IEICE TRANS. INF. & SYST., VOL.E106–D, NO.5 MAY 2023

PAPER Special Section on Deep Learning Technologies: Architecture, Optimization, Techniques, and Applications

Detection Method of Fat Content in Pig B-Ultrasound Based on Deep Learning

Wenxin DONG[†], **Jianxun ZHANG**[†a)], **Shuqiu TAN**[†], ***and*** **Xinyue ZHANG**[††], ***Nonmembers***

SUMMARY In the pork fat content detection task, traditional physical or chemical methods are strongly destructive, have substantial technical requirements and cannot achieve nondestructive detection without slaughtering. To solve these problems, we propose a novel, convenient and economical method for detecting the fat content of pig B-ultrasound images based on hybrid attention and multiscale fusion learning, which extracts and fuses shallow detail information and deep semantic information at multiple scales. First, a deep learning network is constructed to learn the salient features of fat images through a hybrid attention mechanism. Then, the information describing pork fat is extracted at multiple scales, and the detailed information expressed in the shallow layer and the semantic information expressed in the deep layer are fused later. Finally, a deep convolution network is used to predict the fat content compared with the real label. The experimental results show that the determination coefficient is greater than 0.95 on the 130 groups of pork B-ultrasound image data sets, which is 2.90, 6.10 and 5.13 percentage points higher than that of VGGNet, ResNet and DenseNet, respectively. It indicats that the model could effectively identify the B-ultrasound image of pigs and predict the fat content with high accuracy.

key words: *B-ultrasound image, convolutional neural network, deep learning, fat content detection, nondestructive testing*

1. Introduction

With the notable improvement of material life, people's demand for pork quality is growing rapidly. Because the content of pork fat and its distribution are uniform or not determine the quality of meat, thus affecting the quality of pork varieties, pork fat content detection is a major issue of great significance for scientific research on pig breeding.

In the traditional breeding process, in addition to experienced animal husbandry staff who make observations based on appearance, technical personnel are also required to carry out layer-by-layer detection after slaughtering pigs [1]. This process requires a closed professional environment. The slaughterhouse and storage need to be closely linked with a laboratory, and no error is allowed. Once the sample is contaminated during collection, it will be invalid. This traditional method of artificial or physical and chemical detection after slaughter is strongly destructive, has substantial technical requirements and cannot evaluate the meat traits of pigs in vivo [2].

To solve this problem, this paper proposes a fat content prediction method based on deep learning for pig B-ultrasound images. B-ultrasound is an economical and practical technology. [3] analyzes the comparative effect of B-ultrasound and MRI, and points out that B-ultrasound is not only easier to obtain and use, but also has the same effect as MRI, and has the advantages of reproducibility. Our model skillfully combines B-ultrasound images with convolutional neural networks to complete the task of pork fat content detection, adding a novel, convenient and economical nondestructive detection method. It makes efficient use of frontier computer technology, will greatly assist in solving practical market application problems, simplify the detection process, reduce labor costs, save detection time, reduce detection costs, improve detection accuracy and improve breeding effects. It will have great significance in industrial production and human life.

In the second part, we introduce the development of pork fat detection technology, from a simple linear model to a complex machine learning model and then to a deep learning model. The third part proposes the overall framework of our convolutional neural network, HAFFNet. The fourth part is the experiment carried out on pig B-ultrasound images, and the last part is the summary.

2. Related Work

2.1 Detection Technology for Fat in Live Pork

Compared with traditional detection technology, live detection technology is more scientific. At present, a widely used live detection method is ultrasonic detection [4], [5]. There is a significant difference in resistance between the fat and lean meat of pigs. Thus, the reflection wave will be extremely different. Depending on ultrasonication, we can obtain effective parameters of fat content, backfat thickness, eye muscle area and other related traits of live pigs. However, this technology has problems such as obvious noise in images and distortion in transmission displays, so accurate fat content cannot be obtained only by ultrasonic technology.

To achieve more effective nondestructive testing, many scholars have conducted relevant research. A prediction model based on statistical analysis was proposed by [6]. However, this method only selects several parameters, such as the area of porcine eye muscle, backfat thickness and

Manuscript received October 11, 2021.
Manuscript revised January 14, 2022.
Manuscript publicized February 7, 2022.

[†]The authors are with the College of Computer Science and Engineering, Chongqing University of Technology, Chongqing, 400054 China.
[††]The author is with Sydney Smart Technology College, Northeastern University, Hebei, 066004 China.
a) E-mail: zjx@cqut.edu.cn

DOI: 10.1587/transinf.2022DLP0022

depth of eye muscle, to construct a linear model with fat content, which cannot effectively use the complete features of B-ultrasound images. The determination coefficient (R^2) is still at a low level, cannot achieve high accuracy, and has great limitations for practical applications. In addition to the prediction of pork fat content based on certain parameters, some scholars also analyzed pork fat content based on the overall traits of pigs. For example, there are some studies to detect pork fat content based on shape [7], but these researches rely on a variety of measurement parameters, such as body length, body height, chest depth, abdominal length, hip width, and waist width. The sampling process is complex, and there are errors in the determination, so it may not be the most ideal method.

2.2 Application of Machine Learning

A pig fat content detection technology based on a support vector machine (SVM) was proposed by [8]. SVM is one of the better supervised learning models and can effectively deal with high-dimensional data sets. This technique can be used to classify the fat categories in B-ultrasound images of pigs, but an important drawback of the model is that it cannot directly provide probability estimation. Otherwise, the feature extraction rules of the support vector machine are set manually, which cannot be applied to the feature extraction of large data. In practical applications, an excessively small amount of data is not representative and cannot make the model learn well. If a large amount of data is only expected to manually extract feature variables, the workload is too large and does not have practical significance.

