ISSN 2616-5910 Vol. 5, Issue 1: 102-112, DOI: 10.25236/AJALS.2024.050114

Target localisation of rotating frame in ear root area of pigs based on improved YOLOv8_OBB

Qi Li, Dong Ren*

School of Electronic Information and Artificial Intelligence, Shaanxi University of Science & Technology, Xi'an, 710021, China
1195092541@qq.com
**Corresponding author*

Abstract: *In the thermal infrared temperature measurement scenario of pigs, the temperature at the root of the ear is closest to the body temperature. Due to the weak light in the pig house scene, the inconspicuous characteristics of the pig's ear area, and the mutual blockage caused by the pig's posture changing at any time, it is difficult to accurately locate the pig's ear root. Therefore, in order to reduce the impact of low light conditions on the positioning of the pig's ear root, the Retinex algorithm based on bilateral filtering is used to preprocess the input image; an improved rotating target detection method of YOLOv8_OBB is proposed, and the Shuffle Attention mechanism is introduced in the neck, which effectively integrates spatial and channel attention mechanisms to enhance the feature extraction capability of the model; the original feature extraction method is replaced by the Shuffle Attention mechanism, which effectively integrates spatial and channel attention mechanisms to enhance the feature extraction capability of the model. Feature extraction capability; the original backbone network is replaced with MobileNetV3 network structure to reduce the number of parameters and computation of the model. The improved YOLOv8_OBB algorithm improves the accuracy by 7.3%, recall by 7.6%, and average accuracy by 7.1% compared with the baseline algorithm, and is more robust under low light conditions and target occlusion, which provides the basis for intelligent temperature measurement in pigs. Effective positioning out of the ear root is of practical significance for modernised intelligent temperature measurement in pigs.*

Keywords: *Pig population, YOLOv8_OBB, MobileNetV3, Hybrid attention mechanism*

1. Introduction

In the process of pig breeding, the health of pigs often determines the development and economic benefits of pig farming. The morbidity of pigs is often accompanied by changes in body temperature, so real-time temperature monitoring of pigs can help screen out unhealthy pigs in pens and detect and deal with abnormalities in a timely manner, which is of great significance for the enhancement of breeding and production efficiency, the refined management of farms, and the control of the quality of pig breeding [1-3].

Traditional artificial temperature monitoring mode, need to rely on a large number of labour costs and difficult to achieve long-term continuous temperature monitoring, at present, most of the pig farms, intelligent, automated feeding has been basically popularized with the help of infrared imaging technology combined with computer vision, can avoid the shortcomings of the traditional temperature measurement methods, can be achieved without contact, low-cost real-time monitoring of pig temperature, to achieve the wisdom of the pig's breeding; due to weak light, poor contrast, poor contrast, and poor quality control of pig breeding, it is of great significance to the improvement of breeding and production efficiency. Due to the weak light and poor contrast in the pig house, the pig ear root feature information cannot be effectively captured, resulting in low detection accuracy, so there are certain difficulties in the detection of pig ear root targets under low light conditions.

Computer vision-based breeding detection methods are characterised by non-contact and continuous automatic monitoring. At present, they are mainly divided into two categories, traditional target detection algorithms and target detection algorithms based on deep learning, traditional detection algorithms are mainly through manual feature extraction, poor robustness in the case of complex scenes and high computational complexity, target detection algorithms based on deep learning is to use deep learning technology to automatically extract the hidden feature information of the input image, and to classify and predict the samples with higher accuracy, at present, target detection algorithms based on Deep learning detection algorithms are mainly divided into two kinds of the first is based on one-stage target detection algorithms[4], the target task is considered as the regression task of the whole image, the target recognition accuracy is low but relatively fast, common algorithms are YOLO (you only look once) series, SSD (single shot multibox detector) algorithm; the second is based on two-stage target detection algorithms[5], the second is based on two-stage target detection algorithms[6].) algorithm; the second is based on two-stage (two-stage) target detection algorithm is in the first stage to generate the region proposal, in the second stage of the region of interest of the content of the classification and regression, the recognition of the target accuracy is relatively high but the detection speed is relatively slow R-CNN (region-convolutional neural network, regional convolutional neural network) [7]. convolutional neural network)[8], SPP-Net (spatial pyramid pooling network),Fast-RCNN[9], Faster-RCNN [10] and so on. The use of infrared imaging in combination with the commonly used target detection networks Faster-RCNN, SSD, and YOLO generally anchors the target by a horizontal detection frame, but it is not applicable to pig ear root target detection. Since the pig's activities are random, the horizontal detection frame will cause the target range to be inaccurate (containing a large amount of background area), and the pig's ear posture will change morphologically with the head rotation when the pig is ingesting, tilting the head, and bowing the head, etc., which will reduce the detection accuracy and make it impossible to accurately obtain the exact position of the ear root of the pig. To address the above problems, scholars at home and abroad have done relevant research, Yang et al [11] added an angle prediction branch on the basis of Faster RCNN to form a rotating region convolutional neural network model, but the model exists in the detection speed is slow, the angle regression is inaccurate and so on; Xu et al [12] used a new rotating frame representation to establish a rotating target detection (gliding-vertex) model, but this model is not suitable for the detection of rotating targets. vertex) model, but this model has problems such as inaccurate detection frame prediction and slow detection speed; Yang et al [13] used a stepwise regression method (R3Det) from coarse-grained to fine-grained to detect the target quickly and accurately, but this method has problems such as slow training speed; Zhou et al [14] used a polar coordinate system to establish an anchorless frame by using a polar coordinate system in the model. rotation-equivariant detector (P-RSDet) by using a polar coordinate system in the model, but the detection accuracy is not high. As a result, it is difficult for the existing rotating target detection methods to simultaneously take into account the detection accuracy, training speed and detection speed[15].

