ITC 2/54

Information Technology and Control
Vol. 54 / No. 2/ 2025
pp. 536-559
DOI 10.5755/j01.itc.54.2.35918

A Two-stage Cattle Face Recognition Method Based on Target Detection and Recognition Network

Received 2024-01-02 | Accepted after revision 2024-04-25

HOW TO CITE: Zheng, P., Deng, M., Gong, J., Li, G., Yin, Y. (2025). A Two-stage Cattle Face Recognition Method Based on Target Detection and Recognition Network. *Information Technology and Control*, 54(2), 536-559. https://doi.org/10.5755/j01.itc.54.2.35918

A Two-stage Cattle Face Recognition Method Based on Target Detection and Recognition Network

Piaoyi Zheng, Minghui Deng*, Junjie Gong, Guiping Li and Yanling Yin
College of Electrical and Information, Northeast Agricultural University, Harbin 150030, China

Corresponding author: Minghui Deng, markdmh@163.com

Traditional methods of cattle management have problems such as high error rates, easy failure of tags, and the need to consume a lot of time and manpower costs. However, as one of the biological characteristics, the recognition of cattle face is one of the important technical means to achieve intelligent farming, accurate feeding, and health management of cattle. Thus, the article proposed improved algorithms based on YOLOv7 and VoVNet for cattle face detection and recognition using a contactless approach. For the improved YOLOv7 cattle face detection model, the efficient layer aggregation networks (ELAN) structures in the backbone and neck networks were replaced with the ConvNeXt network and CoTNet Transformer module, respectively, aiming to improve the detection speed and robustness while reducing computation. The SimAM (A Simple, Parameter-Free Attention Module) attention mechanism, considering both spatial and channel dimensions, was introduced in the neck network to enhance feature representation without adding extra parameters to the original network. Experimental results on the constructed facial detection dataset of Holstein and Simmental beef cattle showed that the improved CCS-YOLOv7 cattle face detection model achieved a precision of 99.43% and a recall rate of 99.10%, with significantly improved detection speed and reduced model size. As for the improved VoVNet cattle face recognition model, residual connections (RC) were added from the input to the output of the One-Shot Aggregation (OSA) modules of VoVNet to enhance the representation of deep features. The Efficient Channel Attention (ECA) was added to the final feature extraction layer of the OSA modules to improve the feature extraction capability for cattle face image classification. Experimental results on the facial recognition dataset of Holstein dairy cows and Simmental beef cattle, built upon the improved CCS-YOLOv7 cattle face detection model, demonstrated that the VoVNet-ECA-RC model achieved a precision of 99.37% for cattle face recognition with a final model size of 41.4MB. Therefore, the proposed research structures can provide a reference for non-contact individual recognition in the process of intelligent farming.

KEYWORDS: Cattle face recognition; Cattle face detection; YOLOv7; VoVNet; Attention mechanism

1.Introduction

According to the USDA Statistical Report and other data, in 2022, global cattle slaughter was 300 million head; beef production was expected to be 73.9 million tonnes, an increase of 1.4% from the previous year; and milk production was 549,356 kilotonnes, an increase of 0.97%. And beef ranked second in total pig, cattle, lamb and poultry production. In summary, cattle management is crucial. Modern and efficient scientific, systematic and intelligent farming methods have become the key issues necessary to ensure the healthy development of the cattle industry. Accurate detection and identification of different cattle is important in cattle management, including precision farming, health traceability and animal welfare [8]. As farming scales and densities increase, so does the risk of disease. Accurate and rapid detection of target cattle can help managers focus on abnormal behaviours and take timely countermeasures to reduce disease incidence [5]. In addition, cattle face recognition technology has the application value of preventing fraudulent insurance [1], achieving win-win cooperation between farmers and insurance companies, and promoting the process of systematic management of the livestock industry [29].

In the livestock industry, cattle individual identification methods can be categorised into four methods: permanent methods, temporary methods, electronic identification and biometric identification [1]. Permanent method pins, which include ear grooves, ear tattoos, branding irons and freeze-engraved numbers, cause permanent damage to cattle and affect their welfare [36]. In addition, over time and with hair growth, these marks may become covered and difficult to recognize. Temporary methods, mainly ear tags, are used, and the materials used for ear tags may affect the health of the cattle as well as the high loss rate of ear tags. Electronic identification, i.e., RFID, is simple to use, but the cost of purchasing and maintaining the equipment is prohibitive and there are some information security risks. Biometric identification, e.g., nose prints, iris [15], retinal blood vessels [27], etc., suffers from low feasibility, high technical feasibility and high cost. Achieving recognition for different individuals in the same category is an area of potential, usefulness and development. Contactless recognition technology based on facial features can be applied in this field, with a promising future and application value.

Therefore, scholars have carried out extensive research on the recognition of livestock using image processing techniques. Xia et al. [30] showed that the recognition of cattle face images can be done by using Sparse Representation Classifier (SRC), extracting the features by Principal Component Analysis (PCA) and recognizing them using SRC. Chen Juanjuan et al. [3] proposed the use of improved bag of features model (BOF) for cattle head image recognition, on the BOF model based on spatial pyramid matching principle (SPM), the SIFT features are changed to HOG features, which reduces the complexity of cattle image features and improves the recognition performance. With the rise of deep learning, convolutional neural network performs well in image recognition, and many scholars use this technique for individual cattle recognition and achieve good results. Li et al. [13] applied a lightweight modification of neural network to recognize the faces of cattle, and conducted test experiments on the model on a Raspberry Pi, and the experimental results show that the model has good performance in terms of recognition accuracy and recognition speed. The model has good performance in terms of accuracy and speed. Yao et al. [36] combined cow face detection with cow recognition and used Faster CNN with PANdasnet to realize cow face detection and recognition in a breeding environment. Xu et al. [34] applied RetinaNet combined with ResNet50 to cattle face recognition, and the experimental results showed that the model has excellent performance in terms of accuracy and speed. Xu et al. [33] combined lightweight RetinaFace-mobilenet with additive angular spacing loss (ArcFace) combination to achieve 91.3% accuracy and recognition speed of 24 frames per second on a real scene dataset. Ma et al. [16] used a combination of offline knowledge distillation and lightweight networks for pig face recognition, which effectively balances accuracy and computational efficiency; they proposed a decoupled approach to pig face recognition that significantly reduces the need for exhaustive re-annotation when applying the model to new datasets. Cattle face recognition as a multi-classification problem, Połap et al. [17] used bilinear pooling with poisoning detection, which not only improves the accuracy of the model, but also enhances the security and robustness of the model. The living conditions of domestic animals in the actual environment are complex, and the process of collecting images has a low degree of domestic animal cooperation and a high degree of intrusiveness of background information. In order to reduce the extraction of redundant information as well as to improve the accuracy of the recognition results, most researchers adopt a multi-level recognition network model. Firstly, the more fine face image is extracted by the face detection model and input to the recognition network, so that the ratio of effective information between the target and background of the cattle face is increased, thus achieving high accuracy of livestock face recognition. Ali et al. [18] firstly detected and extracted the muzzle region of the cattle by using YOLOv3, and then applied ResNet-50 for biometric recognition, and the final recognition accuracy reached 99.11%, with an average inference time of 0.0259s per image. Bergman et al. [2] used YOLOv5 to detect the face of the cattle in the first stage, and Vision-Transformer was used to recognise the cattle in the second stage, and the study achieved an accuracy of 96.3% with an average inference time of 20ms per image.

In summary, the use of computer vision and pattern recognition methods to extract facial features of cattle to recognize individual identity has great advantages in terms of animal welfare, accuracy and practicality [32]. In this paper, we use a target detection algorithm to detect the cattle face, and then use a convolutional network to identify the cattle face using the cattle face detection result. Meanwhile, in order to improve the recognition accuracy of cattle faces in different light and from different angles, this paper proposes algorithmic models based on YOLOv7 and VoVNet to achieve the detection and recognition of cattle faces respectively. The above models not only improve the detection speed and recognition accuracy of cattle faces, but also enhance the robustness of the models by detecting and recognizing cattle faces from different angles and light.

We summarize our main contributions and innovations in the model as follows:

1 In order to better work on cattle face recognition, this paper performs cattle face image acquisition and constructs cattle face detection dataset and cattle face recognition dataset respectively.

2 In the cattle face detection stage, in order to reduce the computational cost of the network, the ELAN-A module in the Backbone and the ELAN-H module in the Neck of YOLOv7 are replaced with the ConvNeXt network structure and the COT module of the CoTNet Transtomer, respectively, with the aim of further reducing the number of parameters in the network, increasing the detection of image speed as well as enhance the extraction and representation of cattle face features; in order to further improve the feature representation capability of the network, the SimAM attention mechanism is added to the Neck part of YOLOv7, which achieves the suppression of irrelevant features of the image and improves the focusing of beneficial features for target detection.

3 In the stage of cattle face recognition, in order to adapt to a more complex network, residual connections are added from the input to the output of the OSA module of the VOVNet network for better extraction of the target′s deeper features; in order to improve the recognition performance of the network, the ECA attention mechanism is added to the final feature extraction of the OSA module, so that the network learns to recognise the effective features of the target and expresses them in a better way.

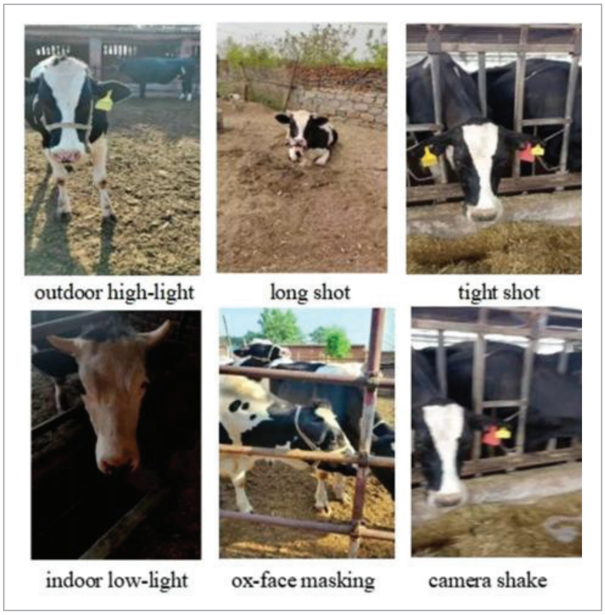
2.Materials and Methods

2.1 Dataset Acquisition

The experimental data were collected from cattle farms in Baoding City, Hebei Province and Tangshan City, Hebei Province from June 2023 to July 2023. Because of the diverse activity states and postures of cattle, it is difficult to directly capture high-quality images of cattle faces. To solve this problem, the cattle face video could be captured by hand-held smartphone first and processed by frame-by-frame interception. In order to improve the generalisation ability of the model, the dataset was used to collect the cattle face data from different angles and lighting when feeding and looking at the wind. A total of 55 video data were captured, which contained Holstein dairy cattle and Simmental beef cattle and the original video resolution was 544 × 960 (frame width × frame height), and some of the samples in the dataset are shown in Figure 1.