2.3 Application of Deep Learning

To date, image prediction methods in computer vision can be roughly divided into two types: the method of manually extracting features based on traditional machine learning and the method of convolutional neural networks (CNN) based on deep learning [9], [10]. Traditional methods usually rely on manual extraction of features such as scale invariant feature transform (SIFT), histogram of oriented gradient (HOG), and then a traditional neural network or support vector machine classifier [11]–[14] to complete the classification; however, such algorithms, as previously analyzed, have strong dependence on manual extraction features, and it is often difficult to handle deeper and more abundant information from the image. Thus, the main difficulty is often low recognition accuracy.

A convolutional neural network is a kind of feed-forward neural network with convolution calculations and deep structures and is one of the representative algorithms of deep learning [15], [16]. The study of convolutional neural networks began in the 1980s and 1990s. LeNet was the earliest convolutional neural networks in practice [17]. After the turn of the twenty-first century, with the proposal of deep learning theory and the improvement of numerical computing equipment, CNNs have been developed rapidly.

In recent years, deep learning methods have gradually been widely used in the field of computer vision. AlexNet won the championship in the 2012 ImageNet large-scale visual recognition challenge (ILSVRC12). The error rate of Top-5 was only 15.3%, which was significantly improved compared with those previous works from 26.2% [18]. Since then, various advanced convolutional neural network structures, such as GoogLeNet, VGGNet and ResNet [19]–[22], have been proposed, and they have made great progress in tasks related to the field of computer vision. Some scholars have carried out research on the detection of pig body size using deep learning [23]. A method based on deep learning can automatically learn image features in the case of a large amount of data indeed.

In view of the limitations of the support vector machine method for detecting pig fat content, this paper will apply deep learning methods to detect pig fat content in B-ultrasound images. Compared with traditional machine learning, CNN has a deep hidden layer network structure and rich feature expression and does not require much manual information extraction. With the help of forward and backward propagation, it automatically learns image features from data sets and obtains more accurate, high-dimensional and abstract features on the basis of a deep network architecture. These features will be more conducive to improving the accuracy of regression prediction.

3. The Overall Architecture

The structure of CNN is mainly divided into a convolution layer, pooling layer and full connection layer. Each layer plays a different role, where the convolution layer extracts features from the input image by convolution, the pooling layer reduces the size of the input feature map, accelerates the calculation and reduces the probability of overfitting, and the full connection layer connects all learned features and maps them into the markup space. Based on the above, we proposed HAFFNet (hybrid attention feature fusion network), which is a convolutional neural network model.

Our improved model is based on VGG16. However, in the last two stages, we use two continuous convolutions to replace the original three convolutions, which can reduce the number of parameters and accelerate network training on the basis of maintaining simplicity and efficiency. To further improve network performance, HAFFNet uses a depthwise overparameterized convolutional (DO-Conv) [24] layer instead of traditional convolution. To obtain more important fat feature information from pig B-ultrasound images, the model adds the CBAM [25] feature extraction module. We also adopt upsampling on the small feature map and fuse it with the previous feature map. With this design, the feature map, to be learned at different stages, is endowed with both shallow and deep features, which will strengthen the generalization ability of the model. To overcome overfitting to some extent, we have tried the adaptive activation function ACON [26], L^2 regularization [27] and the dropout algorithm [28].

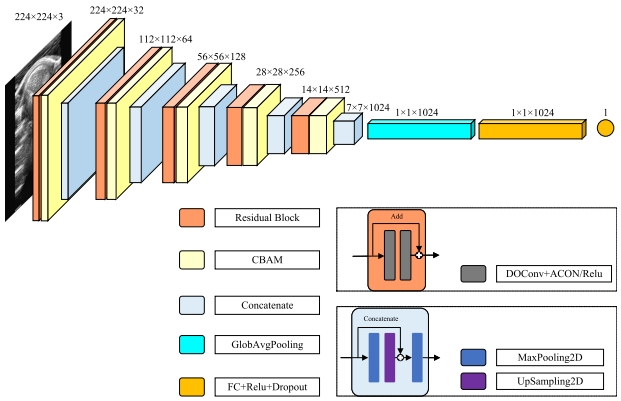


Fig. 1 Overall network structure of HAFFNet.

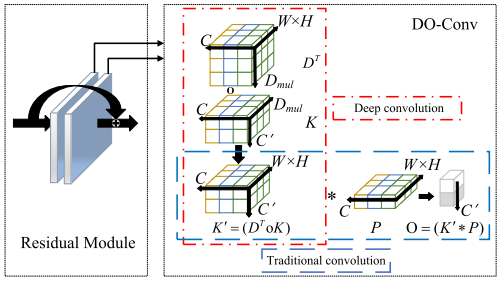


Fig. 2 Residual module with DO-Conv.

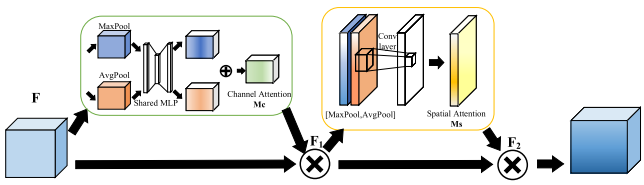


Fig. 3 Hybrid domain feature extraction mechanism.