In order to address the above problems, taking the pig as the research object, Retinex algorithm is introduced to enhance the image with low light; the pig head posture in visible light image is detected by the improved YOLOv8_OBB model, and the use of rotary target detection method with angular information can overcome the limitations of the detection method of the horizontal detection frame localisation target detection method, improve the accuracy and detection speed of the tracking and localisation of the ear, in order to meet the demand for temperature monitoring in modern pig farming.

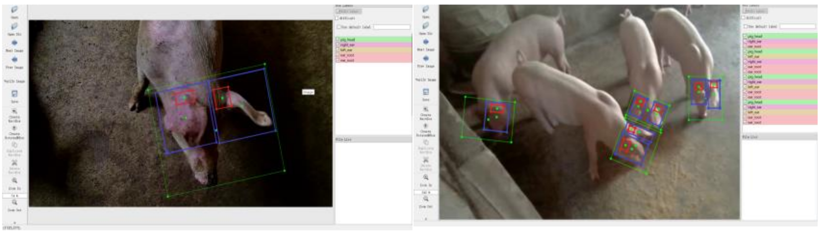
2. Materials and Methods

2.1. Dataset Creation

The experimental data were collected in a pig barn of Henan Cattle Industry Co. In order to solve the problem of the lack of stability and adaptability of the light detector due to the images collected by a single device, the visible light data were collected using a 100-million-pixel mobile phone camera and a 200W pixel industrial camera of JERUSALEM as the collection devices, of which the resolutions were 1920*1080 and 640*480, respectively. 100 segments of the visible light video of multiple pigs were collected in total under the complex background. A dataset of 2700 pig head images was finally formed.

RoLabelme annotation tool was used to annotate the target frame of the collected dataset. The green target box is pig head area (pig_head), the blue box is pig ear (pig_ear), the red box is pig ear root (ear_root), and the pig annotation is shown in Figure 1 below.

ISSN 2616-5910 Vol. 5, Issue 1: 102-112, DOI: 10.25236/AJALS.2024.050114



（a）Single pig labelling *（b）Multiple pigs labelled*

Figure 1: Dataset collection and annotation

After completing the dataset annotation, in order to improve the model generalization and robustness, the original annotated dataset was subjected to data enhancement operations such as cropping, scaling, adding Gaussian noise, masking, etc., and finally formed 14250 sheets of dataset. As shown in Table 1 the enhanced dataset was divided into three parts according to the ratio of 7:2:1;9975 sheets for the training set, 2850 sheets for the validation set and 1425 sheets for the test set, and the number of pig heads was categorized.

Table 1: Dataset Partitioning

Type	Training set /piece	Validation set /piece	Test set /piece
single pig	4928	1423	712
Two pigs	2365	711	356
many pigs	2682	716	357

2.2. Retinex algorithm image enhancement based on bilateral filtering

Retinex is an image enhancement method that adjusts the colour and brightness of an image by simulating the human visual system[16].If the original image is $S(x,y)$ The luminance image is $L(x,y)$ The reflected image is $R(x,y)$ .Due to the similarity between the human visual model and the logarithmic domain model, the single-scale Retinex algorithm formula can be obtained as shown in equation (1)

$$R_i(x,y) = \log S_i(x,y) - \log\left[G(x,y) * S_i(x,y)\right] \tag{1}$$

In equation(1), i is the i^{th} colour channel, i=1 for single-channel images, * denotes the convolution operation; is the standard Gaussian surround function, which is calculated as shown in equation(2).

$$G(x,y) = \lambda \cdot \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{2}$$

In equation (2), λ the needs of $\int\int G(x,y)dxdy = 1$ σ is the scale parameter. The kernel function of bilateral filtering is the combined result of the spatial domain kernel and the pixel range domain kernel[17], which can preserve the edge effect better. The definition of bilateral filter is shown in equation (3).

$$h(x) = k^{-1}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(\xi)c(\xi,x)s[f(\xi), f(x)]d\xi dx \tag{3}$$

Normalisation parameters for equation (3) are as in equation (4)

$$k(x) = k^{-1}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} c(\xi,x)s[f(\xi), f(x)]d\xi dx \tag{4}$$

In equations (3)and(4),denotes the distance between ξ the proximity point and the centre point x The bilateral filtering is extended to a Gaussian kernel[18],specifically equation.(5) $s[f(\xi), f(x)]$ denotes the brightness similarity between the proximity point ξ and the centre point as shown x in equation (6) $f(x)$ and $h(x)$ denote the luminance values of the input and output images at the centre point, respectively.