Figure 1

Sample raw data of cattle face.

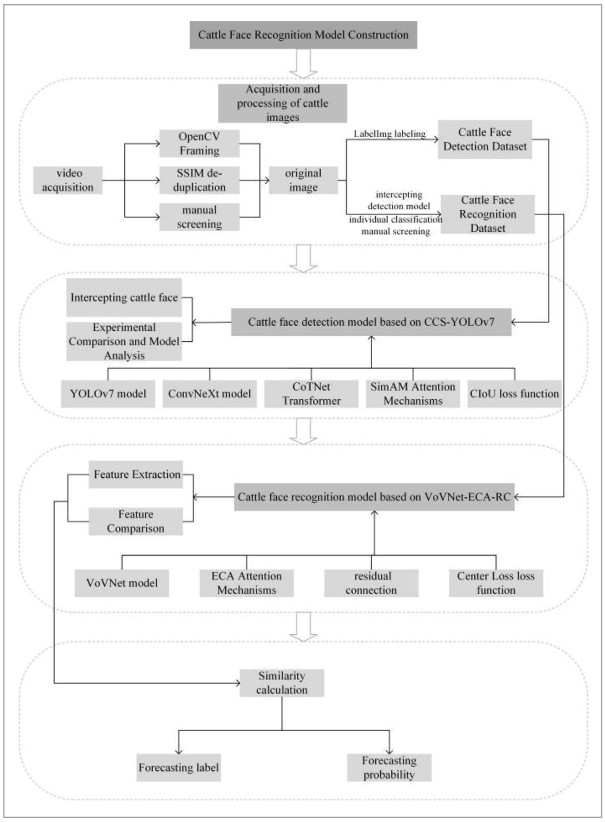


outdoor high-light long shot tight shot

indoor low-light ox-face masking camera shake

Figure 2

Technology roadmap.


Cattle Face Recognition Model Construction
Acquisition and processing of cattle images
video acquisition
OpenCV Framing
SSIM de-duplication
manual screening
original image
LabelImg labeling
Cattle Face Detection Dataset
intercepting detection model
individual classification
manual screening
Cattle Face Recognition Dataset
Intercepting cattle face
Experimental Comparison and Model Analysis
Cattle face detection model based on CCS-YOLOv7
YOLOv7 model
ConvNeXt model
CoTNet Transformer
SimAM Attention Mechanisms
CIoU loss function
Feature Extraction
Feature Comparison
Cattle face recognition model based on VoVNet-ECA-RC
VoVNet model
ECA Attention Mechanisms
residual connection
Center Loss loss function
Similarity calculation
Forecasting label
Forecasting probability

2.2 Technological Routes

It mainly includes four steps:

1. **Data set construction**

   Firstly, photos and videos of the cattle face were taken to obtain the original data of the cattle face. The original image of the cattle face was formed by using OpenCV frame-splitting, SSIM de-emphasis, and manual filtering of the target image without the cattle face. Corresponding datasets for different sub-tasks were constructed respectively.

2. **Cattle face detection**

   The LabelImg was used to construct cattle face detection dataset by annotating the cattle face region of the original cattle face image, and then the cattle face detection dataset was used to train the cattle face detection model based on the improved YOLOv7 model, which could be used to detect the cattle face in the actual application scenarios on one hand, and on the other hand, it could be used as the basis for the subsequent cattle face recognition dataset to be constructed automatically.

3. **Cattle face recognition**

   The original cattle face recognition dataset generated by the cattle face detection model was used for individual classification and data filtering to form the cattle face recognition dataset, which was used as the basis for training the VoVNet-ECA-RC cattle face recognition model proposed in this paper.

4. **Result Output**

   The data for cattle face recognition was prepared, and the trained cattle face recognition model was called. The recognition result, including the number of recognized cattle and the probability, was output based on the calculation of the feature vector and similarity.

2.3 Construction of Dataset

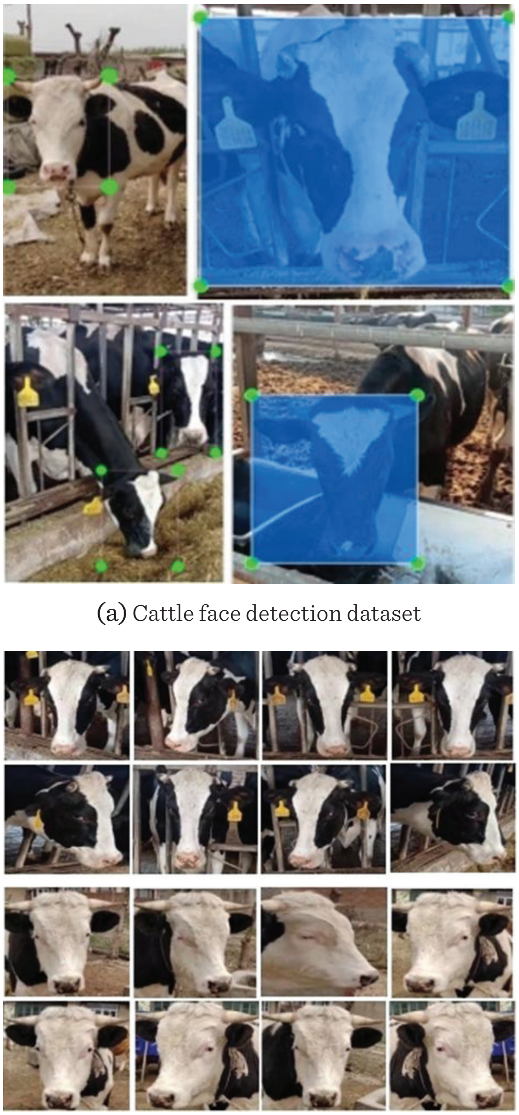
Firstly, one image is intercepted every five video frames using OpenCV; the cattle face dataset constructed by the above method is obtained directly from the video frames, which makes the similarity between the front and back frames high, and in order to avoid the overfitting situation in the later trained detection and recognition model, the extracted images need to be processed. Based on the above structural similarity method (SSIM) [23] can be used to check the images and delete the front and back frame images with high similarity. Given two images x and y, the expression is shown in Equation (1):

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\delta_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\delta_x^2+\delta_y^2+c_2)} \quad (1)$$

where μ_x, μ_y are the average, δ_x, δ_y are the variance and δ_{xy} is the covariance of image x and y. c_1, c_2 are constants to avoid instability when the denominator is close to zero.

Figure 3

Sample cattle face dataset.



(a) Cattle face detection dataset

(b) Cattle face recognition dataset

Two datasets need to be created for this experiment. All images in the cattle face detection dataset were manually annotated for accuracy. Using LabelImg, an open source tool, single frame images extracted from the video were labeled and the annotation format was saved as an XML file that corresponds to the standard format of the Pascal Visual Object Classification dataset. Information such as the sequence number of each cattle and the location of the rectangular frame was recorded in these annotation files. Then, data enhancement techniques such as Gaussian noise, horizontal flipping, and random cropping were applied to expand the cattle face detection data for some of the cattle with insufficient data, and a total of 12,156 images were obtained after the above. Figure 3(a) presents some of the raw data of cattle face detection and the corresponding labelled data. On this basis, a cattle face recognition dataset was established, which firstly required manual operation to classify individual cattle faces detected by the cattle face detection model. In order to ensure the accuracy rate of the ensuing cattle face recognition, the cattle face images in each folder were manually screened to make them conform to the requirements of the cattle face recognition dataset. During the screening process, it was necessary to eliminate other cattle faces that had blurring, false detections, or did not belong to the category to ensure the accuracy and reliability of the results. The above measures were taken in order to ensure that the cattle face images in each folder are presented in high resolution. A sample cattle face recognition dataset is shown in Figure 3(b).

2.4 Cattle Face Detection

2.4.1 Introduction to YOLOv7 Algorithm

The YOLOv7 network model consists of four main parts, which are Input, Backbone, Neck and Head.

The Backbone network module consists of multiple convolutional layers, including three convolutional layers, BConv, Extended-ELAN (E-ELAN) and MP-Conv. Each convolutional layer has an independent deep neural network structure and achieves classification and identification of input data by connecting weights to differentiate the type of network applications and predict future task states.The BConv convolutional layer is a complex structure composed of convolutional layers, a BatchNormalization (BN) layer, and a LeakyReLU activation function [9], which serves to extract image features at different scales, thus achieving comprehensive analysis and recognition of images; E-ELAN's convolutional laWyer continues to extend the network architecture of ELAN18, after the manipulation of the longest and shortest gradient paths in this structure, so that the feature extraction network can learn a greater variety of features, to achieve the purpose of improving the network's learning ability; by adding the Maxpool layer on top of BConv layer to form MPConv, we have achieved the goal of improving the learning ability of the network; and by adding the Maxpool layer to BConv layer to form MPConv. By adding a Maxpool layer to the BConv layer to form the MPConv convolutional layer, the structure forms two branches, the upper branch uses Maxpool to halve the length and width of the image, while the BConv layer halves the length and width of the image channel as well. In order to avoid the shortcomings present in the traditional algorithm, a new parameter, maximum entropy, is introduced to determine the size of each neuron's contribution to the data information. By applying the first layer of BConv processing to the image channel, the lower branch achieves the effect of halving the length of the channel. This allows two parts of the training data with the same structure and parameters to be jointly trained, resulting in a better model. By applying the Cat operation, the features extracted from the two branches are fused in order to further optimise the effect of the convolutional layer, thus improving the feature extraction capability of the network [14].

In order to effectively fuse the extracted features at different levels, YOLOv7 adopts the Path Aggregation Feature Pyramid Network (PAFPN) structure in the Head network, which goes through the bottom-up path, so that the information at the higher level can better receive the information at the bottom level to achieve the purpose. The Prediction module uses the REP (RepVGG Block) structure [4] to adjust the number of image channels for three different scales of features such as P3, P4, and P5 output from PAFPN, and finally the 1×1 convolution is used for the prediction of confidence, category, and anchor frame.

2.4.2 Introduction to Other Target Detection Algorithms

Among the target detection algorithms, in addition to the YOLOv7 algorithm mentioned above, this paper briefly introduced two common target detection algorithms, Faster R-CNN (Region-based Convolutional Neural Network) and SSD (Single Shot MultiBox Detector), for subsequent comparison of cow face detection experiments.

1 **Faster R-CNN**

Faster R-CNN is a deep learning algorithm for object detection. It extracts image features by convolutional neural network and generates candidate object regions using Region Proposal Network (RPN). Then, these candidate regions are corresponded to the feature map to extract the region features. Next, target classification and bounding box regression are performed for each candidate region by means of a fully connected layer. The training process uses a multi-task loss function, including classification loss and bounding box regression loss. Finally, the detection results are filtered using non-maximal suppression (NMS) to output the final target object and its accurate bounding box. Faster R-CNN improves the detection speed and accuracy by introducing RPN and end-to-end training.

