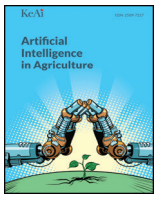
Artificial Intelligence in Agriculture 12 (2024) 19–28

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



A novel approach based on a modified mask R-CNN for the weight prediction of live pigs

Chuanqi Xie [a], Yuji Cang [b], Xizhong Lou [b], Hua Xiao [c], Xing Xu [a], Xiangjun Li [b], Weidong Zhou [a,*]

[a] State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-Products, Institute of Animal Husbandry and Veterinary Science, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China
[b] College of Information Engineering, China Jiliang University, Hangzhou 310018, China
[c] College of Hydraulic Engineering, Zhejiang Tongji Vocational College of Science and Technology, Hangzhou 311231, China

ARTICLE INFO

Article history:
Received 13 November 2023
Received in revised form 1 March 2024
Accepted 3 March 2024
Available online 04 March 2024

Keywords:
Deep learning
Modified mask R-CNN
Image processing
Pig weight
Prediction

ABSTRACT

Since determining the weight of pigs during large-scale breeding and production is challenging, using non-contact estimation methods is vital. This study proposed a novel pig weight prediction method based on a modified mask region-convolutional neural network (mask R-CNN). The modified approach used ResNeSt as the backbone feature extraction network to enhance the image feature extraction ability. The feature pyramid network (FPN) was added to the backbone feature extraction network for multi-scale feature fusion. The channel attention mechanism (CAM) and spatial attention mechanism (SAM) were introduced in the region proposal network (RPN) for the adaptive integration of local features and their global dependencies to capture global information, ultimately improving image segmentation accuracy. The modified network obtained a precision rate (P), recall rate (R), and mean average precision (MAP) of 90.33%, 89.85%, and 95.21%, respectively, effectively segmenting the pig regions in the images. Five image features, namely the back area, body length, body width, average depth, and eccentricity, were investigated. The pig depth images were used to build five regression algorithms (ordinary least squares (OLS), AdaBoost, CatBoost, XGBoost, and random forest (RF)) for weight value prediction. AdaBoost achieved the best prediction result with a coefficient of determination (R^2) of 0.987, a mean absolute error (MAE) of 2.96 kg, a mean square error (MSE) of 12.87 kg^2, and a mean absolute percentage error (MAPE) of 8.45%. The results demonstrated that the machine learning models effectively predicted the weight values of the pigs, providing technical support for intelligent pig farm management.

1. Introduction

Body weight is a significant index for monitoring pig growth during pig breeding (Liu et al., 2023). Rapidly and accurately estimating pig weight can help producers timeously adjust breeding strategies to reduce labor and feed costs (He et al., 2021a). Pig weight is traditionally measured using direct and indirect contact techniques (Bhoj et al., 2022). Although the direct method is most commonly employed to record values using a weighing scale, it is time-consuming, inefficient, and labor-intensive (Wang et al., 2021), while pigs are prone to severe stress reactions during the measurement (He et al., 2021a). The indirect method is used to measure the body size characteristics of pigs, such as body length, hip width, hip height, and heart girth, and calculates the relationships between the characteristics and body weight (Wang et al., 2021). Panda et al. (2021) demonstrated that body length, heart girth, paunch girth, height at wither, height at back, rump width, thigh circumference, neck circumference, and body depth are highly correlated with body weight (correlation coefficients (r) = 0.8–0.97). However, this method presents the same disadvantages as the direct method and is impractical for large-scale pig weight detection. Consequently, non-contact pig weight determination is more suitable for large-scale pig breeding.

In recent years, studies have explored non-contact pig weight estimation using machine vision techniques (1) to extract weight-related features from 2D pig images and establish correlation models between image features and body weight (Kashiha et al., 2014; Suwannakhun and Daungmala, 2018) and (2) to extract characteristics from 3D pig images, such as the back area and body size, and establish prediction models for the characteristics and body weight (Kongsro, 2014; Shi et al., 2016). Kashiha et al. (2014) collected top-view pig body area images. The heads and bodies of the pigs were fitted using the ellipse fitting method to calculate the pig body area. The transfer function model for the relationship between the body area and body weight obtained an accuracy of 96.2% at the individual level (standard error = 1.23 kg). To prevent edge detection errors caused by the dark contrast

* Corresponding author.
E-mail address: zhouwd@zaas.ac.cn (W. Zhou).

https://doi.org/10.1016/j.aiia.2024.03.001

Table 1
The weight statistics of the pigs.

Type	Number	Weight range/kg	Average weight/kg	Standard deviation
Nursery pigs	47	3.9–17.5	9.15	4.28
Finishing pigs	85	18.3–104.0	54.23	22.76
All	132	3.9–104.0	38.18	28.43

between the background and the pig image, Wongsriworaphon et al. (2015) extracted two features from the images: the average distance from the pig's center of mass to its boundary point and perimeter. The autoregressive model, locally linear embedding, and vector-quantized temporal associative memory (VQTAM) were used to improve the model, yielding an average error of <3%. He et al. (2021b) predicted pig weights using 3D images and a regression network based on BotNet. The single 3 × 3 convolution of the fourth block in ResNet was replaced by the dual branch of 3 × 3 convolution and multi-head self-attention, obtaining a mean absolute error (MAE) of 6.366 kg during testing. Another study used three different algorithms (least absolute shrinkage and selection operator (LASSO), random forest (RF), and long short-term memory (LSTM)) based on feeding behavior data to predict the body weight of growing pigs (He et al., 2021a). The individual-informed predictive scenario obtained the best result with a correlation of 0.87 and an accuracy of 0.89. This study established a promising link between feeding behavior dynamics and pig body growth to predict the body weights of group-housed pigs. Additionally, research was conducted on depth images for pig weight prediction. Pezzuolo et al. (2018) collected top- and side-view pig depth images, reconstructing the 3D pig structure to obtain the heart girth, body length, and body height values. Two weight prediction models (linear and non-linear) were constructed, obtaining the coefficients of determination (R^2) of >0.95. Jun et al. (2018) investigated two new features (curvature and deviation) for pig weight prediction, obtaining an average estimated error of 3.15 kg and an R^2 of 0.79. Compared with 2D images, 3D images provided more phenotypic information about the pigs, such as depth information. The parameter was highly correlated with pig weight, ultimately improving the prediction accuracy.

Studies have also employed the convolutional neural network (CNN) for pig weight estimation. Cang et al. (2019) added a regression branch based on the faster region-convolutional neural network (R-CNN) and integrated the pig detection and live weights regression network into an end-to-end network, obtaining a pig weight prediction MAE of 0.644 kg and a relative error (RE) of 0.374%. Zhang et al. (2021) proposed a multi-output regression CNN to estimate pig weight and body size. DenseNet 201, ResNet152 V2, Xception, and MobileNet V2 were modified into multi-output regression neural networks. The improved Xception was identified as the optimal model, yielding an R^2 ranging from 0.9879 to 0.9973.