3.1 Construction of HAFFNet

Since 245 pixels and 309 pixels is the original resolution of the image, we reduce the width and height to a uniform value considering the size inconsistency caused by the odd value during the process between downsampling and upsampling. The final dimension of the B-ultrasound image is 224×224×3. The numbers of residual layers, feature fusion layers, global average pooling layers and full connection layers used in our model are 5, 5, 1 and 1, respectively. The CBAM module is added between the convolution layer and feature fusion layer. Each residual module contains two DO-Conv operations and will connect the input to the output after convolution. ACON, the activation function, will be added behind the third residual module. The final feature fusion is followed by a global average pooling layer. Additionally, ReLU and dropout are added behind the first fully connected layer. Finally, the predicted fat content of pig B ultrasound images is output by a full connection. The overall model network structure of HAFFNet is shown in Fig. 1.

3.2 Residual Module with DO-Conv

In the residual module, we not only add the residual structure to prevent network degradation, but also use DO-Conv to replace the traditional convolution filter, so that the network will converge faster and can converge to a lower error. DO-Conv enhances the convolution layer by additional depthwise convolution. The size of the input feature map is $W{\times}H{\times}C$, and the size of the output feature map remains unchanged, where W represents the width of the feature map, H represents its height, and C represents the number of channels. There are two continuous DO-Conv in our residual module. DO-Conv has two equivalent training methods, and we use kernel composition with less computation, as shown in Fig. 2. In the process, o represents the deep convolution, * represents the traditional convolution, D represents the depthwise convolution trainable kernel, K represents the traditional convolution trainable kernel and P represents the area of convolution applied to the corresponding size of the feature map. D_{mul} is often referred to as the depth multiplier. In terms of kernel composition, $K' = (D^T \mathrm{o} K)$ is obtained by deep convolution, and then the output feature map $O = (K' * P)$ is obtained by traditional convolution.

3.3 Hybrid Domain Feature Extraction Module: CBAM

CBAM is a hybrid domain attention mechanism. It mainly imitates the important characteristics of human selective attention in the visual system, so that the network pays more attention to the recognition of the target area. In view of the complexity of fat content in B-ultrasound images of pigs, we add the CBAM feature extraction module to assign weights to each channel and each pixel, as shown in Fig. 3. In the channel dimension, a weight is used to represent the importance of the channel in the next step, and then in the spatial dimension, a weight is used to represent the importance of a pixel in the space. The two steps help the network obtain more important feature information.

In this paper, the CBAM is set between each residual layer and feature fusion layer, which can not only stimulate the fat features from the dataset but also amplify the difference between the fat features and nonfat features, which is conducive to better prediction of pork fat content. In CBAM, the dimension of the input feature map, F, is $W{\times}H{\times}C$, and the output feature map, F_2, is $W'{\times}H'{\times}C'$. These dimensions remain unchanged. For F, two $1{\times}1{\times}C$ features are obtained by average pooling and maximum pooling, respectively. Then, they are respectively sent to a neural network with two layers.

The number of neurons in the first layer is C/r, where r is the reduction ratio, and the second layer is C. ReLU is the activation function. The neural network is shared. Then, the weight coefficient Mc is obtained by adding the two obtained features to a sigmoid. Finally, the new feature, F_1, can be obtained by multiplying the weight coefficient with the original F. For feature map F_1 with $H{\times}W{\times}C$, we first perform

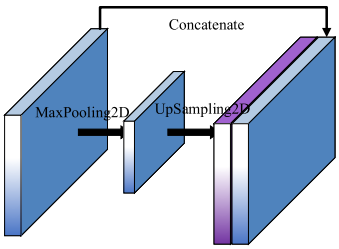


Fig. 4 Feature fusion mechanism.

average pooling and maximum pooling among the channels to obtain two feature maps with $H\times W\times 1$ and connect them together according to the dimension of the channel. Then, after 7×7 convolution, the weight coefficient Ms is obtained from the sigmoid. Finally, the new feature F_2 is obtained by multiplying Ms and F_1.

3.4 Multilayer Feature Fusion Module

To strengthen the feature learning, in different stages after DO-Conv convolution, feature extraction and max-pooling, the small feature map is upsampled and fused with the feature map before max-pooling, as shown in Fig. 4, which will give the feature map to be studied with both shallow and deep features and strengthen the recognition ability of the model. If the size of the feature map before maximum pooling is $H\times W\times C$, the size after that will be $H/2\times W/2\times C$, and after upsampling, the size will be restored to $H\times W\times C$. However, the feature at that time will be a deep feature. After fusing with those shallow features before maximum pooling in a dimension, the size of the feature map will become $H\times W\times 2C$. That is, the feature map at this time has both shallow and deep features, and the features learned by the network will be more abundant.

3.5 Adaptive Activation Function

Activation functions are divided into saturated activation functions, such as sigmoid and tanh, and unsaturated activation functions, such as ReLU and its variants. The unsaturated activation function, such as ReLU, sets all negative numbers in the matrix to 0, which can solve the gradient disappearance problem to a certain extent and accelerate the convergence; therefore, it is widely used. Although ReLU is commonly used, ReLU has the problem of training vulnerability. When a large gradient flows through a ReLU neuron and the parameters are updated, the neuron gradient will always be zero. The neuron will no longer activate any data. To avoid such a problem and to improve the nonlinear expression ability of the network, we try to use the adaptive activation function ACON to replace the traditional ReLU. The expression of ACON is shown in (1).

$$(p_1 - p_2)x \cdot \sigma(\beta(p_1 - p_2)x) + p_2 x \tag{1}$$

p_1 and p_2 are two learnable parameters used for adaptive adjustments that are represented by two 1×1 traditional convolutions in the network. The value of β will control whether neurons are activated ($\beta = 0$, i.e., not activated). σ is the sigmoid activation function.