2 **SSD**

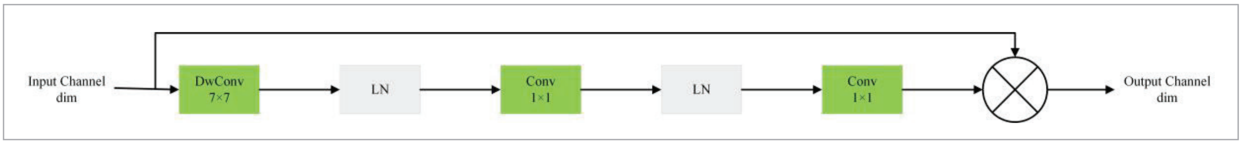
SSD is a deep learning algorithm for target detection. Unlike traditional two-stage detection algorithms, SSD employs a single-stage detection framework, resulting in faster and more accurate detection.SSD's algorithm performs multilevel feature extraction through convolutional neural networks and uses densely sampled anchor frames to predict the location and class of the target in the image. By using multi-layer feature fusion, SSD is able to perform target detection at different scales. It uses a multi-task loss function to train the network, including classification loss and bounding box regression loss.The advantages of SSD are its efficient speed and accurate performance, which makes it widely used for target detection tasks in real-time applications and mobile devices.

2.4.3 Introduction to ConvNeXt

In 2022, the ConvNeXt structure was first proposed, a network that incorporates the main improvements

Figure 4

ConvNeXt network structure.



that have been successfully integrated in convolutional neural networks in recent years. Figure 4 illustrates the structure of the ConvNeXt network, where the input data to the convolutional block is a feature map of the channel number dimension (dim). Firstly, the number of channels is achieved by a deep convolutional layer with a convolutional block size of 7×7, followed by normalisation at the Layer Normalization (LN) layer, and finally by a convolutional layer with a convolutional block size of 1×1, where the number of channels is quadrupled from the original number of input channels. After the above operations, then the nonlinear characteristics are given by the Gaussian Error Linear Unit (GELU) activation function and then the output channels are restored to the original dimensions by a convolutional block size of 1×1. Finally, the input and output data are integrated by residual joining to obtain the final output.

In the application of deep convolutional neural networks, the semantic information of an image is significantly enhanced, however, the precise features are on the trend of weakening as the process of image convolution continues. On the contrary, shallow images have a strong representation of details, but a biased representation of semantic information. In order to make the network model more optimised with less network structure, the ELAN layer in the BackBone network is replaced with a ConvNeXt layer.

2.4.4 Introduction to CoTNet Transformer

The rise of Transformer technology has significantly advanced the field of natural language processing, and the network architecture it employs has recently created a buzz in the field of computer vision with highly competitive results. Nevertheless, existing frameworks implement self-attention and generate attention matrices directly on 2D feature maps, but are under-exploited for the rich contextual information around them.

The CoT module is an attention mechanism that can fully utilise the contextual information of 2D feature graphs, and its structure is shown in Figure 5.

From Figure 5 [12] it can be seen that assuming the input 2D feature $X \epsilon R^{H\times W\times C}$, and the key vector K, the input feature Q and the value vector V are defined as:

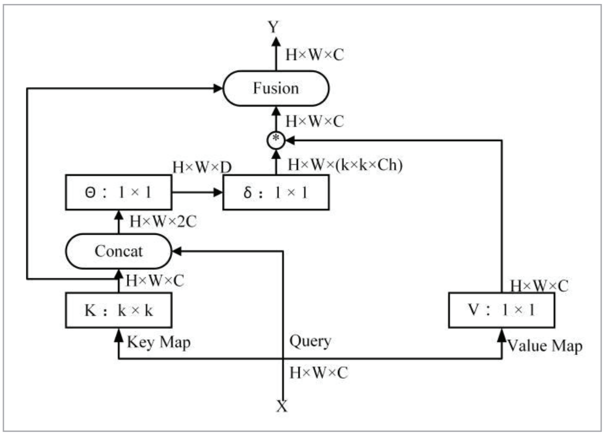
$$K = X, Q = X, V = XW_v \ , \qquad (2)$$

where: W_v means that X has been mapped with features to obtain a new V. The CoT module first uses the $K\times K$ convolution method to extract contextual information, and the resulting $K^1 \epsilon R^{H\times W\times C}$ is naturally able to represent the contextual information between the immediate neighbours, which we regard as the static contextual representation of the input X. On this basis, the text features in this dynamic contextual representation are further analysed using an attention model based on deep learning algorithms. The previously mentioned K^1 and Q are combined and then implemented by two consecutive 1×1 convolutions W_θ and W_δ; to perform the computation and determine the attention matrix A:

$$A = \left[K^1, Q\right] W_\theta W_\delta \, . \qquad (3)$$

Figure 5

CoT module.



Next, based on the attention matrix A and the value vector V obtained above, we generate the augmented feature K^2: the

$$K^2 = V * A. \tag{4}$$

The above obtained enhancement feature K^2 can capture the dynamic feature interaction about the input, which we call the dynamic contextual representation of the input X. The output of the final CoT module can be realised by fusing the above two contextual representations with the attention mechanism.

By integrating contextual information mining and self-attention learning into a unified framework, the CoT module is able to effectively mine contextual information in the neighbouring environments, thus improving the efficiency of self-attention learning and further enhancing the output features and visual representations.In this paper, the module is applied several times in the Head network with the improved YOLOv7 model, which helps to globally extract the image depth features and enhance the focus information, so that the global features of the cattle face can be enhanced, and the network structure can be further reduced.

2.4.5 Introduction to SimAM Attention Mechanisms

The attention mechanism allows the neural network to pay more attention to the important information needed for the cattle face detection task, ignoring non-important information such as background, thus improving the model's performance. Some of the existing attention modules extract features in only one of the spatial or channel dimensions, but lack flexibility in simultaneously changing spaces and channels [22]. Inspired by the attention mechanism of YANG et al. [35], SimAM Attention Module proposes a neuroscience theory-based energy function design method for building 3D attention modules that calculates the weights of attention by studying the importance of each neuron. To further improve the efficiency of the algorithm, a multi-scale convolutional neural network is also used for feature extraction.SimAM searches for the linear separability of the target neuron and other neurons by modelling an energy function. The energy function et [21] is defined as:

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1}\sum_{i=1}^{M-1}\left[-1-(w_t x_i + b_t)\right]^2 + [1-(w_t t + b_t)]^2 + \lambda w_t^2 \tag{5}$$

where: t, x_i – Target neurons and other neurons for input feature X

i – Sole guidance in the spatial dimension

M – The number of all neurons on a given channel

y – Tagged value indicating whether or not it is a significant neuron

w_t, b_t – Weighting and biasing

λ – Regularisation coefficient.

The minimum energy function is calculated. The lower the energy, the more t neurons are distinguished from other neurons and the more important they are. Compared with existing mechanisms such as Squeeze-and-Excitation (SE) [7], Convolutional Block Attention Module (CBAM) [28] and ECA [22], which generate attention weights through additional self-networks, SimAM derives 3D attention weights from input feature maps, giving greater weights to important neurons. In this study, it is embedded into the Neck of YOLOv7 to find important neurons while suppressing peripheral neurons and enhancing relevant features to focus on feature representations more favourable for cattle target detection.

2.4.6 Cattle Face Detection Network Model

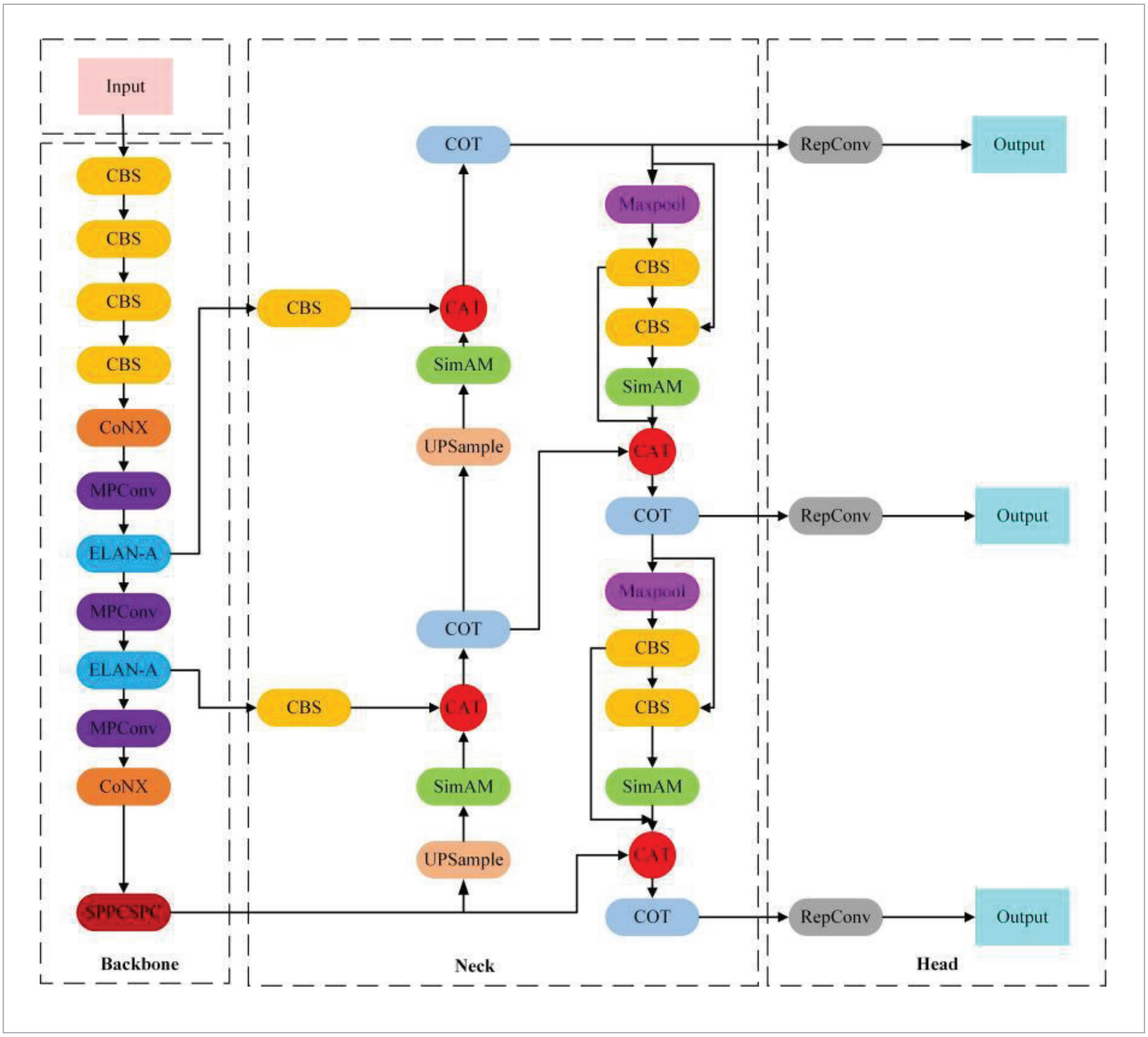
In summary, based on the YOLOv7 target detection model [20], the following changes were made to better adapt to the task of cattle face detection:

1. The ELAN-A module in the Backbone part and the ELAN-H module in the Neck part were replaced with the ConvNeXt network structure and the COT module of the CoTNet Transformer, respectively. This was done to enhance the extraction and expression of the bull's face features on one hand and to optimize the network structure and improve the speed of cattle face detection on the other.
2. The SimAM attention mechanism was added to the Neck part to improve the expression of effective features of the cattle face.