Compared with traditional image processing, the deep learning image segmentation network is more accurate and efficient for complicated image processing (Qin et al., 2022; Bi et al., 2023). To segment the field of group-raised pigs, Hu et al. (2023) introduced channel and spatial attention blocks into the top-performing UNet and LinkNet decoders. Therefore, the basic model was built using ResNext50 as the encoder and Unet as the decoder, while adding two attention blocks simultaneously achieved 98.30% and 96.71% on the F1 and intersection over union (IoU) metrics, respectively. Therefore, our study uses a modified network based on the mask R-CNN for image processing, in which ResNeSt is employed as the feature extraction network to achieve multi-scale feature fusion by adding a feature pyramid network (FPN) to the backbone feature extraction network. Furthermore, the channel attention mechanism (CAM) and spatial attention mechanism (SAM) are introduced into the region proposal network (RPN) for the adaptive integration of local features and global dependencies to capture global information. This improves pig region extraction accuracy in complex pig house environments and provides theoretical support for precise pig weight estimation models. The specific objectives include (1) depth image segmentation using a modified mask R-CNN, (2) essential feature selection for prediction, and (3) optimal model identification for pig weight prediction. Finally, this study can provide a novel approach for weight determination on large-scale pig farms.

2. Materials and methods

2.1. Live pig back image collection

The experiment was performed at the Zhejiang Qingzhu Agriculture and Animal Husbandry Co., Ltd., Hangzhou, China. The pig depth images were collected from August 4, 2022, to August 13, 2022. This experiment included 132 pigs (47 nursery pigs and 85 finishing pigs) aged 30 d to 140 d and weighing 3.9 kg to 104.0 kg. Table 1 presents the weight statistics of the pigs.

An image acquisition device was positioned in the pig-driving passage (Fig. 1). A depth camera (Microsoft Azure Kinect DK, Redmond,

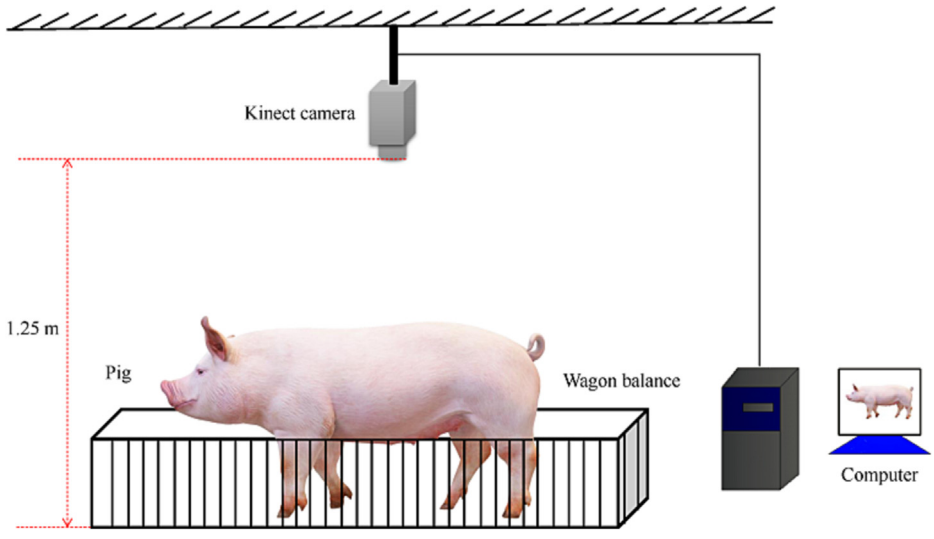


Fig. 1. A diagram of the pig depth image collection platform.

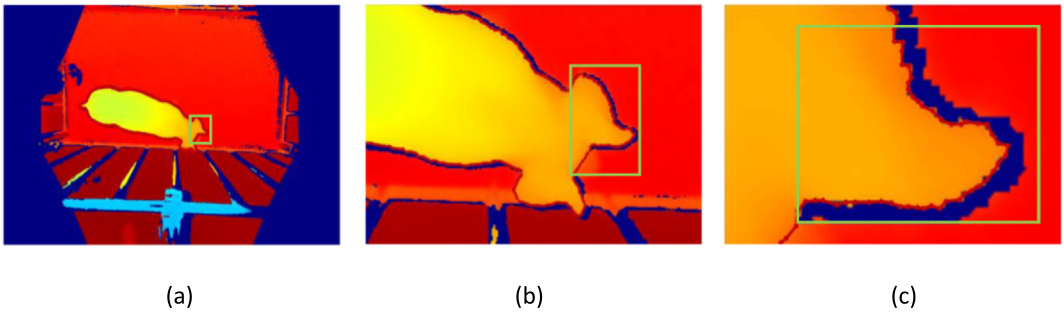


Fig. 2. The pig depth image annotation details: (a) the original depth image, (b) the enlarged depth image, and (c) further depth image enlargement.

WA, USA) was installed on the bracket and enclosed by a fence to capture images of pigs in a vertically downward direction. The camera was positioned 1.25 m above the ground of the wagon balance and captured video at a frame rate of 30 fps. The image collection process included the following steps:

- The pig driving board was used to drive one pig at a time into the wagon balance.
- The weight of the pig was recorded when it was calm.
- A video was taken of the back of the pig for 2 min.

The actual weight was obtained via a positive peak load meter at an accuracy of 0.1 kg.

Incomplete images caused by the movement of the pigs were removed, and the remaining pictures were enhanced via rotation, translation, and noise addition, ultimately obtaining 2684 images (from 132 pigs). Before enhancement, the LabelMe software was used to manually annotate the live pig depth images. To achieve ideal segmentation results, the depth images of the live pigs were enlarged to the limit and then manually annotated via the boundary contours (Fig. 2 (a-c)). All images were shuffled and randomly divided into the training, validation, and testing sets at around the ratio of 6:2:2, after which 1551, 572, and 561 images were obtained for each set, respectively.

2.2. Construction of the image segmentation network using the modified mask R-CNN

The mask R-CNN structure was used to examine the back image segmentation, providing technical support for acquiring the characteristic bodyweight parameters of live pigs. Segmentation aims to locate and divide each object using boundary boxes. The mask R-CNN used in this study represents a classic segmentation network (He et al., 2020). Its basic structure is shown in Fig. 3, which is divided into three parts: the backbone feature extraction network, the RPN, and the header detection network. First, the image features were extracted using the backbone feature extraction network, after which the RPN was employed to generate the image region of interest (ROI) (Jiao et al., 2017). A bilinear interpolation algorithm was used in ROIAlign to replace the quantization operation for image scaling. Therefore, the feature and original image pixels were entirely aligned, improving the segmentation accuracy. Finally, these ROIs were sent to the head detection network for target category prediction, boundary box regression, and mask generation.

This study used the ResNeSt backbone feature extraction network (Zhang et al., 2022), which retained the residual ResNet structure to avoid gradient disappearance during deep convolutional network training. Due to the limited receptive field of the ResNet network and the lack of cross-channel information interaction, it employed SENet (Hu et al., 2018) and SKNet (Li et al., 2019) when designing feature extraction units to select the more critical information for the objective. The Split-Attention module was introduced into the two attention mechanisms, allowing the network to extract semantic information across feature images and effectively improving the feature extraction ability (Zhang et al., 2022). Although the number of ResNeSt parameters did not increase significantly, it was more accurate than ResNet. This study added a modified FPN to the backbone feature extraction network to merge multi-scale characteristics and enrich image feature information. The CAM and SAM were added to the RPN to reduce background interference, such as railings, and obtain a more accurate target anchor frame.