Table 1 Parameter configuration.

name	type	kernel size	strides	kernel number	output
I(data)	Input				
Conv1_1	DO-Conv	3×3	1	32	224×224×32
Conv1_2	DO-Conv	3×3	1	32	224×224×32
CBAM					224×224×32
Concat1	Concatenate				112×112×64
Conv2_1	DO-Conv	3×3	1	64	112×112×64
Conv2_2	DO-Conv	3×3	1	64	112×112×64
CBAM					112×112×64
Concat2	Concatenate				56×56×128
Conv3_1	DO-Conv	3×3	1	128	56×56×128
Conv3_2	DO-Conv	3×3	1	128	56×56×128
CBAM					56×56×128
Concat3	Concatenate				28×28×256
Conv4_1	DO-Conv	3×3	1	256	28×28×256
Conv4_2	DO-Conv	3×3	1	256	28×28×256
CBAM					28×28×256
Concat4	Concatenate				14×14×512
Conv5_1	DO-Conv	3×3	1	512	14×14×512
Conv5_2	DO-Conv	3×3	1	512	14×14×512
CBAM					14×14×512
Concat5	Concatenate				7×7×1024
GAP	GlobalAverage Pooling				1024
FC1	Dense				1024
Out	Dense				1

3.6 Dropout Algorithm

Dropout can reduce the amount of calculation of the network to a certain extent and improve the efficiency of the network to predict the fat content of pig B-ultrasound images. The specific content of the algorithm makes the neurons in the network stop working at a certain probability when training. However, when the next sample is input, since the neurons stop working at a certain probability, the neurons that last did not work may start working again in this training process. Therefore, each input of the sample is equivalent to randomly selecting a different network from the original network for training. In other words, dropout can reduce the probability of overfitting to a certain extent. In this paper, each neuron has a 50% probability of temporarily not working at each training so that the emergence of one neuron does not depend on another neuron and further improves the generalization ability of the network.

3.7 Parameter Configuration of HAFFNet

The model parameter configuration is shown in Table 1. The I(Data) layer is the input layer. It will prepare three-channel color images with 224 pixels and 224 pixels for the whole network. Conv1_1 is the first convolution layer, which is composed of 32 feature maps, and its convolution kernel size is 3 pixels and 3 pixels. Conv1_2 is the second con-

nected convolution layer of 3 pixels and 3 pixels, and the number of convolution kernels in this layer is set to 32. Since the number of convolution kernels needs to be consistent with the number of channels, the output is 32 feature maps of 224 pixels and 224 pixels. For the first CBAM module added, because the output dimension processed by CBAM is consistent with the input dimension, the output is also 32 feature maps of 224 pixels and 224 pixels. Concat1 represents the first feature fusion layer, and the output dimension should satisfy the logical relationship described in (2). FC1 and Out are full connected layers. Finally, HAFFNet will output a one-dimensional vector to predict the fat content of pig B-ultrasound images.

$$\begin{cases} W^i = \frac{W^{i-1}+2\times R-F+S}{S} \\ H^i = \frac{H^{i-1}+2\times R-F+S}{S} \end{cases} \tag{2}$$

In (2), (W^{i-1}, H^{i-1}) means the feature map of the previous convolution layer as the input to the current layer. R is the boundary width added to the current feature map. F is the size of the convolution kernel of the current convolution layer. S is the step length of the pooling layer. (W^i, H^i) stands for the feature map of the current layer after pooling.

3.8 Model Optimization

Ideally, we hope that their model can quickly correct errors and obtain more accurate results, but it is often difficult to achieve the expected results in practice. The loss function can estimate the distance between the predicted results and the correct labels. The smaller the value of the loss function, the better the effect of the model prediction. The selection of the loss function needs to be based on the specific network and the problems to be solved.

The mean squared error (MSE), one of the regression loss functions, is the average of the squared distance between the estimated value and the correct value, as shown in (3). In (3), y is the actual output of the network structure model, and the probability distribution $\hat{y}$ is the expected output (i.e., real label). However, MSE will give the unreasonable average due to outliers and reduce the overall performance of the network model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

MAE, as another regression loss function shown in (4), is used to measure the average value of the distance between y and $\hat{y}$ of the sample. Compared with MSE, MAE is more inclusive for outliers with data damage or wrong sampling. However, due to the existence of non differentiable cusps, MAE is not completely conducive to the convergence of functions and the training of models.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

In HAFFNet, we chose Huber loss, like (5). δ is a parameter, usually 0.1 or 1. Huber combines the advantages of MSE and MAE, and has strong anti-interference ability to outliers. At the same time, it can provide convenience for obtaining derivative everywhere, which is more conducive to model training.

$$H(y,\hat{y}) = \begin{cases} \frac{1}{2}(y-\hat{y})^2 & ,\text{for } |y-\hat{y}| \le \delta \\ \delta\cdot(|y-\hat{y}| - \frac{1}{2}\delta) & ,\text{otherwise.} \end{cases} \tag{5}$$

4. Experiments

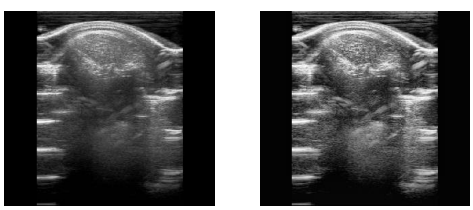
To better predict and analyze the fat in the B-ultrasound images of pigs, we collected images from the Key Laboratory of Pig Science, Academy of Animal Husbandry Sciences, as the experimental data in this paper. Then, several groups of experiments were conducted. The first group of experiments compares those loss functions to our model. The second group of experiments compares the detection accuracy of multiple network models to verify that the network structure proposed in this paper has a good effect. In the third group of experiments, ablation was carried out in our model to compare and analyze how different structures influence the whole model and to tell us whether they have a greater impact on the overall performance.