ISSN 2616-5910 Vol. 5, Issue 1: 102-112, DOI: 10.25236/AJALS.2024.050114

$$c(\xi, x) = \exp^{-\frac{d(\xi,x)^2}{2\sigma^2}} \tag{5}$$

$$s[f(\xi), f(x)] = \exp^{-\frac{\delta[f(\xi),f(x)]^2}{2\sigma^2}} \tag{6}$$

In equation (5), $d(\xi, x)$ denotes the Euclidean distance between 2 points;In equation (6), $\delta[f(\xi), f(x)]$ The difference between the 2 luminance values indicated.

The implementation of Retinex algorithm based on bilateral filtering is firstly to take the logarithm of the image in the piggery by single scale Retinex algorithm before inputting the image to the network, and then do the local contrast enhancement to improve the contrast of the image, and for the better control of the image brightness effect[19], at the same time, the brightness estimation method based on the brightness estimation method based on bilateral filtering is used to estimate the brightness for the image of the piggery after the taking of the logarithm. The workflow of Retinex algorithm based on bilateral filtering is shown in Fig 2.

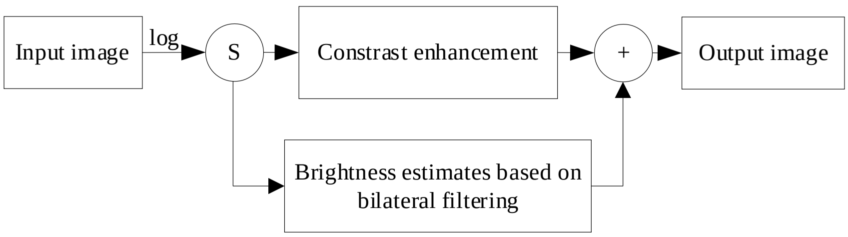


Figure 2: Retinex piggery image enhancement algorithm based on bilateral filtering

Using Retinex image enhancement algorithm based on bilateral filtering, the images under low light conditions are enhanced and the results are shown in Fig. 3 (a is before image enhancement and b is after image enhancement).