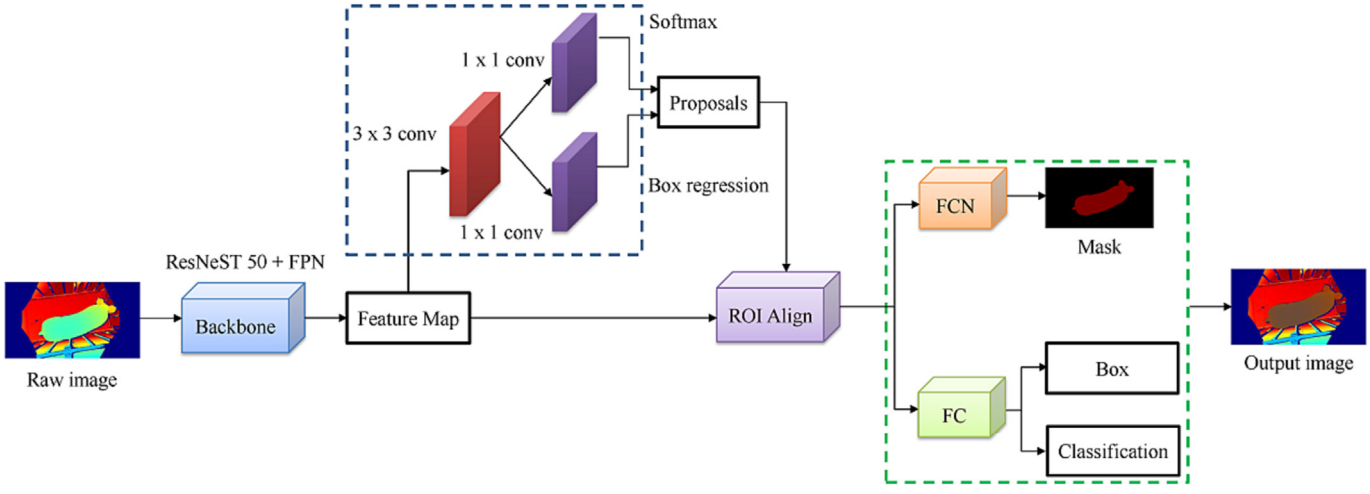


Fig. 3. A diagram of the mask R-CNN structure.

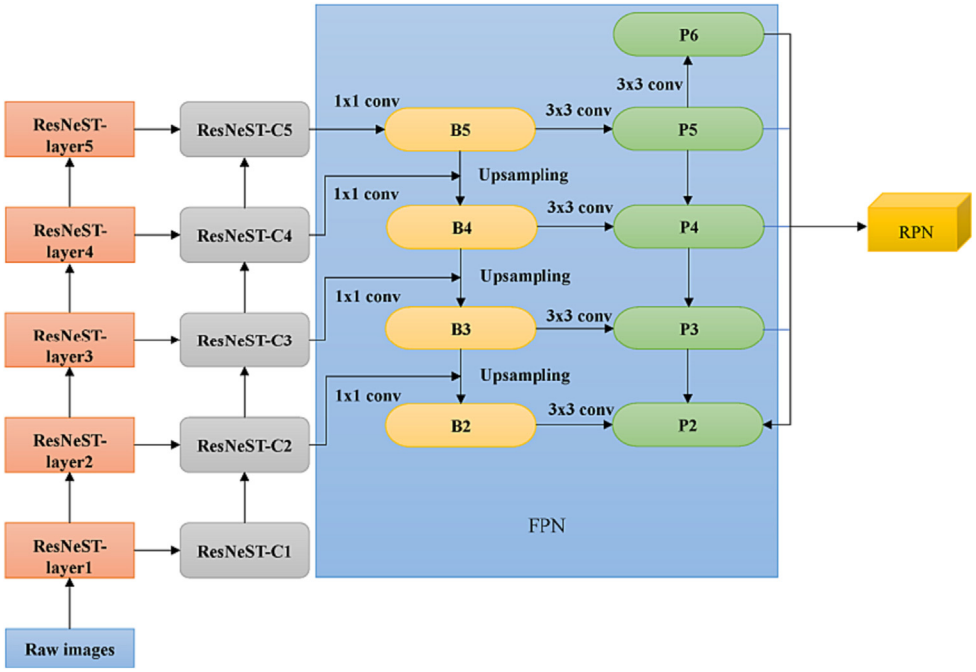


Fig. 4. A diagram of the FPN structure.

2.3. Multi-scale feature fusion FPN

ResNeSt was used as the backbone feature extraction network in the mask R-CNN to extract the pig image characteristics. The low-level features were extracted using the shallow network, increasing the geometric information and resolution. Although the semantic ability was weak, the small-target recognition ability was good. The high-level features were extracted using the deep network and presented contrary results. The receptive field gradually became more pronounced during the convolutional and pooling operations, resulting in minimal geometric information and low resolution. However, significant semantic information was evident, while the small-target extraction ability was poor. In this study, the sizes of the pigs differed significantly. Obtaining high-precision segmentation images was challenging when only single-scale feature images were entered into the RPN and ROIAlign layer for prediction. Therefore, a multi-scale fusion FPN structure was added to the ResNeSt-50 backbone feature extraction network. Fig. 4 illustrates the specific process. First, the feature images {ResNeSt-C1, ResNeSt-C2, ResNeSt-C3, ResNeSt-C4, and ResNeSt-C5} were generated via a convolution operation using the input images. The output images were subjected to a 1×1 convolution operation and fused with those obtained via up-sampling to generate the {B5, B4, B3, and B2} feature images. To eliminate the aliasing effect caused by the up-sampling process, these feature images were subjected to a 3×3 convolution operation to obtain the {P6, P5, P4, P3, and P2} feature images.

2.4. Enhanced RPN

The RPN model represents a region-generating network. Sliding windows are used to generate k anchor frames at each pixel in the feature images produced in the previous step. The anchor frames containing target objects are regressed to obtain accurate target anchor frames. The RPN consists of two branches: one for sorting tasks and the other for bounding box regression tasks. Assuming that the size of an image is $w \times h$, there are $w \times h \times k$ anchor frames for this image. However, since many anchor frames may not contain target objects, it is necessary to determine whether the generated anchor frame represents the background or the target. The function of the second branch involves anchor frame regression to obtain a more accurate boundary frame.

By adding CAM and SAM, the enhanced RPN reduces the number of boxes unrelated to the targets, improving the regression and classification results. Since the CAM can provide different weights according to

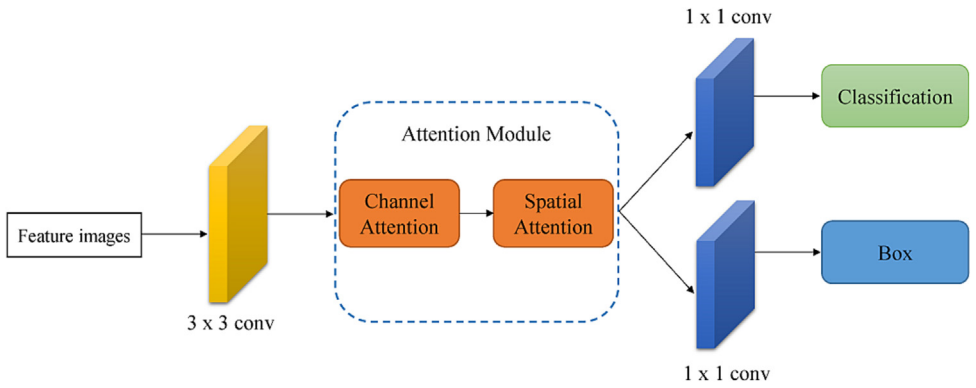


Fig. 5. A diagram of the attention mechanism structure added to the RPN.