The framework diagram of the cattle face detection model was shown in Figure 6.

Figure 6

Framework diagram of cattle face detection network model.



2.4.7 Test Setup for Cattle Face Detection

Adopt the cattle face detection network model built based on the above and set the relevant parameters for the dataset used in this trial: (1) recalculate the anchor frame size using the K-means++ clustering method, adopt the 3-channel processing strategy, and input the detection frame size of 640 pixels × 640 pixels; (2) the weight decay regularity coefficient of 1×10^{-6}, adjust the learning rate of 1×10^{-3}, and take the learning rate of 10 times decay for model iteration; (3) the number of iteration rounds is set to 100; (4) the loss function is adopted as CIOU, and non-maximal value suppression is carried out.

2.4.8 Evaluation Indexes of Cattle Face Detection

Precision(P), Recall(R), mean average precision (mAP), mean frame rate (Frame Per Second, FPS) and model memory usage are selected as evaluation metrics to measure the model's performance. Where FPS is the number of frames transmitted per second and refers to the number of image frames detected per second.

2.5 Cattle Face Recognition

2.5.1 Introduction to the VoVNet Network Architecture

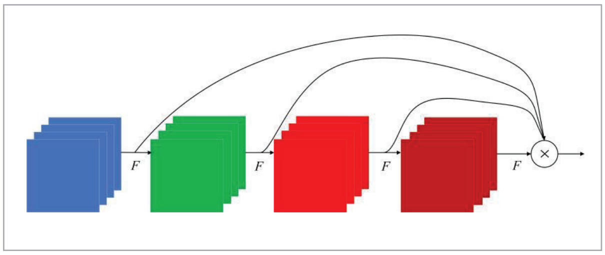
The number of parameters and computational complexity are very critical elements in the process of building a lightweight network, however, reducing the size of the model and the computational complexity does not mean that the inference time and energy consumption can be reduced. For this reason, a lightweight network structure design method based on multi-core parallel algorithms is proposed. The VoVNet network is designed with a comprehensive trade-off between the memory access cost (MAC) and the computational power of GPUs, aiming to achieve more efficient data transmission. Through experimental tests on different datasets, the algorithm was found to be effective in reducing the storage space and energy consumption required for computation. The following is the calculation method for convolutional layer MAC:

$$MAC = hw(c_i + c_o) + k^2 c_i c_o . \tag{6}$$

The computational requirement of the convolutional layer is $F = k^2 hwc_i c_o$, where k represents the size of the convolutional kernel, h, w is the height and width of the features, and c_i, c_o is the number of input and output channels. The computational cost can be reduced by optimising the network structure. If the computational volume F is kept constant, then the $MAC = F(c_i + c_o) / k^2 c_i c_o + Fhw$ will be calculated based on the mean value inequality $c_i^2 + c_o^2 \geq 2c_i c_o$. The $MAC \geq 2\sqrt{hwF / k^2} + F / hw$ will choose the minimum value only if $c_i = c_o$, that is, the number of input and output channels are equal, so that the network can reach the most efficient state. The superior performance of GPUs is reflected in their efficient parallel computing mechanism, and the computational potential of GPUs can be maximised only when the computational dimensions are relatively large. is maximised only when the computational dimensions are relatively large. Therefore, in order to increase the speed of computation, parallelisation techniques can be used to accelerate the process of convolutional layer computation. Although splitting a larger convolutional layer into multiple smaller convolutional layers can achieve similar results, the computational efficiency of GPUs fails to meet expectations. This is because when the objective function is nonlinear, traditional linear processing methods are no longer applicable. Therefore, it is necessary to do a lot of preliminary preparation before starting large-scale computation. More interest is placed on the computational requirements per second than the amount of computation. This entails dividing the overall computation volume by the total GPU inference time to increase the efficiency of GPU usage as the performance metrics improve.

Figure 7

Structure of OSA module.



VoVNet is composed of OSA modules, and the structure of OSA modules is shown in Figure 7 [11].

As can be seen in Figure 7, VoVNet is composed of OSA modules. Different from DenseNet, the OSA module only features all layers before the last aggregation, while each convolutional layer has two different connections. The experimental results show that the algorithm in this paper can effectively reduce the network training time and reduce the number of network trainings while guaranteeing a higher recognition accuracy. (1) directly connect it to the next convolutional layer, thus creating features that can produce a larger sensory field; (2) connect the feature map to the output position of the last layer. Thus, this approach effectively solves the redundancy problem of DenseNet features. With the integration strategy of the OSA module, that the number of input and output channels in each layer was fixed, thus achieving the smallest MAC value. In addition, the utilization of 1×1 convolution to compress the features was circumvented, resulting in a notable enhancement in the GPU computational efficiency of the OSA module. After the optimisation of the connection method, VoVNet enhances its feature representation and extraction, which greatly improves the detection accuracy of the model.

2.5.2 Introduction of the ECA Attention Mechanism

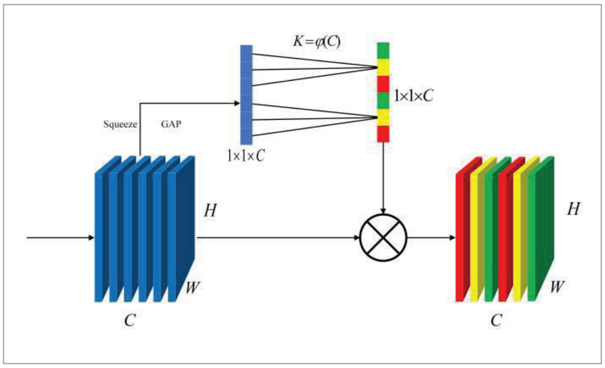
As shown in Figure 8 [19], the composition and structure of the ECA mechanism can be expressed as follows: H represents the height of the input image, W represents the width of the input image, and C represents the number of channels of the feature map. In order to avoid the dimensionality shrinkage of the SE attention mechanism, a global average pooling operation is used after compressing the feature maps of each channel. Then, two fully connected layers are replaced by an adaptive one-dimensional convolutional kernel size K to determine the coverage for cross-channel information exchange. Finally, recalibration of the feature map is accomplished by multiplying this result with the uncompressed original feature map. This approach enables the network to selectively emphasise important features while suppressing the influence of useless features. The one-dimensional convolutional kernel size K is positively correlated with the number of channels C, satisfying the:

$$K = \varphi(C) = \left| \frac{lbC}{\Upsilon} + \frac{b}{\Upsilon} \right|_{odd}. \tag{7}$$

Following the experience in the literature [22], the parameters are mapped as $\Upsilon = 2$ and $b = 1$. For this network model, a local cross-channel interaction strategy without dimensionality reduction can significantly reduce the complexity of the network model with only a small number of parameter additions, while achieving a significant performance improvement. This strategy can be regarded as a lightweight channel attention mechanism.

Figure 8

Structure of the ECA attention mechanism.



The ECA mechanism uses one-dimensional convolution, which can effectively prevent the side effects of feature reduction in the fully connected layer, and by interacting with the information of neighbouring channels to adapt to the convolution kernel size K, it can effectively obtain the small targets that are easy to be ignored and missed in the cattle face image. At the same time, by interacting the information across channels, the redundant information of non-target features in the cattle face image can be avoided without significantly increasing the memory overhead and network depth, thus effectively maintaining the learning results of important features. Combining the ECA mechanism with the VOVNet network model can effectively improve the performance of cattle face recognition.

2.5.3 Cattle Face Recognition Network Model

After the above formulation, the following optimizations were made to the cattle face recognition network:

1 Residual connections were added from the input to the output of the OSA modules of the VOVNet network to better extract deep features of the cattle face for more complex network structures.

2 The ECA mechanism was added to the last feature extraction layer of the OSA modules, so that the network could better learn the effective features of cattle face recognition and express them to improve the recognition performance of the network.

The framework diagram of the cattle face recognition model was shown in Figure 9 [11].

2.5.4 Test setup for Cattle Face Recognition

1 **Image pre-processing**

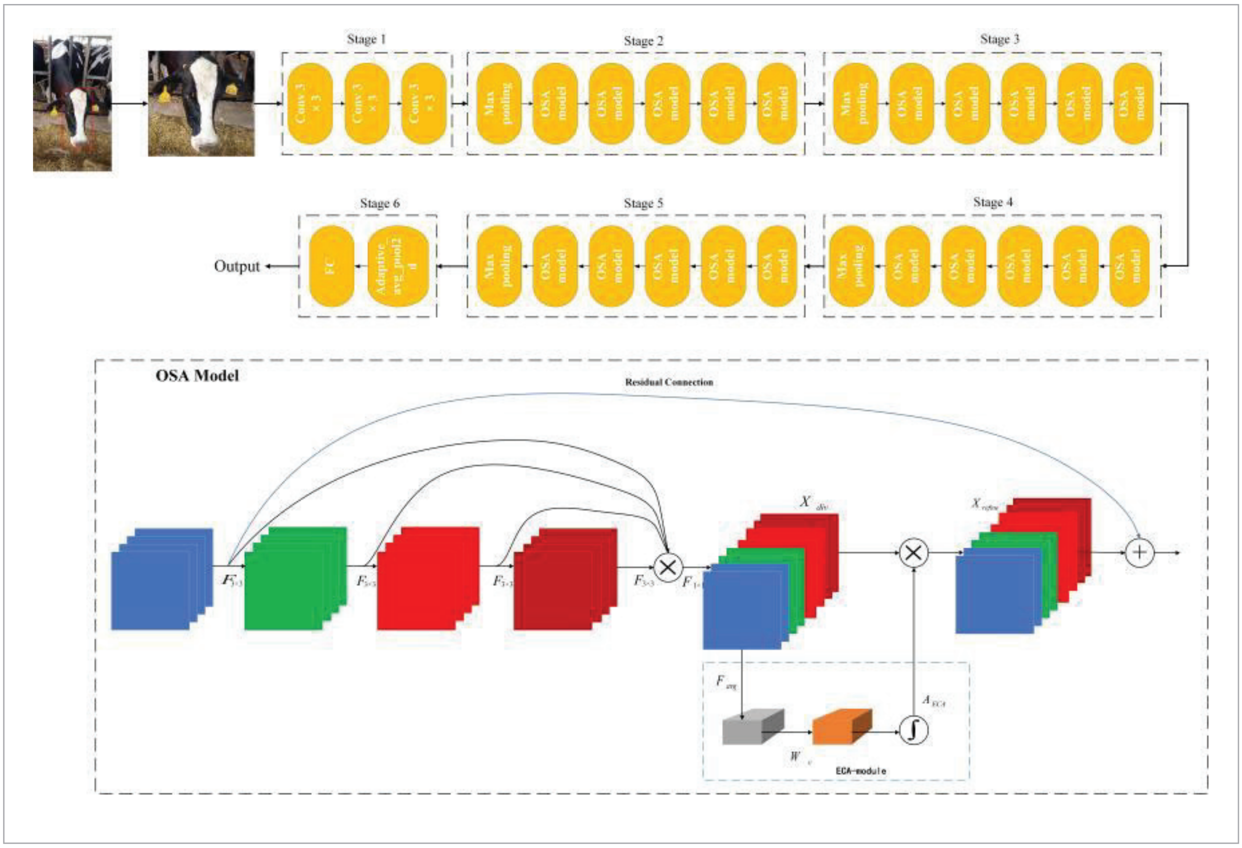
Before training the cattle face recognition model, it was necessary to pre-process the input images appropriately. With the mean and variance calculated on the ImageNet dataset, the image could be normalized according to the channel, thus increasing the convergence speed of the model.