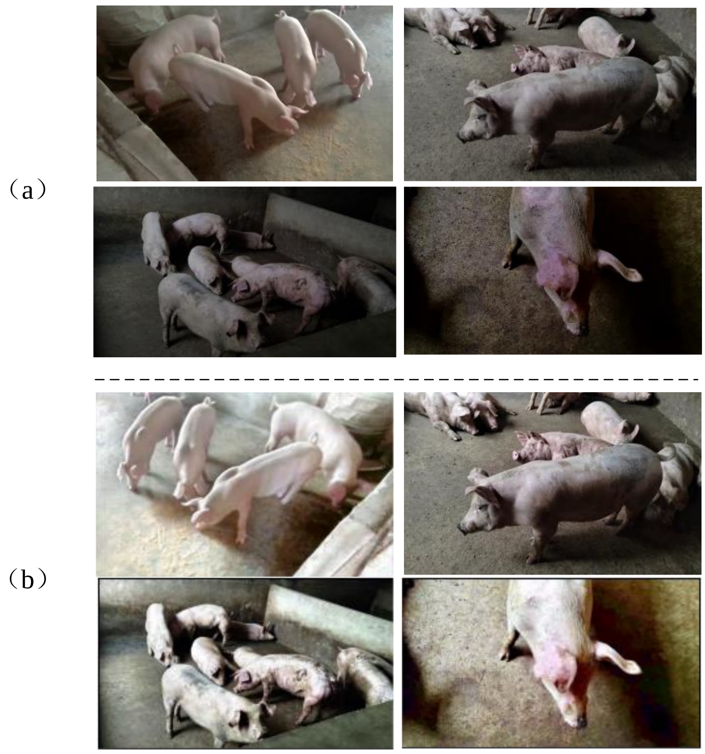


Figure 3: Low light enhancement

ISSN 2616-5910 Vol. 5, Issue 1: 102-112, DOI: 10.25236/AJALS.2024.050114

2.3. Comparison of target detection algorithm parameters

In computer vision, there are various target detection algorithms, the more common of which is the one-stage SSD algorithm, which detects targets by employing multiple predefined anchoring frames on feature maps at different scales, and is less effective in detecting pig ear root targets. The two-stage algorithm Faster R-CNN divides the image into multiple candidate regions, extracts each region and uses a classifier to identify the target, but the classification regression of these regions requires two independent steps, so it is slow in the actual detection of pig ear root.YOLOv8 in the YOLO (You Only Look Once) series of algorithms is a fast YOLOv8 in the YOLO (You Only Look Once) series of algorithms is a fast single-stage target detection method, especially in image recognition with high accuracy while ensuring a small number of model parameters, based on the scaling coefficient, the model is divided into five different scales of the model N/S/M/L/X, the comprehensive model volume and detection efficiency comparison can be seen in this paper, the selection of the S model.As can be seen in Table 2 below the Faster R-CNN and SSD algorithm structure model is large in size and the accuracy is not high enough. the original model of YOLOv5s can realize lightweight, but the accuracy still needs to be improved, and the comprehensive comparison of YOLOv8s model is more in line with the model selection of this experiment.

Table 2: Comparison of parameters of each algorithm

network structure	P/%	R/%	AP_{50}/%	$AP_{50,95}$/%	model volume(MB)	parameters
Faster R-CNN	37.0	52.4	43.8	19.0	108.17	137.09
SSD	91.1	13.6	31.4	11.3	90.61	26.29
YOLOv5s	78.1	51.2	59.7	33.0	13.68	7.05
YOLOv8s	80.1	54.5	62.8	44.9	21.45	11.2

2.4. Introduction of lightweight network architecture MobileNetV3

Without affecting the accuracy of the premise, the speed and volume of the two main aspects of the network to improve the lightweight, lightweight network can largely reduce the size of the model, reduce the number of redundant parameters, so as to obtain a more streamlined model.MobileNetV3 in the mobile image classification, target detection, semantic segmentation and other fields have achieved excellent performance[20].MobileNetV3 combines neural architecture search (NAS) and manually designed network structure. The Platform-Aware NAS is used for global search to get the overall framework, and then the complementary strategy of local fine-tuning through NetAdapt search is used to effectively match the model with the best effect, based on which hswish is used instead of the swish function, thus reducing the amount of model operations and thus improving the detection performance. The formula is shown in (7) below.

$$h - swish[x] = x\frac{ReLU6(x+3)}{6} \tag{7}$$

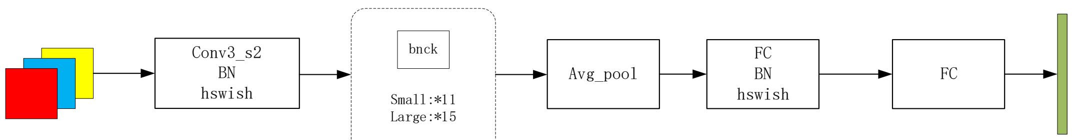


Figure 4: MobileNetV3 network architecture

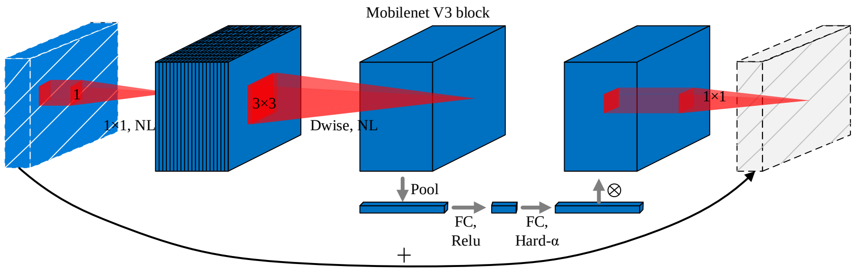


Figure 5: Bneck network structure

The MobileNetv3 network structure is shown in Fig. 4. As shown in Fig. 5, the bneck network

ISSN 2616-5910 Vol. 5, Issue 1: 102-112, DOI: 10.25236/AJALS.2024.050114

structure is a basic module in MobileNetv3, whose main function is to implement the channel separable convolution, SE channel attention mechanism and residual connection.

2.5. Introducing the Shuffle Attention Module

The SA (Shuffle Attention) module operation steps are divided into four segments: feature grouping, channel attention, spatial attention, and feature aggregation. The input feature map is first divided into multiple sub-features according to the channel dimension, then channel statistics are generated by global average pooling, group normalization is used to generate spatial channel statistics, and finally all the sub-features are aggregated, and then communication between different sub-features is achieved by the channel shuffle operation. As the process of accurately locating the ear root of the pig firstly needs to identify the head of the pig, then divide the ear of the pig based on the identified head region, and finally accurately locate the ear root of the pig. Where the pig head pixels average 331pixel*280pixel pig ear pixels average 84pixel*63pixel pig ear root area 28pixel*16pixel (usually using the standard provided by the COCO dataset, objects smaller than 32*32 pixels are defined as small targets) scale difference is large, (the network for the relatively small scale difference of the can be accurately recognised, the recognition ability of relatively large or relatively small scales decreases for scales that are relatively well recognised, leading to a decrease in the average detection accuracy of the network).The SA attention mechanism module uses a mixing unit to combine the channel attention module and the spatial attention module, thus enhancing the information of the pig ear root features, suppressing the noise in the source image, and using channel random mixing to achieve the transfer of information between the different sub-features. Information transfer, increasing the robustness of the model and effectively reducing the risk of overfitting. The Shuffle Attention mechanism is introduced into the neck part of the YOLOv8_OBB model to improve the detection performance of the model when dealing with multi-scale information in multiple poses of the pig, and save computational resources, and the structure of the introduced Shuffle Attention is shown in Fig. 6.