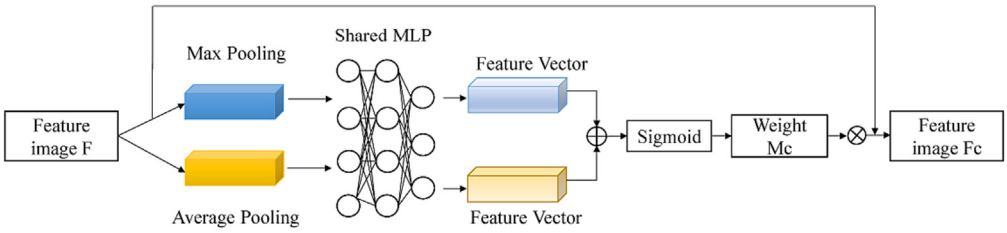


Fig. 6. The CAM structure.

the importance of each channel, it can produce higher weights for the channels containing pig information and lower weights for those containing background data. The regression branch outputs the spatial position of the target in the feature image. Contrarily, since the SAM mainly focuses on the valuable information in the image and can show the spatial position of the pigs, it yields better regression results. Since this study added both CAM and SAM to the RPN, the importance of different channels and varying positions in the same channel were considered. Fig. 5 shows a diagram of the attention mechanism structure added to the RPN.

Fig. 6 shows the CAM structure. The feature image produced by the convolutional layer represents the CAM input, while the number of channels is denoted by C. Two 1 × 1 x C feature images were obtained after average and max layer pooling, which were entered into a shared multilayer perceptron (MLP) to obtain the corresponding feature vectors. The two feature vectors were added, and the channel weight coefficient Mc was produced via a sigmoid activation function. Finally, Mc and input feature image F were multiplied to obtain feature image Fc.

Fig. 7 shows the SAM structure. Feature image Fc produced via CAM represented the SAM input. Two 1-channel feature images were obtained after average and max layer pooling, while a single-channel image was acquired after 7 × 7 convolution and dimensionality reduction. The weight coefficient Ms. was obtained via a sigmoid activation function, while the feature image Fc_s based on CAM and SAM was acquired by multiplying Ms. and feature image Fc.

2.5. Segmentation evaluation

This study used the precision rate (P), recall rate (R), and mean average precision (MAP) to evaluate the image segmentation performance via the following equations:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$MAP = \frac{1}{N}\sum_{i}^{N}\frac{P_i + R_i}{2} \tag{3}$$

where the true positive (TP) represents the number of positive samples predicted as positive, the false positive (FP) denotes the number of negative samples predicted as positive, and the false negative (FN) signifies the number of positive samples predicted as negative.

2.6. Image segmentation parameters

The Windows 10 64-bit operating system was employed for image segmentation, using a Tesla P40 (40 G video memory) graphics card and 24 G memory. Graphics processing unit (GPU) acceleration was employed to establish, train, and debug the segmentation models via PyCharm, Python, and Pytorch. The segmentation network parameters were initialized using the criteria listed in Table 2, while the pig contour segmentation model was trained via random gradient descent. The training, validation, and testing sets contained 1551, 572, and 561 images, respectively. Therefore, the segmentation model is capable of rapid convergence due to data enhancement. The model was trained with a total of 120 epochs, while the segmentation model convergence was evaluated by observing the loss value change.

2.7. Body feature extraction

The feature characteristics were extracted after identifying the pig region in the segmentation image using the mask R-CNN segmentation algorithm. The pigs in the images only displayed posture differences in areas such as the head and tail, which significantly impacted the projected area in the 2D plane. The protruding portions of the image were removed via a morphological opening operation.

The characteristic body weight features of the pigs were extracted from the images without the heads and tails and included the back area, body length, body width, and average depth. Since the acquired images displayed various pig postures, some straight and some twisted, eccentricity was introduced to evaluate the degree of body curvature. The characteristic features were calculated as follows:

2.7.1. Back area (A)

The A was defined by the number of pixels in the region, namely, the sum of pixels within the boundary, and was calculated as follows:

$$A = \sum_{x=1}^{N}\sum_{y=1}^{M} f(x,y) \tag{4}$$

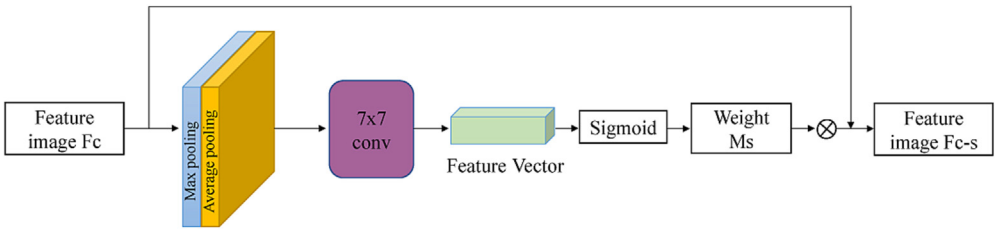


Fig. 7. The SAM structure.

Table 2
The initial training parameters of the mask R-CNN networks.

Parameters	Values
Batch size	5
Initial learning rate	0.001
Learning rate attenuation	0.1
Factor of momentum	0.9
Regularized weight attenuation coefficient	0.0001
Epoch	120

Since this was a binary image, 1 represented the pig region and 0 denoted the background. Therefore, A was f(x, y) = 1.

2.7.2. Body length (L) and body width (W)

The L and W were expressed by the length of the long and short sides of the smallest outer rectangle of the back outline of the pig, respectively.

2.7.3. Average depth (AD)

The AD represented the average pixel value in the fitting ellipse and was calculated as follows:

$$AD = \frac{\sum_{n=1}^{i} d_n}{i} \tag{5}$$

where i represents the total number of pixels in the ellipse and d_n denotes the value of each pixel in the ellipse.

2.7.4. Eccentricity (E)

The E reflected the pig posture curvature. An E value closer to 1 indicated a straighter pig posture, while a value further from 1 showed increased curvature. The E was calculated as follows:

$$E = \sqrt{1 - \left(\frac{b}{a}\right)^2} \tag{6}$$

where a denotes the major axis and b signifies the minor axis of the fitting ellipse.