4.1 Dataset

The original data set used has a total of 130 groups of B-ultrasound images of porcine eye muscles, which are sampled by animal full digital B-ultrasound instruments. Pigs are slaughtered 24 hours after live B-ultrasound image acquisition. After slaughter, eye muscles of the 10th-12th thoracic vertebrae of each test pig were collected, and the data, such as pig number, measurement time and place, were labeled and sent to the Academy of Animal Science. The Soxhlet extraction method was used to determine and analyze the fat content of porcine eye muscle to form labels for fat content prediction.

The analysis of the collected samples shows that the original data have noise or low contrast due to the small range of gray levels of the image. To solve these problems, we first tried method A, contrast-limited adaptive histogram equalization.

Part a of Fig. 5 is one of the original images of the pig B-ultrasound image dataset, and Part b is the image after method A. Figure 6 shows the change in the gray histogram. It was obvious that the range of gray values became larger after treatment, increasing from 0 to 100 and 0 to 150, and



a b

Fig. 5 Image comparison after cpplication method A.

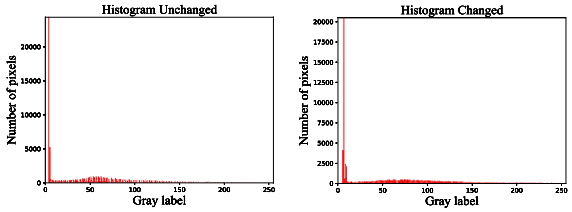


Fig. 6 Histogram comparison after application method A.

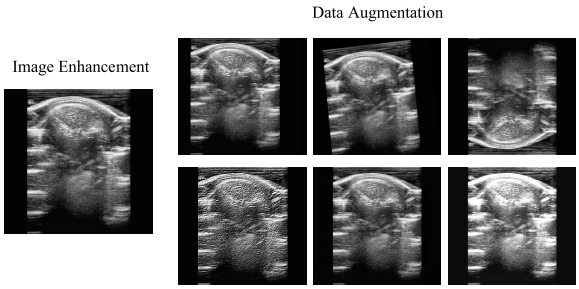


Fig. 7 Comparison of image data before and after enhancement.

the middle area was also more dispersed and uniform. This shows that the contrast of the image is more obvious and the image is clearer. Figure 5 confirmed our analysis.

Due to the complex structure of convolutional neural networks, a large amount of training data are needed to support training to avoid overfitting. However, it is difficult to obtain large amounts of image data for practical applications, so data enhancement, as an effective method to obtain large amounts of data, arises at a historic moment. Currently, data enhancement is a very common and effective method to improve robustness and reduce overfitting. In this paper, the dataset is expanded by translation, rotation, mirroring, sharpening, changing pixel values and brightness. An example is shown in Fig. 7.

Each image enhancement method can double each image of the original data set. If the original image has 130 sets of image data, after image enhancement and then after six kinds of image enhancement steps including five random translations, five random rotations, five random mirrors, five random sharpenings, five random changes in pixel value and five random changes in brightness, the 130 sets of data will be expanded to 130+130×5×6=4030 sets of image data. In addition, we divide the data set into training, verification and testing according to the ratio of 6:2:2. In other words, there are 2580 images in training, 806 in verification and 806 in testing.

4.2 Comparative Experiment of Loss Function

To achieve a fair comparison between algorithms and prevent the network from reaching the error threshold and ending the training in advance, the minimum value of the network loss function should be set as 0, and the updated mini-batch size can be 32. There are 500 training epochs to better observe the change in the loss value of different loss functions when the number of epochs increases. In this paper, MSE, MAE and Huber are added to HAFFNet for comparative experiments, and the results are shown in Fig. 8. Combined with the changes between loss value (part a) and accuracy (part b) in Fig. 8, the loss values of MSE, MAE and Huber decrease with the increase of epoch times. Figure 8 shows that the loss value of Huber decreases rapidly, the loss value tends to zero after 125 epochs, and the network converges quickly. However, the loss values for the other two functions still fluctuated after 200 epochs. After the change in accuracy, when the network with Huber is trained, the prediction accuracy of the training set can reach 98.34%, and the verification set can reach 96.49%. In the experiment, it can be observed that Huber is more suitable for HAFFNet.

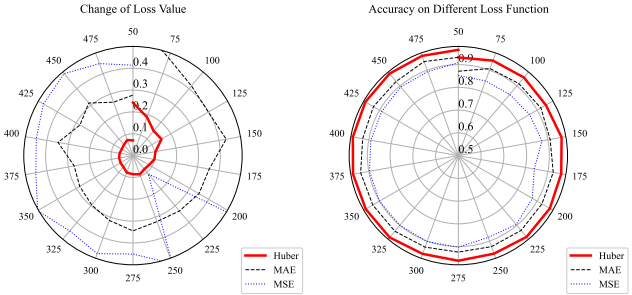


Fig. 8 Comparison on three loss functions.