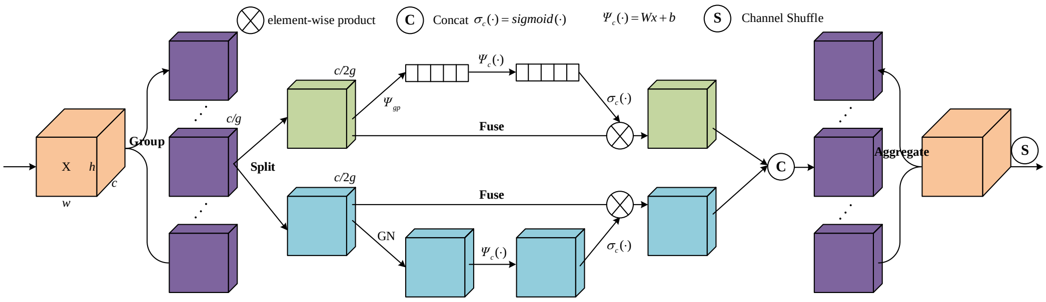


Figure 6: Shuffle Attention Attention Mechanism Module

2.6. Introduction of the BiFPN module

Due to the differences in target size in the pig head, ear, and ear root detection tasks with multiple different scales weighted bi-directional feature pyramid network (BiFPN) is a novel network structure for computer vision such as target detection and semantic segmentation, and its core idea is composed of two aspects: efficient bidirectional cross-scale connectivity and weighted feature map fusion. In convolutional networks, different features correspond to different network layers. Shallow networks can learn high-resolution features and pixel-level features, while deep networks can learn semantic features. However, in a real-world scenario of detecting a small target such as the ear root area of a pig, small-scale feature maps cannot provide enough resolution information. To compensate for this loss, larger scale feature maps need to be combined to enhance the complementarity between feature layers. The combination of shallow pixel-level features and deep semantic features can be used to locate the ear root area of pigs more accurately in scenes with strong lighting.

The BiFPN structure is shown in Fig. 7, the blue arrows represent the top-down path of the table, which is used to convey the semantic information of the upper layer features. The red arrows represent bottom-up paths for conveying the positional information of the underlying features. The purple part represents a new edge added between an input node and an output node in the same layer. Features from different scales are fused through the dual top-down and bottom-up paths to reduce the loss of feature information due to too many layers of downsampling in the model. BiFPN can effectively handle the information interaction between features from different scales and the fusion between different feature maps when constructing the network, thus improving the performance and efficiency

ISSN 2616-5910 Vol. 5, Issue 1: 102-112, DOI: 10.25236/AJALS.2024.050114

of the model.

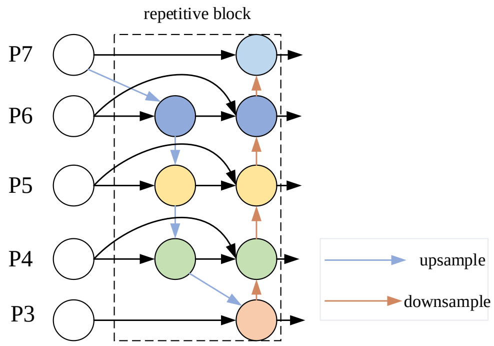


Figure 7: BiFPN structure

2.7. Improved YOLOv8 network

The original network needs to use a large model to ensure the accuracy of model detection when detecting pig ear roots, which is not easy to be deployed on small AI devices. Therefore, an improved YOLOV8_OBB network is proposed, which introduces a lightweight MobileNetV3 network structure for reducing the model parameters, which in turn improves the prediction speed, and adds a Shuffle Attention mechanism. The channel attention module and spatial attention module are combined through the shuffle unit to achieve pig ear root detection accuracy. Figure 8 below shows the improved YOLOv8 network.