2.8. Regression models

This study used ordinary least squares (OLS), AdaBoost, CatBoost, XGBoost, and RF regression algorithms for feature training. AdaBoost was adapted to increase the weights of the samples misclassified by the previous primary classifier and decrease those of the correctly classified samples, which could be used to train the next primary classifier. Furthermore, a new weak classifier was added to each iteration. The final robust classifier was not established until a predetermined, small enough error rate or a maximum number of iterations was reached (Freund and Schapire, 1997). The CatBoost parameters represented a set of multiple decision trees (Prokhorenkova et al., 2018). Every decision tree denoted a weak learner whose structure could be optimized by minimizing the square loss function. Specifically, CatBoost used a greedy algorithm to build each decision tree, employing the Gini coefficient or logarithmic likelihood as split criteria in decision tree construction. XGBoost was improved using the gradient lifting tree model (Chen and Guestrin, 2016). The improvements were all optimized in the loss function. First, standard terms were added to avoid overly complex models and reduce the risk of overfitting. Second, the second-order Taylor formula was used to expand the loss function, improving the speed and accuracy of model fitting. It also represented an additional model, while the original internal trainer had multiple regression tree models. The trainer parameters were optimized individually using the forward stagewise algorithm, and the training ceased when it reached the target error value. An RF is an integrated learning model with multiple decision trees (Breiman, 2001) that improves prediction accuracy and stability. Each decision tree is built according to random samples and features, which can avoid overfitting and ensure good robustness.

3. Results and discussion

3.1. Image segmentation model performance

As is shown in Fig. 8(a-b), the mask R-CNN-ResNeSt50-FPN-AM displayed the lowest loss value of 0.19 during training and 0.22 during validation compared with the other three networks when trained for about 23 epochs, after which it remained mostly stable, indicating that it reached the convergence state.

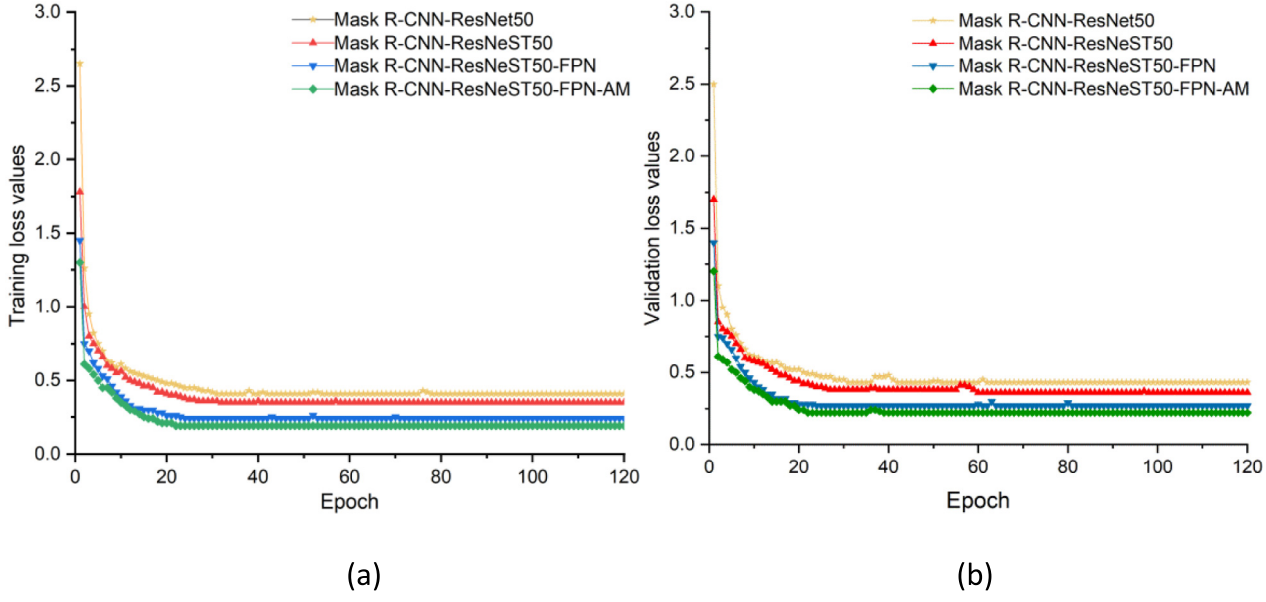


Fig. 8. The loss values during (a) training and (b) validation of the different networks.

Table 3
The performance of the different segmentation networks.

Segmentation models	P (%)	R (%)	MAP (%)
ResNet50	82.63	84.31	86.34
ResNeSt50	84.39	85.42	87.26
ResNeSt50-FPN	88.46	88.23	89.58
ResNeSt50-FPN-AM	90.33	89.85	95.21

The parameters listed in Table 2 were used to test the pig contour segmentation models constructed using different algorithms. The specific feature extraction networks included (1) ResNet50, (2) ResNeSt50, (3) ResNeSt50 and FPN (ResNeSt50-FPN), and (4) ResNeSt50-FPN with RPN added as an attention mechanism (ResNeSt50-FPN-AM).

Table 3 presents the performance of the four segmentation models. The ResNeSt50 evaluation indexes were superior to those of ResNet 50, indicating that the split-attention residual network improved segmentation performance. When the multi-scale FPN was added to the ResNeSt50 backbone feature extraction network, the P, R, and MAP values increased from 84.39%, 85.42%, and 87.26% to 88.46%, 88.23%, and 89.58%, respectively. Furthermore, although the segmentation model performance was enhanced by the feature fusion of different scales, an adhesion phenomenon was evident at the boundary when the pig was close to the railing. The P, R, and MAP of the modified ResNeSt50-FPN-AM segmentation model reached 90.33%, 89.85%, and 95.21%, respectively, after adding CAM and SAM to the RPN, with the MAP value 8.87% higher than that of ResNet 50. The segmentation results of ResNeSt 50-FPN-AM are shown in Fig. 9, presenting the annotated depth image, annotated segmentation image, predicted depth image, and predicted segmentation image.

3.2. Characteristic features

The head and tail areas were removed from the pig images via morphological calculation (Fig. 10(a-b). An s × s circular kernel was used to remove the protruding head and tail portions from the mask image of the live pig. A larger s value indicated the removal of more pixels from the protruding area. The pigs requiring segmentation displayed different shapes and sizes, making it difficult to choose an s value suitable for all cases. Therefore, different kernel sizes should be selected for pigs of varying sizes. Since the head and tail sizes were related to the overall size of the pig, the s value could be associated with the pixel values of the original pig contour. The initial s value was set to remove interference, after which it was adjusted according to the results. The findings indicated that using kernels for pig contour opening operations can successfully eliminate interference by selecting s for five intervals, which were set as follows:

Fig. 10. (a) the raw image and (b) the image after head and tail removal.

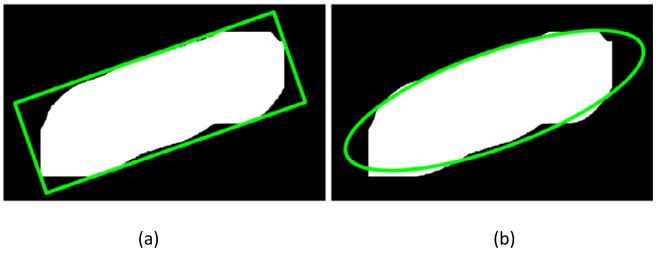


Fig. 11. (a) the smallest outer rectangle of the back outline of the pig and (b) the fitting ellipse.

