect abnormal pig behavior, indicating that the YOLOX model has better detection effect and can provide technical reference for real-time monitoring of scientific pig breeding.

ACKNOWLEDGEMENT

This research, titled ‘Anomaly Detection for Herd Pigs Based on YOLOX’, was funded by the Basic Research Program of Shanxi Province (202103021223241), the National Natural Science Foundation of China (31772651).

REFERENCES

[1] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv*:2004.10934. https://doi.org/10.48550/arXiv.2004.10934

[2] Cao, Y., Chen, K., Loy, C. C., & Lin, D. (2020). Prime sample attention in object detection. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp.11583-11591. https://doi.org/10.48550/arXiv.1904.04821

[3] Chen, C., Zhu, W., Ma, C., Guo, Y., Huang, W., & Ruan, C. (2017). Image motion feature extraction for recognition of aggressive behaviors among group-housed pigs. *Computers and Electronics in Agriculture*, 142, 380-387. https://doi.org/10.1016/j.compag.2017.09.013

[4] Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., & Tian, Q. (2019). Centernet: Keypoint triplets for object detection. *In Proceedings of the IEEE/CVF international conference on computer vision*. pp. 6569-6578. https://doi.org/10.48550/arXiv.1904.08189

[5] Fang, J., Hu, Y., Dai, B., & Wu, Z. (2021). Detection of group-housed pigs based on improved CenterNet model. *Transactions of the Chinese Society of Agricultural Engineering*, 37(16), 136-144. https://doi.org/10.11975/j.issn.1002-6819.2021.16.017

[6] Gao Y, Chen B, Liao H M, et al. (2019). Recognition method for aggressive behavior of group pigs based on deep learning. *Transactions of the Chinese Society of Agricultural Engineering*, 35(23), 192-200. https://doi.org/10.11975/j.issn.1002-6819.2019.23.024

[7] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv*:2107.08430. https://doi.org/10.48550/arXiv.2107.08430

[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770-778.

[9] https://doi.org/10.1109/CVPR.2016.90

[10] He, Z., Yu, S., Li, H., Li, S. & Li, M. (2019). Biological characteristics of African swine fever virus and its prevention and control measures in pork food production. *Transactions of the Chinese Society of Agricultural Engineering*, (10), 299-304. https://doi.org/10.11975/j.issn.1002-6819.2019.10.038

[11] Lee, J., Jin, L., Park, D., & Chung, Y. (2016). Automatic recognition of aggressive behavior in pigs using a Kinect depth sensor. *Sensors*, 16(5), 631 – 641. https://doi.org/10.3390/s16050631

[12] Li, D., Zhang, K., Li, X., Chen, Y., Li, Z., & Pu, D. (2019). Mounting Behavior Recognition for Pigs Based on Mask R-CNN. *Transactions of the Chinese Society for Agricultural Machinery*, 50(S1), 261-266+275. https://doi.org/10.6041/j.issn.1000-1298.2019.S0.041

[13] Li, J., Li, Y., Niu, F., Li, R., Zhang, T., & Jing, B. (2021). Pig Diet Behavior Detection Method Based on YOLOv4. *Transactions of the Chinese Society for Agricultural Machinery*, 52(03), 251-256.

[14] https://doi.org/10.6041 / j.issn.1000-1298.2021.03.028

[15] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *In Proceedings of the IEEE international conference on computer vision.* pp. 2980-2988.

[16] https://doi.org/10.1109/TPAMI.2018.2858826

[17] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot MultiBox detector. *In European conference on computer vision*. pp. 21-37. Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0_2

[18] Oczak, M., Viazzi, S., Ismayilova, G., Sonoda, L. T., Roulston, N., Fels, M., & Vranken, E. (2014). Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network. *Biosystems Engineering*, 119, 89-97. https://doi.org/10.1016/j.biosystemseng.2014.01.005

[19] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 39(6), 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[20] Tan, M., Pang, R., & Le, Q. V. (2020). EfficientDet: Scalable and efficient object detection. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp.10781-10790. https://doi.org/10.48550/arXiv.1911.09070

[21] Viazzi, S., Ismayilova, G., Oczak, M., Sonoda, L. T., Fels, M., Guarino, M., Berckmans, D. (2014). Image feature extraction for classification of aggressive interactions among pigs. *Computers and Electronics in Agriculture*, 104, 57-62. https://doi.org/10.1016/j.compag.2014.03.010

[22] Xue, Y., Li, S., Zheng, C., Gan, H., Li, C., & Jiu, H. (2021). Posture change recognition of lactating sow by using 2D-3D convolution feature fusion [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 37(09), 230-237. https://doi.org/10.11975/j.issn.1002-6819.2021.09.026

[23] Xue, Y., Zhu, X., Zheng, C., Mao, L., Yang, A. et al. (2018). Lactating sow postures recognition from depth image of videos based on improved Faster R-CNN. *Transactions of the Chinese Society of Agricultural Engineering*, 34(9), 189-196. https://doi.org/10.11975/j.issn.1002-6819.2018.09.023

[24] Yang, Q., Xiao, D., & Lin, S. (2018). Feeding behavior recognition for group-housed pigs with the Faster R-CNN. *Computers and electronics in agriculture*, 155, 453-460. https://doi.org/10.1016/j.compag.2018.11.002

[25] Yan, H., Lu, Z., Cui, Q., Hu, Z., & Li, Y. (2019). Detection of facial gestures of group pigs based on improved Tiny-YOLO. *Transactions of the Chinese Society of Agricultural Engineering*, 35(18), 169-179. https://doi.org/10.11975/j.issn.1002-6819.2019.18.021

[26] Yan, H., Lu, Z., Cui, Q., & Hu, Z. (2020). Multi-target detection based on feature pyramid attention and deep convolution network for pigs. *Transactions of the Chinese Society of Agricultural Engineering*, 36(11), 193-202. https://doi.org/10.11975/j.issn.1002-6819.2020.11.022

[27] Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *In European conference on computer vision*. Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_53

[28] Zhang, Z., Fidler, S., & Urtasun, R. (2016). Instance-level segmentation for autonomous driving with deep densely connected MRFs. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* pp. 669-677. https://doi.org/10.1007/978-3-319-46466-4_19

[29] Zheng, C., Zhu, X., Yang, X., Wang, L., Tu, S., & Xue, Y. (2018). Automatic recognition of lactating sow postures from depth images by deep learning detector. *Computers and electronics in agriculture*, 147, 51-63. https://doi.org/10.1016/j.compag.2018.01.023