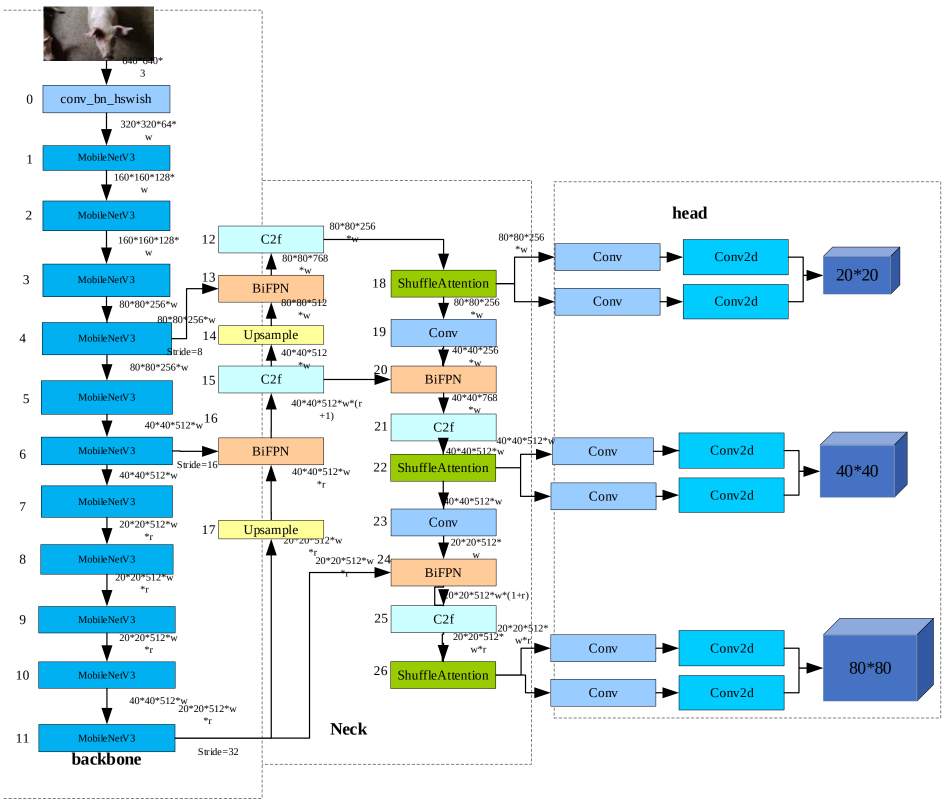


Figure 8: Improved YOLOv8_OBB network structure

3. Results and analyses

3.1. Experimental data and environment

In order to objectively evaluate the performance of the improved YOLOv8_OBB model, the graphics card was configured as an NVIDIA GeForce RTX 2080 Ti GPU with 16G RAM, the operating

ISSN 2616-5910 Vol. 5, Issue 1: 102-112, DOI: 10.25236/AJALS.2024.050114

system was Ubuntu 18.04, and the algorithm was implemented based on Python language. The Faster R-CNN, SSD, YOLOv5s, original YOLOv8 and improved YOLOv8_OBB models were tested using the self-constructed pig dataset, respectively.

3.2. Evaluation indicators

The evaluation criteria adopted in this experiment for ear root detection in pigs include precision (P),recall(R),average precision(AP_{50})when the IOU threshold is gradually increased between 0.5 and 0.95, and calculated parameters.

$$P=\frac{T_P}{T_p+F_P} \tag{8}$$

$$R=\frac{T_P}{T_p+F_N} \tag{9}$$

$$AP_{50}=\int_0^1 P(R)dR \tag{10}$$

$$\mathrm{m}AP=\frac{1}{N}\sum_{i=1}^{N}A_{p_i} \tag{11}$$

In Eqs. (8)~(11): P denotes detection accuracy; R denotes recall; AP_{50} denotes average accuracy; mAP denotes average precision.

3.3. Ablation experiment

In order to assess the degree of optimization of each part of the improvement on the overall algorithm performance, this paper designs ablation experiments to verify the effectiveness of each part of the improvement, and the experimental results are shown in Table 3. Performing image enhancement adding Shuffle Attention attention mechanism and multi-scale BiFPN structure both improve the detection performance. Performing image enhancement before the pig images are collected and inputted into the detection network reduces the influence of low light on the detection accuracy, and improves the mAP index of the model by 0.8 percentage points; by adding the Shuffle Attention attention mechanism module, fusing the attention weights in both channel and spatial dimensions, increasing the sensory field of the network, and strengthening the model's ability to pay attention to the key information in the feature map, which is It is conducive to the expression of feature map information, so that the network can learn more effective information, so that the mAP index of the network model rises by 3.8 percentage points;finally, by incorporating the BiFPN module, it effectively handles the feature information of different scales of the pig's head, ear, and ear root parts, as well as the effective fusion of the different feature maps, so that it can locate more accurately in the pig's ear root, and the model improves the mAP by 2.3 percentage points on the original basis. 2.3 percentage points. From the above, it can be seen that all the three improvement methods have improved the mAP of the model, and the mAP of the model after combining the three improvement methods is 97.0%, which is 7.4 percentage points higher than that of the original YOLOv8. It proves that the improved method has good effect on the weak light and the existence of multi-scale detection targets.