4.3 Comparative Experiment and Analysis

We chose the current mainstream deep learning networks VGG16, ResNet, DenseNet and HAFFNet to compare the prediction accuracy. The structure of VGG16 is famous for its neat and concise structures, and there are few hyperparameters. The improvement of performance is due to deepening of the network structure. ResNet maintains performance by stacking a large number of residual structures. DenseNet uses concatenation to learn huge features in exchange for performance.

Based on the regression loss function, we add regularization to further reduce the risk of overfitting. Here, we adopt L^2 regularization, and the calculation is shown as (6).

$$\|\tilde{y}\|_2 = (\sum_{i=1}^{n} |\tilde{y}_i|^2)^{\frac{1}{2}}, \tilde{y} = y_i - \hat{y}_i \tag{6}$$

The optimization algorithm used in this paper is Adam (Adaptive Moment Estimation) instead of random gradient descent, which is a traditional method. It can iteratively update the weights of neural networks based on training data. The essence of Adam is to dynamically adjust the learning rate of each parameter by using the first-order and second-order matrix estimates of the gradient. The main advantage is that after bias correction, each iterative learning rate has a range so that the parameters do not undergo a large shock but a more stable change.

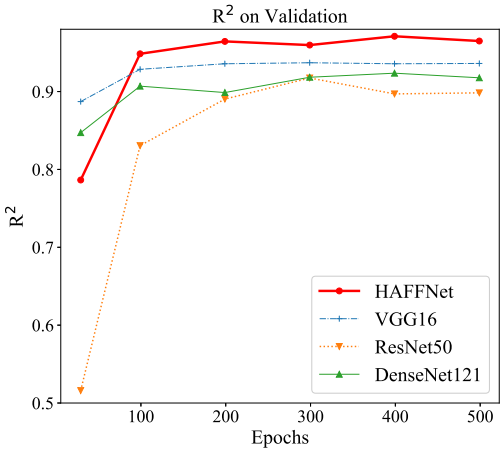


Fig. 9 Changes of R^2.

Table 2 Validation R^2 comparison under different epochs.

epoch	VGGNet	ResNet	DenseNet	HAFFNet
300	0.9370	0.9174	0.9184	0.9597
400	0.9357	0.8970	0.9236	0.9710
500	0.9361	0.8983	0.9177	0.9649
mean	0.9362	0.9042	0.9139	0.9652

The determination coefficient (R^2) is used for our main evaluation criterion. R^2 is a very common statistical method in regression analysis and is often used as a standard to measure the prediction ability of the model. The range of R^2 varies from 0 to 1, as shown in (7), indicating the percentage of squared correlation between the predicted value and the actual value of the target. R^2 may be negative at the beginning when training, and the prediction may not be as good as the average value if directly calculated.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{7}$$

With the same number of epochs, the learning rate is $1e^{-4}$, and the dropout rate is 0.5. NVIDIA 3070 GPU was used to accelerate training. The specific prediction accuracy on validation and testing of each network after 300, 400, and 500 epochs is shown in Fig. 9 and Table 2. In contrast, there are fluctuations in model training, but our model is better in terms of accuracy and stability. Our model is the optimal choice at present.

After 4-fold cross validation, the mean value of R^2 reached 0.9382 and best value reached 0.9629 on test data with HAFFNet. Besides, the mean value of MSE reached 0.3937, best value reached 0.2382. The mean value of MAE reached 0.2489, best value reached 0.2223. It shows that our model performs well on various evaluation metrics.

Considering the parameters and FLOPs from Table 3, the computational time cost of HAFFNet is lower than that of VGG16 and ResNet50. Although the computational time cost of HAFFNet is not the best scheme to control, it can improve the prediction accuracy as much as possible, and the overall effect of HAFFNet is better.

Table 3 The computational time cost under different models.

Model	Total params	FLOPs
HAFFNet	7.48M	8.48 G
VGG16	38.53M	30.8 G
ResNet50	22.50M	7.7 G
DenseNet121	6.71M	5.7 G

Table 4 Comparison with different detection strategies.

Fat content detection strategy	R^2
B-ultrasound+Deep Learning (HAFFNet)	0.93(cross validation)
CT+PLSR(2019) [30]	0.83(optimal value) [30]
Hyperspectral Imaging+MLR(2017) [31]	0.87(cross validation) [31]
Hyperspectral Imaging+ MSC+CARS+PLSR(2021) [29]	0.96(optimal value) [29]

Table 5 Parameter control of ablation experiment (A).

condition	√ means selected				
A		√			
B			√		
C				√	
D					√
R^2	0.8848	+0.33%	+3.98%	+2.00%	+1.65%
time[a](s)	6	6	14	9	10

[a]The time of each epoch.

Table 6 Parameter control of ablation experiment (B).

condition	√ means selected				
A	√	√		√	√
B	√		√	√	√
C		√	√	√	√
D					√
R^2	+3.52%	+0.21%	+4.81%	+0.79%	+8.01%
time(s)	14	10	17	19	22

Comparing the method of this paper with the method of pig fat content detection in recent years, as shown in Table 4, it can be found that the method of this paper is a quite well supplement to the task, and for the index of correlation coefficient, the method of this paper can achieve a relatively ideal effect. Although the multivariate scattering correction (MSC) and competitive adaptive reweighed sampling (CARS) are used in [29] and the determination coefficient of 0.962 can be obtained by combining better quality hyperspectral images, the hyperspectral equipment lacking price advantage is not conducive to promotion at this stage. In contrast, the economic benefit of research is higher on the basis of sampling using B-ultrasound equipment.