2 **Data enhancement**

In order to expand the number of training images and enhance the robustness of the network with certain transformations, the training images were randomly manipulated, resized to 224 pixels × 224 pixels, and the data format was converted to tensor format and normalized.

Figure 9

Framework diagram of the cattle face recognition network model.



3 **Parameter setting**

The Center Loss loss function is used in the experiment, and the dynamic initial learning rate is set to $1×10^{-3}$. In order to prevent the training model from overfitting, the dropout is set to 0.9, the learning momentum is $5×10^{-4}$, the iteration period is 100, and the batch size is 8.

2.5.5 Evaluation Indexes of Cattle Face Recognition Model

Precision (P), Recall (R), Accuracy (A), mean frame rate (Frame Per Second, FPS), Kappa coefficient, model size, F1-score, Parameters, floating point operations per second (FLOPS), Mean pixel Accuracy (MPA) are selected as evaluation metrics to measure the model performance. Among them, Kappa coefficient a multivariate statistical method to evaluate classification accuracy, can reflect the degree of consistency between classification results and actual categories.

3. Experimentation and Result Analysis

3.1 Cattle Face Detection

3.1.1 Comparative Analysis of the Effect of Different Improvement Strategy Models

Since the cattle face detection network was based on YOLOv7 network, integrating ConvNext network, SimAM attention mechanism and COT module, in order to verify the effect of the improvement strategy, the performance of the cattle face detection network before and after the improvement was compared and analysed. Table 1 shows the comparison of the evaluation metrics of the models with different improvement strategies. The initial YOLOv7 algorithm was used as the initial model, denoted as Model A; Model B referred to the replacement of the ELAN module

Table 1

Comparison of evaluation indicators of different improvement strategy models.

Models	Backbone networks	Attention mechanism	COT module	Precision (P/%)	Recall (R/%)	FPS ($f\cdot s^{-1}$)
A	ELAN	–	–	98.31	97.97	68.03
B	ConvNext	–	–	95.28	95.31	80.09
C	ConvNext	–	√	96.75	96.80	83.33
D	ConvNext	√	√	99.43	99.10	96.15

in the backbone network with ConvNext; Model C used the COT module to replace the ELAN module in the neck in Model B; and Model D added SimAM in the neck part on the basis of Model C, which was the final version of the cattle face detection network.

The results show that:

1. For the same task network, after the ELAN module in the YOLOv7 backbone network was replaced by ConvNext, the model parameters were reduced, and the detection speed was improved, but there was a loss in precision. From Table 1, it can be seen that the original YOLOv7 had excellent detection precision in this experiment, with values of P and R reaching 99.31% and 98.11%, respectively. However, the overall network model size of YOLOv7 measured 142MB, as shown in Table 2. The overall parameters of this model were too large, resulting in slightly poorer real-time performance and portability. After replacing the ELAN module in the YOLOv7 backbone network with ConvNext, the detection speed of Model B increased to 80.09 frames/second, which was 12.06 frames/second higher than the original YOLOv7. In terms of detection accuracy, the improved backbone network decreased. Model B′s precision decreased by 3.03%. This was because replacing the ELAN layer in the backbone network with the ConvNext network reduces the network depth, which may lead to relatively weaker computational capability [37] and less effective extraction of corresponding input features. As a result, the detection speed in the model increased while the precision decreased accordingly. Moreover, when the ELAN module in the feature extraction network was replaced by the COT module, Model C had decreased precision but the highest detection speed. Although the speed was greatly improved, the semantical information of the features extracted by the network, such as shallow texture and contour information, was not expressed better, which is crucial for cattle face recognition [39]. This model could be considered integrating attention information into the Head section to ensure precision.

Table 2

Comparison of network structure parameters.

Network models	Network parameter	Model size/MB
YOLOv7	39447429	142
CCS-YOLOv7	32543596	63.7

2. Adding the attention mechanism greatly enhanced the effect of cattle face detection. After incorporating the SimAM attention mechanism into the location of the feature extraction network of model D, the P-value and R-value of model D were improved to 99.43% and 99.10%, respectively, which represented an increase of 2.68% and 2.3% compared to model C. This improvement was possible because the SimAM attention mechanism could be obtained 3D attention weights without the need for additional parameters. It selectively strengthened the effective information by assigning higher weights to the channels containing cattle faces. Furthermore, different features could be assigned to different channels of the feature map.

Through the above analysis, it was proved that the network optimises YOLOv7 to the maximum extent without loss of precision, the number of parameters was greatly reduced, and the speed of model detection was linearly improved. The model can be used for the facial detection of cattle in cattle farm environments, providing technical support for subsequent cattle face recognition and intelligent cattle management.

3.1.2 Comparative Analysis of Experimental Results of Different Models

On the cattle face detection test set, the results of SSD, Faster R-CNN, YOLOv7, and the improved YOLOv7 cattle face detection model were shown in Table 3. It could be observed that the improved cattle face detection model, CCS-YOLOv7, outperformed SSD, Faster R-CNN, and YOLOv7 in cattle face detection. The CCS-YOLOv7 model achieved a precision of 99.43%, a recall of 99.10%, a mAP of 99.80%, and an FPS of 96.15 frames per second. Compared to SSD, Faster R-CNN, and YOLOv7 models, the mAP improved by 5.8, 10.41, and 0.2 percentage points, respectively. Although there was some improvement in precision, recall, and mAP compared to the original model, the difference was not significant. However, the model size was greatly reduced, and the detection speed was improved. Thus, the replacement of the backbone network and the addition of attention mechanisms in the ELAN module and feature extraction layer of the Head network in this study effectively improved the precision and speed of cattle face detection, validating the effectiveness of the proposed method and laying the foundation for subsequent cattle face recognition.

Table 3

Cattle face detection results of different models.

Models	Model size/MB	FPS/(f·s^{-1})	Precision(P/%)	Recall(R/%)	mAP/%
SSD	90.6	62.08	95.40	92.20	94.00
Faster R-CNN	108	6.42	86.30	89.87	89.39
YOLO v7	142	68.03	98.31	97.97	99.60
CCS-YOLOv7	63.7	96.15	99.43	99.10	99.80

3.1.3 Comparative Analysis of Cattle Face Detection Results

In order to more intuitively reflect the superiority of the improved YOLOv7 cattle face detection model proposed in this paper, experiments based on the cattle face detection dataset were conducted to make the comparison in Figure 10.

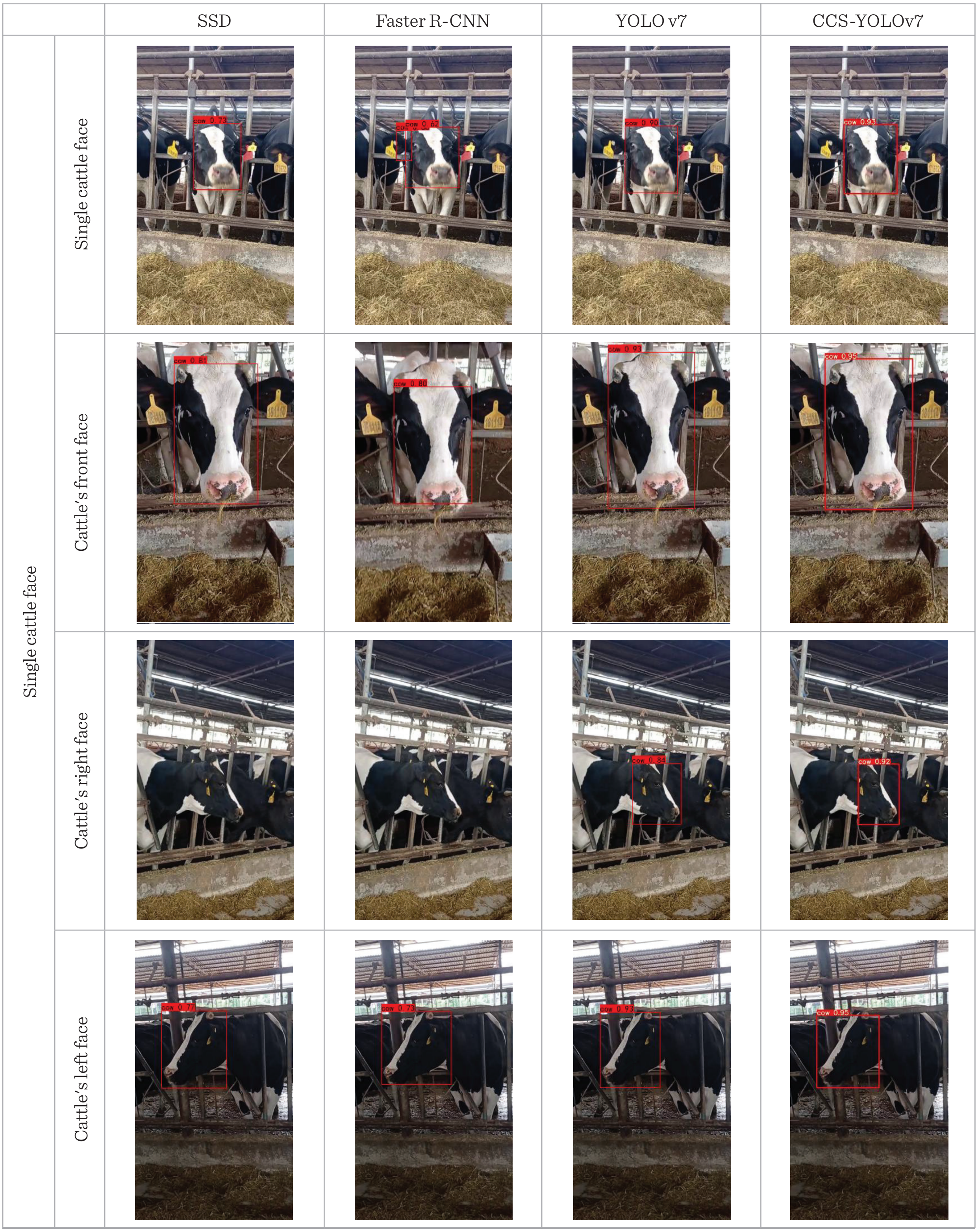
In the case of a single cattle face, due to the influence of the image background, other cattle face detection models encountered some issues, such as missed detection, target detection frame offset, and mistakenly detecting the cattle ear as a cattle face. On the other hand, the CCS-YOLOv7 model not only experienced no leakage detection or wrong detection but also maintained a closely fitting target detection frame with the cattle face, while its confidence level was also improved to some extent. This indicates that the improved CCS-YOLOv7 model demonstrated excellent robustness and accuracy even under the influence of image background.