The smallest outer rectangle of the back outline of the pig and the fitting ellipse were shown in Fig. 11(a-b). Finally, the weight feature parameters were obtained after extracting the pig image, as shown in Table 4.

$$\begin{cases} 50*50 & & area \le 100000 \\ 110*110 & 100000 < & area \le 200000 \\ 140*140 & 200000 < & area \le 300000 \\ 190*190 & 300000 < & area \le 400000 \\ 250*250 & & area > 400000 \end{cases} \quad (7)$$

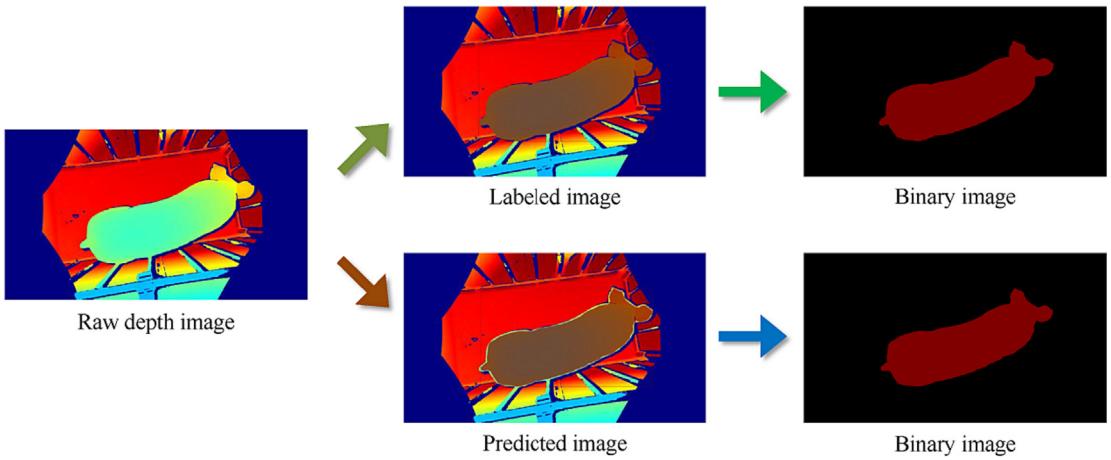


Fig. 9. The ResNeSt 50-FPN-AM segmentation results.

Table 4
The characteristic weight features of the live pigs.

No.	A/pixel	L/pixel	W/pixel	AD/mm	E
1	163,108	775	312	884	0.95
2	334,231	1067	504	795	0.97
3	431,748	1027	675	707	0.92
…	…	…	…	…	…
132	540,310	1372	751	689	0.97

Table 5
The weight statistics of the pigs in the training and testing sets.

Type	Number	Weight range/kg	Average weight/kg	Standard deviation
Training	105	3.9–103.0	37.43	27.95
Testing	27	5.1–104.0	41.10	30.62
All	132	3.9–104.0	38.18	28.43

Table 6
The prediction results of the five models in the testing set.

Models	R^2	MAE/kg	MSE/kg^2	MAPE (%)
OLS	0.979	3.57	19.39	12.44
AdaBoost	0.987	2.96	12.87	8.45
CatBoost	0.974	4.07	23.74	12.19
XGBoost	0.984	3.36	16.42	9.69
RF	0.967	4.19	31.92	12.20

3.3. Pig weight prediction results

Each feature value was normalized to improve model training efficiency. The calculation formula was as follows:

$$x_{normalization} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (8)$$

The normalized data were divided into 4:1 training and testing sets for live pig weight prediction (Duan et al., 2023), as shown in Table 5. Since the weight values of the nursery and finishing pigs differed significantly, all samples were arranged from the smallest to the largest weights, allowing the training set to cover all possible values. Then, one sample was selected randomly from every four and used in the testing set. The back area, body length, body width, average depth, and eccentricity values were used as regression algorithm input variables. The OLS, AdaBoost, CatBoost, XGBoost, and RF regression model prediction results were calculated, as shown in Table 6. AdaBoost achieved the best prediction results with an R^2 of 0.987, an MAE of 2.96 kg, a mean square error (MSE) of 12.87 kg^2, and a mean absolute percentage error (MAPE) of 8.45%. The predicted values of pig weights using AdaBoost in training and testing sets can be seen in Fig. 12 (a-b). The RF model exhibited the worst prediction results for the testing set, with an R^2 of 0.967, an MAE of 4.19 kg, an MSE of 31.92 kg^2, and an MAPE of 12.20%. The prediction results of OLS, CatBoost, and XGBoost were acceptable, with R^2 values of 0.979, 0.974, and 0.984, MAE values of 3.57, 4.07, and 3.36 kg, MSE values of 19.39, 23.74, and 16.42 kg^2, and MAPE values of 12.44%, 12.19%, and 9.69%, respectively. Therefore, AdaBoost was selected as the optimal live pig weight estimation model. AdaBoost displayed a high level of precision, possibly because

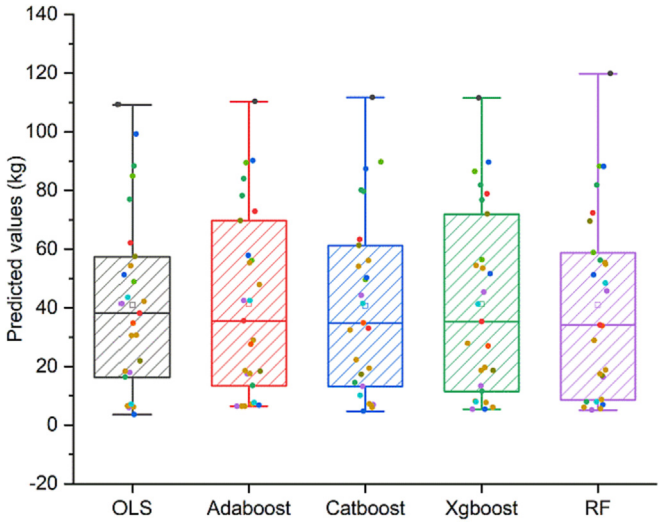


Fig. 13. The boxplot of the prediction results for the different models.