4.4 Ablation Experiment

To verify the effect of each part on the performance of our overall model, we designed ablation experiments. As shown in Table 5 and 6, in the Baseline, we restore DO-Conv to ordinary convolution, temporarily shield CBAM, and do not perform feature fusion. The activation function is unified as ReLu.

To simplify the description, A is used to represent DO-Conv, B is a hybrid domain feature extraction module, C

is feature fusion, and D is an adaptive activation function. Considering the individual effects on the test set, it can be observed that the effect of B is the most obvious. The combined effect of B with C greatly improves the network performance compared with A+B and A+C. Of course, we observed that the effect of the last column is best.

5. Conclusion

In this paper, we combined the self-defined deep learning model on the B-ultrasound images to realize a convenient, economical and popularized pork fat content detection method. Through a comparative test of multiple network structures, such as VGG16, ResNet and DenseNet, the advantages of HAFFNet in predicting fat content in B-ultrasound images of pigs were verified. Through ablation experiments, it was found that different modules have different effects on the entire network. Flexible use of multiple modules can make the entire model have better convergence and improve the accuracy of the model. Compared with the current better fat detection scheme, further research based on B-ultrasound is the most economically feasible strategy. In future research, the network level can be improved on the basis of the network model HAFFNet, the network structure can be improved, and more abundant and representative fat features of pig B-ultrasound images can be extracted to further improve the prediction accuracy. However, this means that the network is more complex and has more parameters, so optimizing the algorithm and improving the processing speed are problems that should be solved.

Acknowledgments

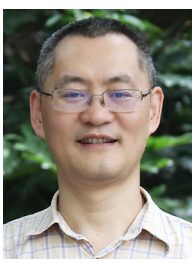
The authors would like to appreciate the anonymous reviewers for their useful comments to improving the quality of this paper. In addition, we thank the College of Computer Science and Engineering, Chongqing University of Technology, China for supporting this research. This research was supported by Science and Technology Research by Chongqing Municipal Education Commission (Project number:KJZD-K201801901) and supported by Chongqing Postgraduate Research Innovation Project Funding (Grant No.CYS21470).

References

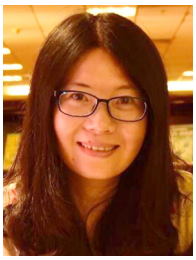
[1] A. Taheri-Garavand, S. Fatahi, M. Omid, and Y. Makino, "Meat quality evaluation based on computer vision technique: a review," Meat Sci., vol.156, pp.183–195, 2019.
[2] C.T. Kucha, L. Liu, and M.O. Ngadi, "Non-destructive spectroscopic techniques and multivariate analysis for assessment of fat quality in pork and pork products: a review," Sensors, vol.18, no.2, p.377, 2018.
[3] D.W. Park, D.C. Park, and S.H. Chung, "Ultrasound signal processing technique for subcutaneous-fat and muscle thicknesses measurements," IEEE Access, vol.7, pp.155203–155208, 2019.
[4] P. Janiszewski, K. Borzuta, D. Lisiak, E. Grześkowiak, and D. Stanisławski, "Prediction of primal cuts by using an automatic ultrasonic device as a new method for estimating a pig-carcass slaughter and commercial value," Animal Production Science, vol.59, no.6, pp.1183–1189, 2019.
[5] Y.Y. Shi, X.C. Wang, M.S. Borhan, J. Young, D. Newman, E. Berg, and X. Sun, "A review on meat quality evaluation methods based on non-destructive computer vision and artificial intelligence technologies," Food Science of Animal Resources, vol.41, no.4, pp.563–588, 2021.
[6] C.J. Yuan, D. Shi, D.Q. Liu, X.B. Lv, and R. Liu, "An improved algorithm of pig's intramuscular fat detection," Mod. Comput., vol.18, pp.31–36, 2013.
[7] M. Zhang, N. Zhong, and Y.Y. Liu, "Estimation method of pig lean meat percentage based on image of pig shape characteristics," Transactions of the Chin. Soc. of Agric. Eng., vol.33, pp.308–314, 2017.
[8] J.X. Zhang, T. Li, Q. Sun, and T.T. Xie, "Texture feature extraction and classification of pork loin ultrasonography images," Journal of Chongqing University of Technology (Natural Science), vol.22, no.2, pp.74–78, 2013.
[9] Q. Wu, Y.G. Liu, Q. Li, S.L. Jin, and F.Z. Li, "The application of deep learning in computer vision," 2017 Chinese Automation Congress (CAC), pp.6522–6527, 2017.
[10] J.Y. Chai, H. Zeng, A. Li, and E.W.T. Ngai, "Deep learning in computer vision: a critical review of emerging techniques and application scenarios," Machine Learning with Applications, vol.6, 100134, 2021.
[11] R.R.P. Kumar, S. Muknahallipatna, and J. McInroy, "An approach to parallelization of sift algorithm on gpus for real-time applications," J. Comput. and Communications, vol.4, no.17, pp.18–50, 2016.
[12] M.-E. Ilas and C. Ilas, "A new method of histogram computation for efficient implementation of the hog algorithm," Computers, vol.7, no.1, p.18, 2018.
[13] V.K. Ojha, A. Abraham, and V. Snášel, "Metaheuristic design of feedforward neural networks: a review of two decades of research," Engineering Applications of Artificial Intelligence, vol.60, pp.97–116, 2017.
[14] Y. Liu, K.W. Wen, Q.X. Gao, X.B. Gao, and F.P. Nie, "Svm based multi-label learning with missing labels for image annotation," Pattern Recognition, vol.78, pp.307–317, 2018.
[15] N. Kriegeskorte and T. Golan, "Neural network models and deep learning," Current Biology, vol.29, no.7, pp.R231–R236, 2019.
[16] L. Alzubaidi, J.L. Zhang, A.J. Humaidi, A. AI-Dujaili, Y. Duan, O. AI-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," Journal of Big Data, vol.8, no.1, pp.1–74, 2021.
[17] S. Pouyanfar, S. Sadiq, Y.L. Yan, H.M. Tian, Y.D. Tao, M.P. Reyes, M.-L. Shyu, S.-C. Chen, and S.S. Iyengar, "A survey on deep learning: algorithms, techniques, and applications," ACM Computing Surveys, vol.51, no.5, pp.1–36, 2018.
[18] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, vol.60, no.6, pp.84–90, 2017.
[19] M.Z. Alom, T. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, B.C.V. Esesn, A.S. Awwal, and V.K. Asari, "The history began from alexnet: a comprehensive survey on deep learning approaches," arXiv preprint arXiv:1803.01164, 2018.
[20] M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, M. Hasan, B.C.V. Esesn, A.A.S. Awwal, and V.K. Asari, "A state-of-the-art survey on deep learning theory and architectures," Electronics, vol.8, no.3, p.292, 2019.
[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
[22] D. McNeely-White, J.R. Beveridge, and B.A. Draper, "Inception and resnet features are (almost) equivalent," Cognitive Systems Research, vol.59, pp.312–318, 2020.
[23] Z. Li, X.D. Du, T.T. Mao, and G.H. Teng, "Pig dimension detection system based on depth image," Transactions of the Chin. Soc. for