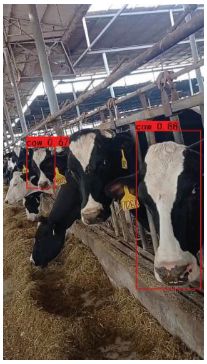
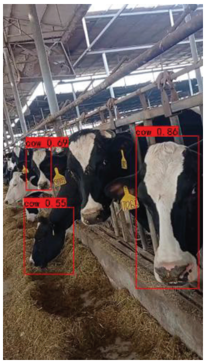
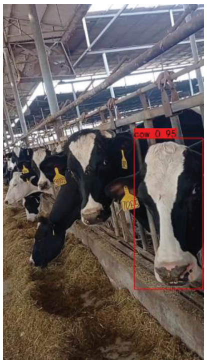
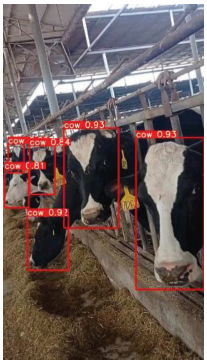
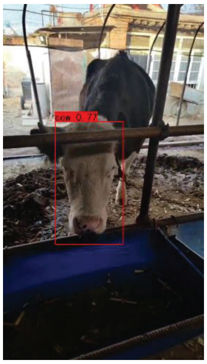
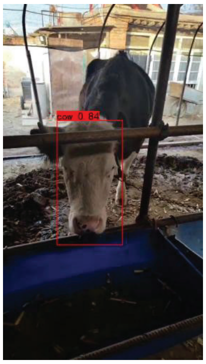
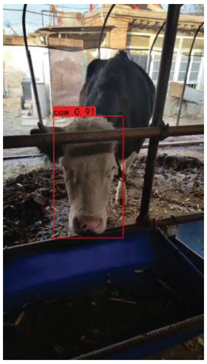
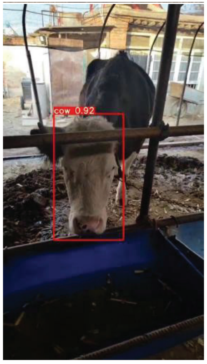
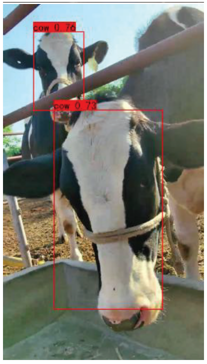
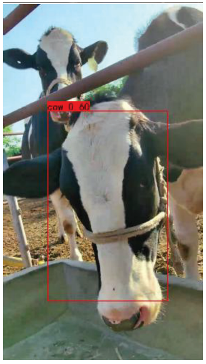
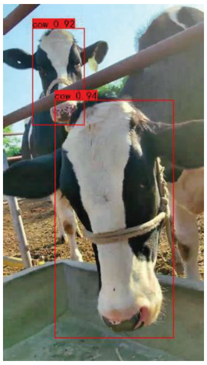
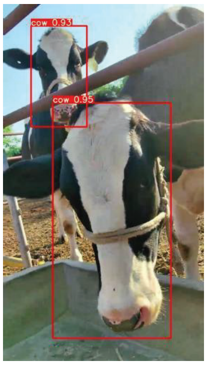
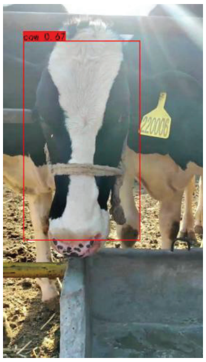
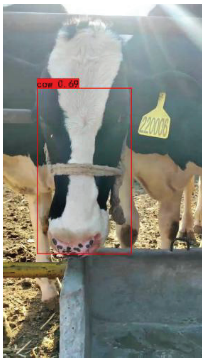
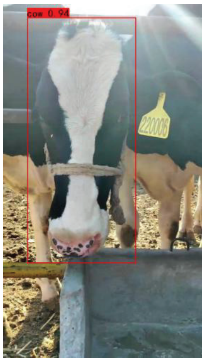
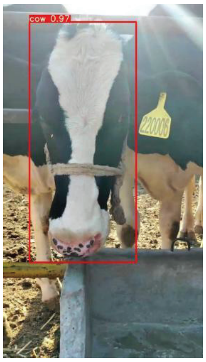
In the scene of multiple cattle faces, other cattle face detection models had the problem of missed detection in the case of dense cattle faces and the presence of blurred cattle faces. In contrast, the CCS-YOLOv7 model detected all the cattle faces in the image, and the confidence of target detection was improved. This also indicated that the improved CCS-YOLOv7 model had improved the learning of cattle face location information and the application of deep and shallow target features. The model was still able to accurately detect and maintain high accuracy even in the case of occluded or blurred cattle faces.

Under varying lighting conditions, other cattle face detection models exhibited missed detections and target detection box misalignment when detecting cattle faces in situations with challenging lighting and obstructions. In contrast, the CCS-YOLOv7 model could be detected cattle faces under different lighting conditions and even when there were obstructions. This effectively reduced the model's missed detection rate, and there was a certain improvement in the classification confidence. This indicates that the improved model had strong generalization capabilities under different lighting and obstruction scenarios.

Figure 10

Comparison of different models of detection.



		SSD	Faster R-CNN	YOLO v7	CCS-YOLOv7
Multiple cattle face	Multiple cattle face				 cow 0.93 cow 0.93 cow cow 0.84 cow 0.81 cow 0.97
Variable light	Dark and sheltered				
	Natural light and multiple				
	Glare				

3.1.4 Analysis of Experimental Results on Public Cattle Face Dataset

In order to further the generalisability of the model, this paper used the cattle face dataset made public by Northwestern University for the experiments. The dataset contains a total of 1866 images, and 606 cattle face images were obtained by eliminating those whose pixels are too blurred, do not belong to and do not contain cattle faces. Next, the cattle face detection dataset was expanded using data enhancement techniques such as horizontal flipping, and the final dataset contained 4848 cattle face images.

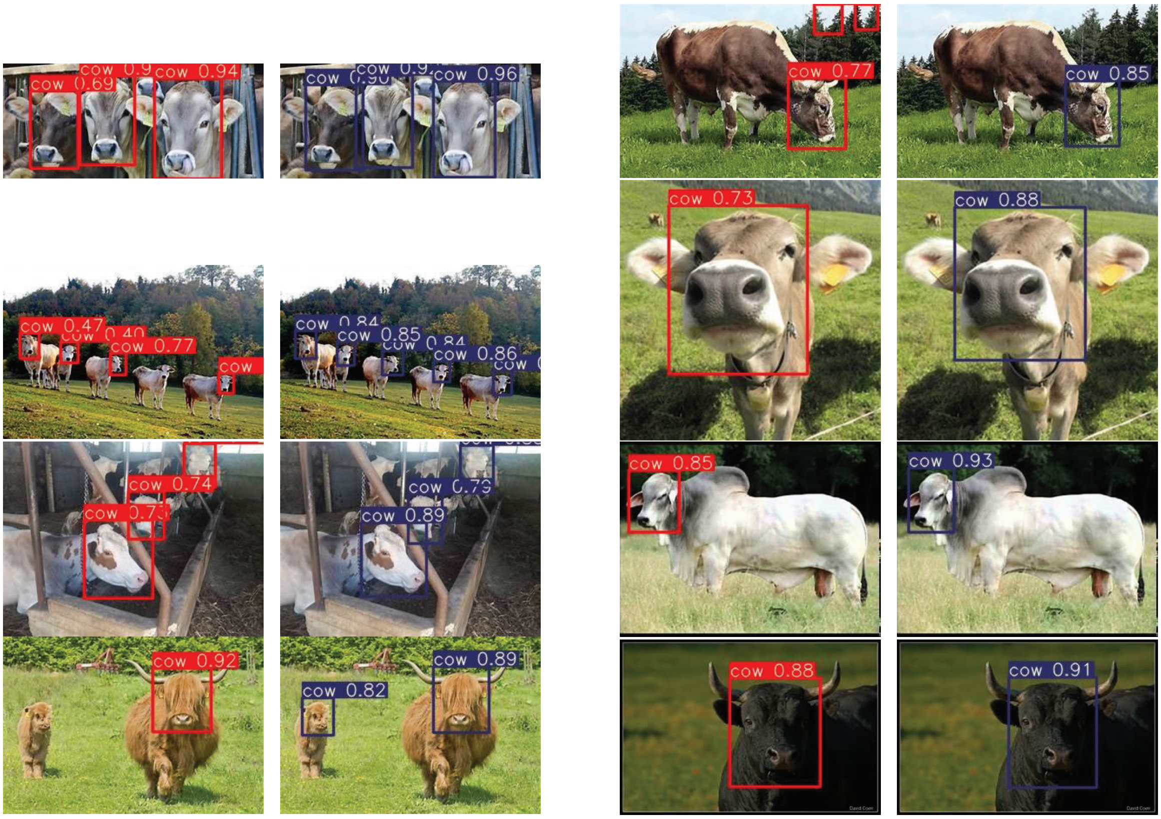
Figure 11 showed the results of the comparison between CCS-YOLOv7 and YOLOv7 cow face detection under the public dataset. From Figure 11(a), in the case of multiple cow faces, the YOLOv7 model showed a missed detection; from Figure 11(b), in the case of a single cow face, the YOLOv7 model showed a misdetection, and its confidence was significantly lower than that of the CCS-YOLOv7 model in both cases.