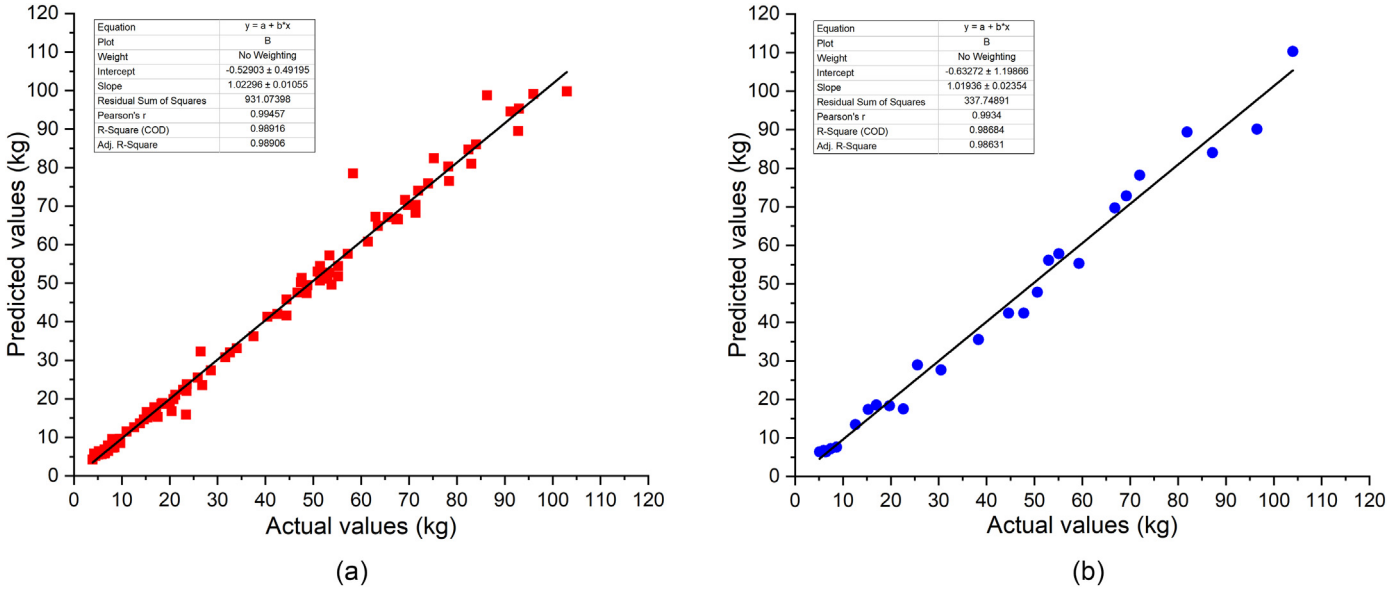


Fig. 12. The actual versus the predicted values of live pig weights using AdaBoost: (a) training and (b) testing.

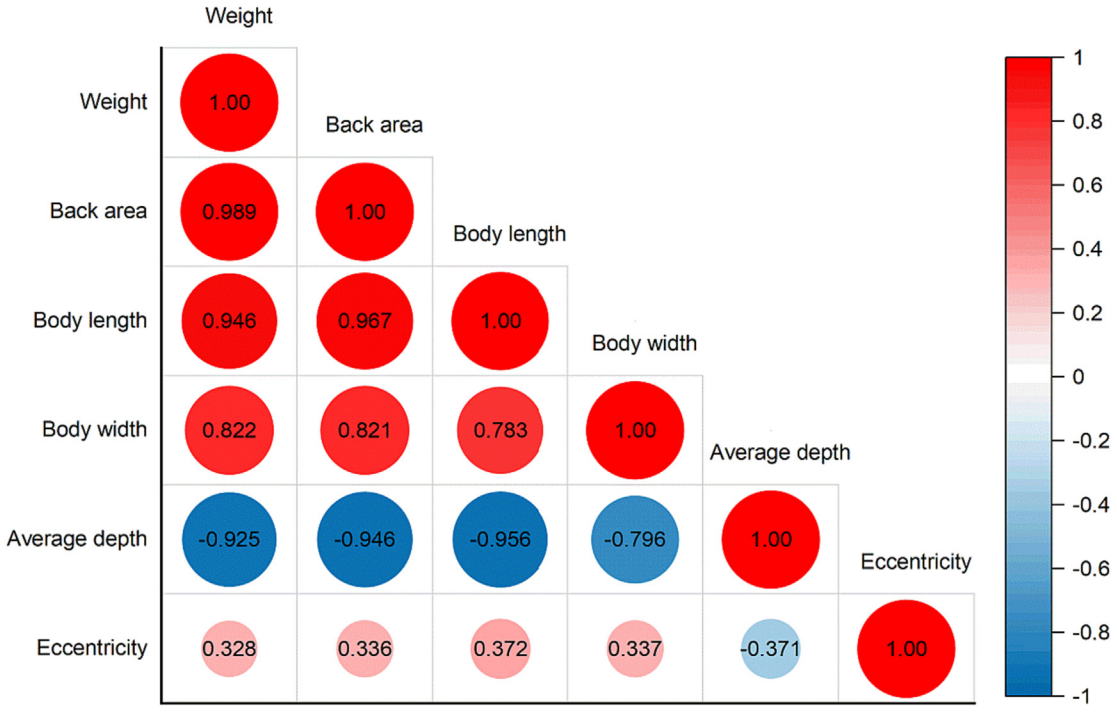


Fig. 14. The heat map of Pearson's correlation analysis.

the training error decreased exponentially while fully considering the weight of each classifier compared with the bagging and RF algorithms. A previous study (Nguyen et al., 2023) used support vector regression (SVR) as the base learner of the MLP, AdaBoost, and SVR algorithms. The model parameters were adjusted using a grid algorithm and cross-validation, yielding good results. However, our study compared the classical regression models, including the linear, boosting, and bagging algorithms, concluding that the Adaboost model was superior to other classic models, which would aid other researchers in selecting base learners.

Fig. 13 shows a boxplot of the prediction results for the different models. Each model obtained good prediction results with no outliers. S. Fig. 1 shows the different model learning curves using MAE, indicating similar MAE curve gradients in each model. Since the AdaBoost gradient was the largest, its MAE loss value was the smallest.

3.4. Pearson's correlation analysis

Fig. 14 shows the pairwise Pearson's correlation heat map between the manually measured body weight and each body size parameter obtained from the images. The back area exhibited the highest correlation with the body weight (0.989), followed by body length (0.946), average depth (−0.925), body width (0.822), and eccentricity (0.328). Among the body size parameters other than body weight, the highest association was evident between the body length and back area (0.967), while eccentricity and back area presented the lowest correlation (0.336). Therefore, back area might be the primary consideration in future body weight prediction studies.

4. Conclusions

This study uses machine learning based on depth images to investigate live pig weight prediction. Depth images of the backs of 132 live pigs are captured, while a modified mask R-CNN is constructed based on multi-scale fusion FPN and enhanced RPN for image segmentation. The P, R, and MAP of the network are 90.33%, 89.85%, and 95.21%, respectively, indicating excellent segmentation performance. Five characteristic body weight features (back area, body length, body width, average depth, and eccentricity) are selected as the input for the regression algorithms (OLS, AdaBoost, CatBoost, XGBoost, and RF). The AdaBoost model yields the best prediction result with an R^2 of 0.987, an MAE of 2.96 kg, an MSE of 12.87 kg^2, and an MAPE of 8.45%. Furthermore, back area plays the most significant role during weight prediction, which is helpful for accelerating animal breeding.

Using traditional image processing methods to extract the pig regions is challenging due to the complex environment of pig pens and various obstructions, such as railings. This study shows that ResNeSt50-FPN-AM image segmentation is more stable and efficient. Furthermore, the average depth information of the pigs can be obtained using deep images, compensating for the lack of 3D information in 2D images. The findings provide a reference for live pig weight detection in large-scale pig farms. However, to increase model robustness, future studies should consider a larger number of pigs and images from multiple scenarios.

CRediT authorship contribution statement

Chuanqi Xie: Conceptualization, Writing – original draft, Writing – review & editing, Funding acquisition, Project administration, Supervision. **Yuji Cang:** Investigation, Formal analysis, Methodology. **Xizhong Lou:** Conceptualization. **Hua Xiao:** Conceptualization. **Xing Xu:** Conceptualization. **Xiangjun Li:** Conceptualization. **Weidong Zhou:** Conceptualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Key R&D Program of Zhejiang (2022C02050) and Zhejiang Provincial Natural Science Foundation of China (ZCLTGN24C1301).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aiia.2024.03.001.