Table 3: Results of ablation experiments

groups	image enhancement	Shuffle Attention	BiFPN	mAP/%
1	×	×	×	89.6
2	√	×	×	90.2
3	×	√	×	93.4
4	×	×	√	90.6
5	√	√	×	94.7
6	√	×	√	91.1
7	×	√	√	95.5
8	√	√	√	97.0

Shuffle Attention, replacement attention mechanism mAP, mean accuracy; ×, indicates that this improvement method was not chosen; √, indicates that this improvement method was chosen.

ISSN 2616-5910 Vol. 5, Issue 1: 102-112, DOI: 10.25236/AJALS.2024.050114

3.4. Comparison between different algorithms

Using the pig ear root test set taken by ourselves in this paper, the algorithm of this paper is compared with Faster R-CNN, SSD, YOLOv8s original network model, and the results are obtained in Table 4, from which it can be seen that, Faster R-CNN is a two-stage network, the detection accuracy is high compared to that of a one-stage network but the real-time performance is poorer; the detection accuracy of a one-stage SSD network is lower, the But the detection speed is high compared to two-stage Faster R-CNN. The algorithm in this paper improves the detection speed, detection accuracy and recall, compared with the two-stage target detection network Faster R-CNN, the average accuracy is improved by 2.7%, the detection speed is improved by 17FPS; compared with the one-stage SSD detection network the average accuracy is improved by 13.9%, based on the dataset of this paper the improved algorithm achieves experimental results are better than Faster R-CNN, SSD and the original YOLOv8 algorithm.

Table 4: Comparison of target detection results of different algorithms

algorithm	P/%	R/%	mAP/%	FPS
Faster R-CNN	89.9	89.8	92.1	14
SSD	84.6	79.6	82.9	27
YOLOv8s	89.1	84.2	89.7	30
The algorithms in this paper	92.2	90.2	93.4	30

P, Precision; R, Recall; mAP, Mean average precision; FPS, Frames per second

Figure 9: Target detection results before and after image preprocessing

In order to verify whether the pre-processing of low-light pig images can improve the detection efficiency, this paper takes the presence of poorly lit images for image enhancement, and the images before and after enhancement are detected separately. Fig. 9 shows the detection results before and after image preprocessing, and from Fig. 9(b) and Fig. 9(c), it can be seen that the average confidence level of the left and right ears of the pig and the root part of the ear is improved by about 0.1%.

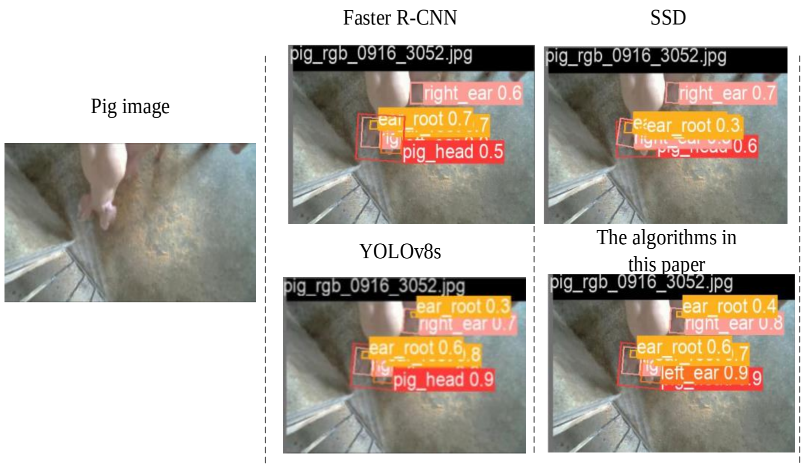


Figure 10: Different algorithms for occlusion case cleaning target detection results

In order to validate the improved YOLOv8 algorithm with the introduction of the attention mechanism and multi-scale BiFPN for the improvement of multi-scale detection accuracy of pig head, pig ear and ear root area, Faster R-CNN, SSD, YOLOv8 and the improved algorithm were compared in the dataset with the presence of occlusion, and the results of the detection are shown in Fig. 10above, in the case of the target without occlusion, the SSD algorithm has the lowest detection accuracy relative to the other three algorithms. Among them, YOLOv8s and the improved YOLOv8_OBB algorithm have

ISSN 2616-5910 Vol. 5, Issue 1: 102-112, DOI: 10.25236/AJALS.2024.050114

higher detection accuracies than Faster R-CNN and SSD for both the head and the root of the ear of the pig, in the case of occlusion, Faster R-CNN and SSD can only detect the relatively large scale ear, and for the small scale root of the ear the YOLOv8 algorithm can detect, but the accuracy is not enough. The improved algorithm can not only detect the target but also further improve the accuracy of occlusion detection, which better solves the situation of missed detection. Thus, it is proved that the algorithm in this paper can improve the detection accuracy and robustness of multi-scale pig head, ear and ear root parts.