Taken together, Figure 11(a) and Figure 11(b) provided good detection results in the case of close-up, distant view, different cow face angles, and different lighting. In addition to this, the CCS-YOLOv7 model was better than the YOLOv7 model for face detection of different breed types of cattle.

As Table 4 showed the experimental results under the open source dataset, the improved CCS-YOLOv7 improved P by 5.32 percentage points, R by 4.51 percentage points and mAP by 3.85 percentage points compared to YOLO v7. The experimental results indicated that CCS-YOLOv7 could be used for face detection work in different cattle face angles, lighting, and multi-breed cattle.

Figure 11

Comparison of cattle face detection under public dataset(the red is YOLOv7 and the blue is CCS-YOLOv7).



(a) Comparison of multiple cattle face detection

(b) Comparison of single cattle face detection

Table 4

Comparison of network structure parameters.

Models	Precision (P/%)	Recall (R/%)	mAP /%
YOLOv7	93.24	93.68	95.38
CCS-YOLOv7	98.56	98.19	99.23

3.2 Cattle Face Recognition

3.2.1 Ablation Experiments

Model ablation experiments were performed as shown in Table 5. The second set of experiments was based on the first set of experiments using VoVNet as the backbone network, and by adding residual connections between the inputs and outputs, the OSA module had the ability to pass the gradients of each stage in reverse, thus allowing the model to better adapt to more complex network structures. Given that the depth of the VoVNet network in this experiment had not yet been reached, there was no significant difference between FPS and accuracy. The third set of experiments involved the addition of only the ECA attention mechanism, with an accuracy rate of 97.941%, which was an improvement of 1.294%, indicating that the ECA attention mechanism was able to better learn the cattle face features and improve the accuracy rate of cattle face recognition. The fourth set of experiments constituted the final cattle face recognition model, which had an improved accuracy rate compared to the state where the detection speed was almost unaffected. Therefore, the improvement of the cattle face recognition model was considered effective and reasonable.

Table 5

Comparison of ablation experiments of cattle face recognition models.

Group	VoVNet	Residual connection	ECA	FPS/($f{\cdot}s^{-1}$)	Accuracy(A/%)
1	√	-	-	92.53	96.647
2	√	√	-	92.79	96.823
3	√	-	√	90.39	97.941
4	√	√	√	92.72	99.470

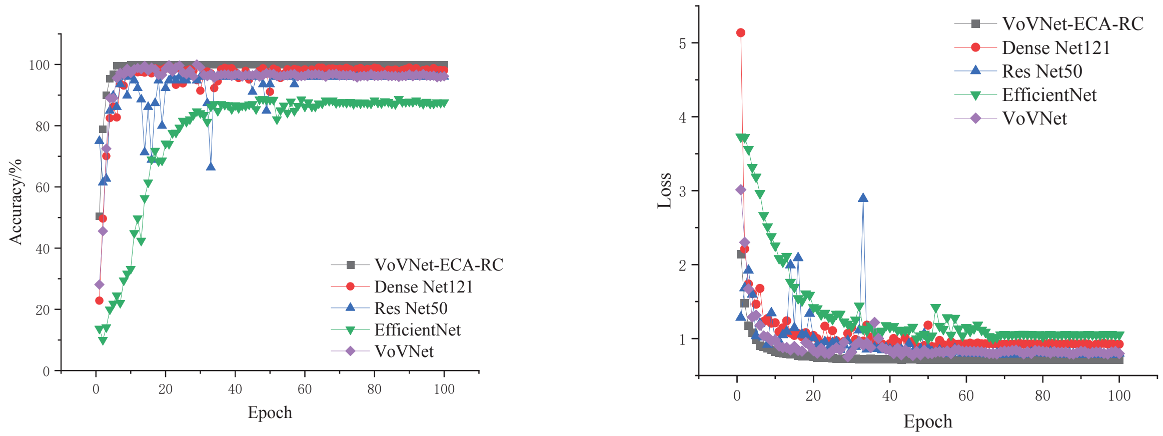
3.2.2 Comparative Analysis of Training Process

In order to verify the performance of the proposed VoVNet-ECA-RC cattle face recognition model, VoVNet-ECA-RC, VoVNet, Dense Net121, Res Net50 and Efficiennet models were trained under the same dataset and test environment. The curves of the accuracy and loss values of the validation set during the training of each model are shown in Figure 12.

As seen from the changes in the graph lines in Figure 12, with the gradual increase in the number of iterations, the accuracy of the above five models on the validation set showed an increasing trend, while their loss values showed a decreasing trend, and finally underwent an oscillation and reached a stable state within a certain range. The Figure 12 showed that VoVNet-ECA-RC exhibited higher accuracy and faster convergence on the validation set compared to other models. When the models entered the convergence stage, the accuracy rate of the VoVNet-ECA-RC model fluctuated in the range of 99.00%, while the value of the loss function fluctuated in the range of about 0.71. Meanwhile, the accuracy rate and loss value of VoVNet were around 96.00% and 0.81, respectively. It was easy to see that the accuracy rate and convergence speed of VoVNet-ECA-RC were better than the original model due to the inclusion of residual connections from input to output and the introduction of the ECA attention mechanism, which strengthened the learning of features. Moreover, after the convergence of VoVNet and ResNet50, the change curves of the accuracy rate and loss value were approximately the same, and the final accuracy rate and loss value were in the range of 96.00% and 0.81, respectively. In addition, the accuracy and loss values of the DenseNet121 model were stable in the range of 98.00% and 0.92, and the accuracy and loss values of the EfficientNet model were stable in the range of 88.00% and 1.05.

Figure 12

Variation curves of accuracy rate (a) and loss values (b) for different models.



(a) Variation curves of accuracy rate

(b) Variation curves of loss values

3.2.3 Comparative Analysis of Test Performance

The trained models were tested on the same test set and the accuracy, recall, precision, Kappa coefficient and average frame rate of the models were calculated based on the classification results, Table 6. shows the statistics of the test results of cattle face recognition on 55 cattle using the different models mentioned above.

The VoVNet-ECA-RC model for cattle face recognition proposed in this paper is based on the basic VoVNet network model with relevant improvements. The accuracy, recall, precision and Kappa coefficient of the two models can be known through the above comparison Table 5, and the evaluation indexes of the VoVNet-ECA-RC model had been improved by 3.149%, 2.824%, 2.823%, and 0.0282, respectively, in comparison with the basic VoVNet network model, which was a significant improvement. And through the data in the above table, it was easy to find that there is no significant loss in terms of detection speed of the VoVNet-ECA-RC model. Therefore, the constructed cattle face recognition model improved the evaluation indexes such as accuracy while ensuring speed. Com-

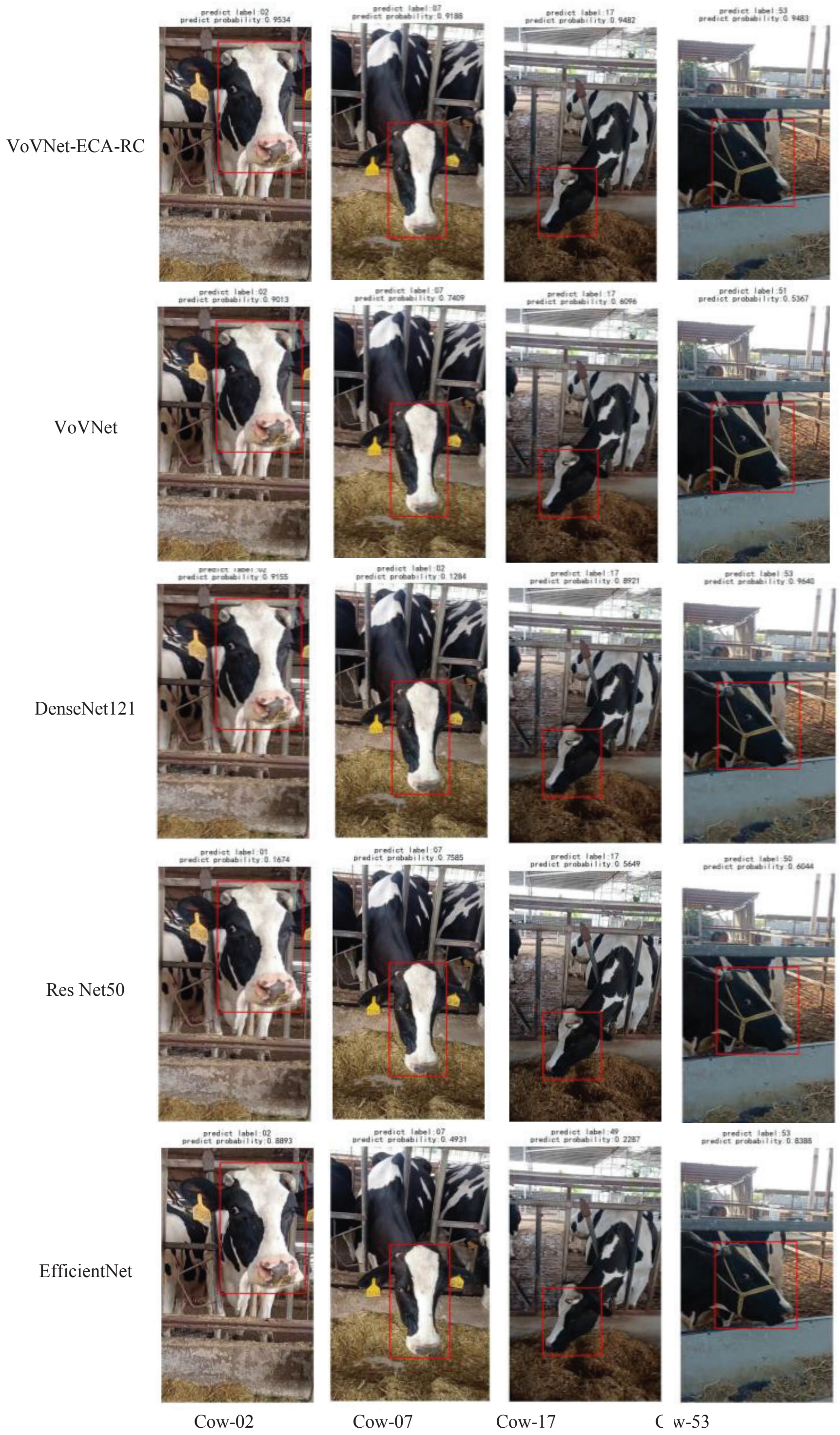
Table 6

Comparison of performance of different models on the test set.