References

Bhoj, S., Tarafdar, A., Chauhan, A., Singh, M., Gaur, G.K., 2022. Image processing strategies for pig liveweight measurement: updates and challenges. Comput. Electron. Agr. 193, 106693. https://doi.org/10.1016/j.compag.2022.106693.

Bi, Y., Campos, L.M., Wang, J., Yu, H.P., Hanigan, M.D., Morota, G., 2023. Depth video data-enabled predictions of longitudinal dairy cow body weight using thresholding and Mask R-CNN algorithms. Smart Agr. Tech. 6, 100352. https://doi.org/10.1016/j.atech.2023.100352.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Cang, Y., He, H.X., Qiao, Y.L., 2019. An intelligent pig weights estimate method based on deep learning in sow stall environments. IEEE Access 7, 164867–164875. https://doi.org/10.1109/ACCESS.2019.2953099.

Chen, T.Q., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA, USA. pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Duan, E.Z., Hao, H.Y., Zhao, S.D., Wang, H.Y., Bai, Z.C., 2023. Estimating body weight in captive rabbits based on improved mask RCNN. Agriculture 13 (4), 791. https://doi.org/10.3390/agriculture13040791.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139. https://doi.org/10.1006/jcss.1997.1504.

He, H.X., Qiao, Y.L., Li, X.M., Chen, C.Y., Zhang, X.F., 2021b. Automatic weight measurement of pigs based on 3D images and regression network. Comput. Electron. Agr. 187, 106299. https://doi.org/10.1016/j.compag.2021.106299.

He, K.M., Gkioxari, G., Dollár, P., Girshick, R., 2020. Mask R-CNN. IEEE T. Pattern Anal. 42 (2), 386–397. https://doi.org/10.1109/TPAMI.2018.2844175.

He, Y.Q., Tiezzi, F., Howard, J., Maltecca, C., 2021a. Predicting body weight in growing pigs from feeding behavior data using machine learning algorithms. Comput. Electron. Agric. 184, 106085. https://doi.org/10.1016/j.compag.2021.106085.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. pp. 7132–7141. https://doi.org/10.1109/CVPR.2018.00745.

Hu, Z.W., Yang, H., Lou, T.T., Yan, H.W., 2023. Concurrent channel and spatial attention in fully convolutional network for individual pig image segmentation. Int. J. Agr. Biol. Eng. 16 (1), 232–242. https://doi.org/10.25165/j.ijabe.20231601.6528.

Jiao, L.C., Liang, M.M., Chen, H., Yang, S.Y., Liu, H.Y., Cao, X.H., 2017. Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification. IEEE T. Geosci. Remote 55 (10), 5585–5599. https://doi.org/10.1109/TGRS.2017.2710079.

Jun, K., Kim, S.J., Ji, H.W., 2018. Estimating pig weights from images without constraint on posture and illumination. Comput. Electron. Agr. 153, 169–176. https://doi.org/10.1016/j.compag.2018.08.006.

Kashiha, M., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Ödberg, F.O., Berckmans, D., 2014. Automatic weight estimation of individual pigs using image analysis. Comput. Electron. Agr. 107, 38–44. https://doi.org/10.1016/j.compag.2014.06.003.

Kongsro, J., 2014. Estimation of pig weight using a Microsoft Kinect prototype imaging system. Comput. Electron. Agr. 109, 32–35. https://doi.org/10.1016/j.compag.2014.08.008.

Li, X., Wang, W.H., Hu, X.L., Yang, J., 2019. Selective kernel networks. 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. pp. 510–519. https://doi.org/10.1109/CVPR.2019.00060.

Liu, Z.X., Zhang, X.F., Ji, B.Y., Banhazi, T., Li, C.C., Zhao, S.H., 2023. Analysis of diurnal variations in body weight of wean-to-finish pigs. Biosyst. Eng. 228, 80–87. https://doi.org/10.1016/j.biosystemseng.2023.02.010.

Nguyen, A.H., Holt, J.P., Knauer, M.T., Abner, V.A., Lobaton, E.J., Young, S.N., 2023. Towards rapid weight assessment of finishing pigs using a handheld, mobile RGB-D camera. Biosyst. Eng. 226, 155–168. https://doi.org/10.1016/j.biosystemseng.2023.01.005.

Panda, S., Gaur, G.K., Chauhan, A., Kar, J., Mehrotra, A., 2021. Accurate assessment of body weights using morphometric measurements in Landlly pigs. Trop. Anim. Health Pro. 53 (3), 362. https://doi.org/10.1007/s11250-021-02803-2.

Pezzuolo, A., Guarino, M., Sartori, L., González, L.A., Marinello, F., 2018. On-barn pig weight estimation based on body measurements by a Kinect v1 depth camera. Comput. Electron. Agr. 148, 29–36. https://doi.org/10.1016/j.compag.2018.03.003.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. 32nd Conference on Neural Information Processing Systems (NIPS). Montreal, Canada, p. 31.

Qin, Q., Dai, D.L., Zhang, C.Y., Zhao, C., Liu, Z.C., Xu, X.L., Lan, M.X., Wang, Z.X., Zhang, Y.J., Su, R., Wang, R.J., Wang, Z.Y., Zhao, Y.H., Li, J.Q., Liu, Z.H., 2022. Identification of body size characteristic points based on the mask R-CNN and correlation with body weight in Ujumqin sheep. Front. Vet. Sci. 9, 995724. https://doi.org/10.3389/fvets.2022.995724.

Shi, C., Teng, G.H., Li, Z., 2016. An approach of pig weight estimation using binocular stereo system based on LabVIEW. Comput. Electron. Agr. 129, 37–43. https://doi.org/10.1016/j.compag.2016.08.012.

Suwannakhun, S., Daungmala, P., 2018. Estimating pig weight with digital image processing using deep learning. In Proceedings of the 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems, pp. 320–326. https://doi.org/10.1109/SITIS.2018.00056.

Wang, Z.Y., Shadpour, S., Chan, E., Rotondo, V., Wood, K.M., Tulpan, D., 2021. ASAS-NANP symposium: applications of machine learning for livestock body weight prediction from digital images. J. Anim. Sci. 99 (2), 1–15. https://doi.org/10.1093/jas/skab022.

Wongsriworaphon, A., Arnonkijpanich, B., Pathumnakul, S., 2015. An approach based on digital image analysis to estimate the live weights of pigs in farm environments. Comput. Electron. Agr. 115, 26–33. https://doi.org/10.1016/j.compag.2015.05.004.

Zhang, H., Wu, C.R., Zhang, Z.Y., Zhu, Y., Lin, H.B., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., Li, M., Smola, A., 2022. ResNeSt: Split-attention networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, pp. 2735–2745. https://doi.org/10.1109/CVPRW56347.2022.00309.

Zhang, J.L., Zhuang, Y.R., Ji, H.Y., Teng, G.H., 2021. Pig weight and body size estimation using a multiple output regression convolutional neural network: a fast and fully automatic method. Sensors 21 (9), 3218. https://doi.org/10.3390/s21093218.