Agric. Mach., vol.47, pp.311–318, 2016.
[24] J.M. Cao, Y.Y. Li, M.C. Sun, Y. Chen, D. Lischinski, D. Cohen-Or, B. Chen, and C.H. Tu, "Do-conv: depthwise over parameterized convolutional layer," Preprint arXiv:2006.12030, 2020.
[25] S. Woo, J. Park, J.-Y. Lee, and I.S. Kweon, "Cbam: convolutional block attention module," Proc. ECCV Conf., Preprint arXiv:1807.06521, 2018.
[26] N.N. Ma, X.Y. Zhang, M. Liu, and J. Sun, "Activate or not: learning customized activation," Proc. CVPR Conf., pp.8032–8042, 2021.
[27] Y.J. Tian and Y.Q. Zhang, "A comprehensive survey on regularization strategies in machine learning," Information Fusion, vol.80, pp.146–166, 2022.
[28] C. Garbin, X.Q. Zhu, and O. Marques, "Dropout vs. batch normalization: an empirical study of their impact to deep learning," Multimedia Tools and Applications, vol.79, no.19-20, pp.12777–12815, 2020.
[29] Y. Li, Y.K. Peng, Y.Y. Li, Q.B. Zhuang, and Q.H. Guo, "Nondestructive detection of pork fat content based on hyperspectral spectroscopy," ASABE Annual International Virtual Meeting, 2021.
[30] M. Font-i-Furnols, A. Brun, and M. Gispert, "Intramuscular fat content in different muscles, locations, weights and genotype-sexes and its prediction in live pigs with computed tomography," Animal, vol.13, no.3, pp.666–674, 2019.
[31] H. Huang, L. Liu, and M.O. Ngadi, "Assessment of intramuscular fat content of pork using nir hyperspectral images of rib end," Journal of Food Engineering, vol.193, pp.29–41, 2017.

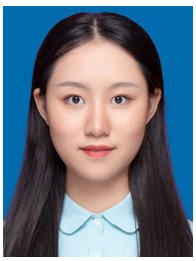
Wenxin Dong received his B.Eng. degree from Yangtze Normal University, China in 2019. He has won the national scholarship twice. Since 2020, he has been a Master course student at College of Computer Science and Engineering, Chongqing University of Technology, China. He used to work as a teaching assistant. He is now a member of CCF. The main research area of him is computer vision.



Jianxun Zhang received his B.S. degree from Chongqing Normal University, China in 1994 and his M.S. and Dr. Eng. degrees from Chongqing University, China in 1997 and 2001. Postdoctoral from Tianjin University in 2005. In 2000, he joined Chongqing University of Technology, where he became full professor in 2006. Prof. Zhang is currently serving as the associate dean of the the College of Computer Science and Engineering at Chongqing University of Technology. Prof. Zhang has published 45 well-known journal papers or conference papers, and has served as the editor-in-chief or editorial board member of textbooks for many times. Prof. Zhang was responsible or participated in more than 20 key projects such as the National Natural Science Foundation. Prof. Zhang's research interests are concentrated in computer application technology, computer graphics and other fields, with an emphasis on efficient image analysis and deep learning.



Shuqiu Tan received her M. Eng. degree from Chongqing University of Technology, China in 2011 and her Dr. Eng. degree from University of Electronic Science and Technology, China in 2017. From 2020 till now, Dr. Tan is doing post-doctoral research at Chongqing University and Chongqing Shanwaishan Blood Purification Technology Co., Ltd. Dr. Tan has already more than 10 high-quality published papers and has been responsible for a number of key projects. Dr. Tan has been a lecturer at College of Computer Science and Engineering, Chongqing University of Technology since 2017. Her main research area includes computer vision, image processing, augmented reality.



Xinyue Zhang is currently an undergraduate student at Sydney Smart Technology College, Northeastern University, China.