4. Conclusion

This paper completes the ear root of pigs for localisation, firstly, by collecting pig data in a Henan livestock company to construct a more comprehensive ear root part data set, meanwhile, the collected data set is divided into training set, validation set and test set; adopting Retinex algorithm to process part of the low-light pig images in order to improve the recognition ability of the detection of the ear root part; and then, through the improvement of the YOLOv8- OBB rotating target frame detection, introducing the Shuffle Attention mechanism to enhance the saliency of pig ear parts and improve the detection accuracy of ear root targets, and fusing BiFPN module to solve the target recognition in the multi-scale cases of pig head, pig ear, and ear root parts in order to improve the detection accuracy. The experimental results show that compared with the YOLOv8 algorithm, the improved YOLOv8-OBB network model has better adaptability in low-light conditions and multi-scale target detection, with an average accuracy of 96.7% for pig ear root site localisation and an average detection speed of 30 FPS, which meets the requirements of pig ear root site localisation, and provides a good foundation for subsequent intelligent infrared ear root site temperature measurement.

References

[1] SOERENSEN D D, CLAUSEN S, MERCER J B, et al. Determining the emissivity of pig skin for accurate infrared thermography[J]. Computers and Electronics in Agriculture, 2014, 109: 52-58.

[2] SYKES D J, COUVILLION J S, CROMIAK A, et al. The use of digital infrared thermal imaging to detect estrus in gilts[J]. Theriogenology, 2012, 78(1):147-152.

[3] TZANIDAKIS C, SIMITZIS P, ARVANITIS K, et al. An overview of the current trends in precision pig farming technologies[J]. Livestock Science, 2021, 249:104530.

[4] Jing Huang, Jian Zhang. A study on an individual pig target detection method based on improved SSD network[J]. Software Engineering, 2022, 25(8): 25-29.

[5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. New York:ACM, 2014:580-587.

[6] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[7] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). December 07-13, 2015, Santiago, Chile. IEEE, 2016: 1440-1448.

[8] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN:towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[9] YANG S, PEI Z, ZHOU F, et al. Rotated faster R-CNN for oriented object detection in aerial image[C]// ZOU J X, Proceedings of the 2020 3rd International Conference on Robot Systems and Applications. Chengdu:Association for Computing Machinery, 2020:35-39. DOI: 10. 1145/3402597. 3402605.

[10] XU YC, FUMT, WANG Q M , et al. Gliding vertex on the horizontal bounding box for muli-oriented object detection[J]. IEEE. Transacions on Pattem Analysis andMachine Intelligence, 2021, 43(4): 1452-1459. DOI; 10. 1109/TPAMI. 2020. 2974745

[11] YANG X, YAN JC, FENG Z M, et al. R3Det ; refined single-stage detector with feature refinement for rotating object[J]. Proceedings of the AAAl Conference on Artificial Intelligence, 2021, 35(4); 3163-3171. DOI:10. 1609/aai. v35i4. 16426.

[12] WANG J, CHEN L Q, HUANG L L, et al. Lane recognition method in weak light condition based on retinex [J]. Computer & Digital Engineering, 2019, 47(2): 451-456. (in Chinese with English abstract)

[13] Hulin Li, Hanbing Wei, Zheng Liu et al. Slim-neck by GSConv: A better design paradigm of

ISSN 2616-5910 Vol. 5, Issue 1: 102-112, DOI: 10.25236/AJALS.2024.050114

detector architectures for autonomous vehicles [J]. Computer Vision and Pattern Recognition (CVPR), 2022.

[14] D. Ouyang et al., "Efficient Multi-Scale Attention Module with Cross-Spatial Learning, " ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5.

[15] Xie Qiuju, Wu Mengru, Bao Jun et al. Individual pig face recognition by fusing attention mechanisms[J]. Journal of Agricultural Engineering, 2022, 38(07):180- 188.

[16] Dongsheng Ruan, Daiyin Wang, Yuan Zheng, Nenggan Zheng, Min Zheng. Gaussian Context Transformer [J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[17] Zhao L, Liu S-P. Global and local image feature adaptive fusion of global and local image features [J]. Control and Decision, 2023, 38(04):935-943.

[18] Lu Jingyu, Zhang Haiyan, Wang Wenxin et al. Global sampling space attention mechanism and its application in image classification and small target detection and recognition[J]. Laser and Optoelectronics Advances in Laser and Optoelectronics, 2023(3):1-23.

[19] Cheng Changxin, Qiao Qingyuan, Luo Xiaoling et al. Target detection in UAV aerial images based on improved YOLOv8 Target Detection Algorithm for UAV Aerial Images[J]. Radio Engineering, 2024(3): 1-10.

[20] Ma Zhigang, Nan Xinyuan, Gao Bingpeng, et al. A pedestrian detection algorithm based on Pedestrian Detection Algorithm Based on Mobilenetv3 [J]. Modern Electronic Technology, 2023, 46(16): 149-154.