Models	VoVNet-ECA-RC	VoVNet	DenseNet121	Res Net50	EfficientNet
Precision(P/%)	99.371	96.222	98.597	95.861	86.847
Recall(R/%)	99.003	96.179	98.114	95.491	86.307
Accuracy(A/%)	99.470	96.647	98.353	96.352	87.763
Kappa	0.9945	0.9663	0.9833	0.9629	0.8768
FPS/(f·s^{-1})	92.72	92.53	65.34	44.28	56.05
Model size/MB	41.4	41.3	51.3	45.2	39.4
F1/%	99.114	96.151	97.151	95.338	86.439
Parameters/M	21.631	21.631	26.673	23.619	20.247
FLOPS/G	7.100	7.098	2.896	4.132	5.445
MPA/%	99.003	96.179	98.114	95.491	86.307

Figure 13

Recognition results of different models.


predict label:02
predict probability:0.9534
predict label:07
predict probability:0.9188
predict label:17
predict probability:0.9482
predict label:53
predict probability:0.9483
VoVNet-ECA-RC
predict label:02
predict probability:0.9013
predict label:07
predict probability:0.7409
predict label:17
predict probability:0.6096
predict label:51
predict probability:0.5367
VoVNet
predict label:02
predict probability:0.9155
predict label:02
predict probability:0.1284
predict label:17
predict probability:0.8921
predict label:53
predict probability:0.9640
DenseNet121
predict label:01
predict probability:0.1674
predict label:07
predict probability:0.7585
predict label:17
predict probability:0.5649
predict label:50
predict probability:0.6044
Res Net50
predict label:02
predict probability:0.8893
predict label:07
predict probability:0.4931
predict label:49
predict probability:0.2287
predict label:53
predict probability:0.8388
EfficientNet
Cow-02
Cow-07
Cow-17
C w-53

pared with Dense Net121, Res Net50 and EfficientNet models, the recognition accuracy of the VoVNet-ECA-RC model for cattle face was 99.470%, which was 1.117, 3.117 and 11.707 percentage points, and its accuracy, recall and Kappa coefficient on the test set were also better than the other models, which were 99.371%, 99.003% and 0.9945, respectively, indicating that the model had better overfitting resistance and generalisation performance [31]. In addition, the VoVNet-ECA-RC model also had an advantage in cattle face detection speed, which was 92.72 frames per second.

In addition to this, the VoVNet-ECA-RC model for cattle face recognition had a model size of 41.4MB when the test values of performance metrics such as accuracy, recall and precision were kept constant or improved.Meanwhile VoVNet-ECA-RC model outperformed the other models in terms of F1-score, number of parameters, computational speed-FLOPS, and MPA while balancing accuracy and speed. The data in Table 6 better illustrated the superiority of the improved cow face recognition model VoVNet-ECA-RC.

3.2.4 Comparative Analysis of Cattle Face Recognition Results

As the Figure 13 showed the comparison of cattle face recognition results of different models. Although the above cattle face recognition models did not show any missed detection, it was not difficult to see that the VoVNet, DenseNet121, Res Net50 and EfficientNet models showed the problem of misidentification and low probability of recognition prediction dued to the similarity of hair colour and pattern of the face of some cattle, and the high degree of facial similarity, and the insufficient information about the cattle's facial features extracted by the models. When in the cattle front face recognition, the Res Net50 model misidentified No.2 as No.1, and the DenseNet121 model misidentified No.7 as No.2; when in the cattle feeding recognition, the EfficientNet model misidentified No.17 as No.49; in addition to the high similarity of the cattle's facial features, the light conditions in the actual scene are also an important factor affecting the recognition results, and the VoVNet model and Res Net50 model misidentified cow No.53 as cow No.51 and cow No.50 in the natural light condition of the outdoor shooting. The VoVNet-ECA-RC model had correct recognition results for all four cattle.

3.2.5 Comparative Analysis with Other Algorithms

In order to further illustrate the accuracy and efficiency of the above proposed cattle face recognition model, the paper selected articles related to cattle face recognition in recent years and classified the methods used in the articles into deep learning algorithms and other algorithms. In the articles, the models A-F applied deep learning algorithms, and the models G-K used other algorithms. The accuracy of the models in the article and the average inference time data of some models were displayed in Table 7. Through the data displayed in the table, it was found that the accuracy rate of model F in this article was higher than that of other models in most cases. When compared with model D, the accuracy rate of model F was lower, but the average inference time of model F was far better than that of model D, despite a small difference in the accuracy rate.

4. Conclusion

For the task of cattle face detection, this study proposed an improved YOLOv7 cattle face detection model. ConvNeXt and CoTNet Transformer modules were used to replace the ELAN structure of the Backbone network and Head network, respectively, to optimize the network structure, reduce the size of the model, and improve the network's recognition of image features. The SimAM attention mechanism was introduced into the Head network of YOLO v7, which effectively improved the expression ability of target features. Experiments on the cattle face detection dataset showed that the improved YOLOv7 cattle face detection model achieved the precision of 99.43%, a recall of 99.10%, an FPS of 96.15 frames per second, and a model size of 63.7 MB, outperforming other models in the detection of cattle faces under

Table 7

Comparison with other algorithms.

Models	Recognition Algorithm	Accuracy (A/%)	Average Reasoning Time/ms	Source
A	CNN+ResNet	95.12	-	Literature [40]
B	ResNet	94.53	-	Literature [24]
C	RetinaFace	91.30	42	Literature [33]
D	Two-Branch CNN	99.71	6820	Literature [26]
E	ResNet	98.42	-	Literature [6]
F	VOVNet	99.47	10.80	This study
G	K-SVD Algorithm	90.00	-	Literature [38]
H	Independent Component Analysis	86.95	-	Literature [10]
I	ORB	23.36	28.15	Literature [25]
J	SURF	30.65	58.01	
K	SIFT	28.58	140.32	

conditions involving multiple cattle, occlusion, and varying lighting. To further the generalisability of the model, this study used a public cattle face dataset for experiments, and the final experimental results validated the accuracy of the model and could be generalised to different breeds and types of cattle. The model could not only be used for cow cattle detection and counting, but also provided support for subsequent cattle face recognition tasks.

For the cattle face recognition task, this study proposed a cattle face recognition model with VoVNet-ECA-RC that introduced the ECA attention mechanism. Residual connections were added from the input to the output of the OSA modules, and the ECA attention mechanism was added to the final feature extraction layer of the OSA modules to improve the expression of deep features and the feature extraction capability of the network. By conducting experiments on the cattle face recognition dataset, the results showed that the performance in all aspects was improved, and the accuracy of the model reached 99.47% with a model size of 41.4 MB. The cattle face recognition experiments showed that the proposed model outperformed other models for different angles of the cattle face, such as cattle sideways, feeding, and under natural lighting. Therefore, the research structure provided a reference for non-contact individual recognition and insurance claims in cattle management in the intelligent breeding process.

Based on the above research results, the model has been well implemented in terms of accuracy and speed, therefore future work will integrate it to mobile applications to build an individual recognition system based on the facial features of a cattle face. In addition, since this research has divided the cattle face recognition into two stages, future work will also focus on further optimising the model to improve the speed and model size.

This study was based on a deep learning approach to cattle face recognition, constructed a cattle face image dataset, and constructed a cattle face detection and recognition network model to achieve individual recognition of Holstein dairy cattle and Simmental beef cattle. However, there are still some problems to be solved in cattle face recognition. Firstly, there is no public standard dataset for cattle face recognition, and the dataset in this paper needs to be further extended in terms of the number of categories and the amount of data in a single category. Secondly, this paper only investigates the individual recognition of cattle categories in the farms where the data were collected. However, the individual recognition of multi-species cattle by the algorithms in this paper needs to be further explored.

Acknowledgements

This study was supported by General Project of National Natural Science Foundation of China, Research on Model Construction and Early Warning Method of Abnormal State and Acoustic Sensing Mechanism of Pigs (32172784). The authors are grateful to anonymous reviewers for their comments.

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.

Data availability

Data will be made available on request.

References

1. Awad, A. I. From Classical Methods to Animal Biometrics: A Review on Cattle Identification and Tracking. Computers and Electronics in Agriculture, 2016, 123, 423-435. https://doi.org/10.1016/j.compag.2016.03.014
2. Bergman, N., Yitzhaky, Y., Halachmi, I. Biometric Identification of Dairy Cows via Real-Time Facial Recognition. Animal, 2024, 18, 101079. https://doi.org/10.1016/j.animal.2024.101079
3. Chen, J. J., Liu, C. X., Gao, Y. F., Liang, Y. Cow Recognition Algorithm Based on Improved Bag of Feature Model. Journal of Computer Applications, 2016, 36, 2346-2351.
4. Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., Sun, J. RepVGG: Making VGG-Style ConvNets Great Again. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, 13728-13737. https://doi.org/10.1109/CVPR46437.2021.01352

5. Disney, W. T., Green, J. W., Forsythe, K. W., Wiemers, J. F., Weber, S. Benefit-Cost Analysis of Animal Identification for Disease Prevention and Control. Revue Scientifique et Technique (International Office of Epizootics), 2001, 20, 385-405. https://doi.org/10.20506/rst.20.2.1277

6. Gong, H., Pan, H. H., Chen, L., Hu, T. L., Li, S. J., Sun, Y., Mu, Y., Guo, Y. Facial Recognition of Cattle Based on SK-ResNet. Scientific Programming, 2022. https://doi.org/10.1155/2022/5773721

7. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E. H. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42, 2011-2023. https://doi.org/10.1109/TPAMI.2019.2913372

8. Hua, L. Z., Feng, Z. X., Zhang, Y. Q., Hao, F., Ge, S. Q., Wang, J., Shao, G. Q. Prevention and Control of African Swine Fever in China: Lessons from Past Outbreaks. Chinese Journal of Animal Infectious Diseases, 2019, 27, 96-104.

9. Jiang, T., Cheng, J. Target Recognition Based on CNN With LeakyReLU and PReLU Activation Functions. 2019. https://doi.org/10.1109/SDPC.2019.00136

10. Kumar, S., Tiwari, S., Singh, S. K. Face Recognition of Cattle: Can It Be Done? Proceedings of the National Academy of Sciences, India, Section A: Physical Sciences, 2016, 86, 137-148. https://doi.org/10.1007/s40010-016-0264-2

11. Lee, Y., Hwang, J.-W., Lee, S., Bae, Y., Park, J. An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, 752-760. https://doi.org/10.1109/CVPRW.2019.00103

12. Li, Y. H., Yao, T., Pan, Y. W., Mei, T. Contextual Transformer Networks for Visual Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45, 1489-1500. https://doi.org/10.1109/TPAMI.2022.3164083

13. Li, Z., Lei, X. M., Liu, S. A Lightweight Deep Learning Model for Cattle Face Recognition. Computers and Electronics in Agriculture, 2022, 195. https://doi.org/10.1016/j.compag.2022.106848

14. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J. Path Aggregation Network for Instance Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 8759-8768. https://doi.org/10.1109/CVPR.2018.00913

15. Lowe, D. G. Object Recognition from Local Scale-Invariant Features. Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, 2, 1150-1157. https://doi.org/10.1109/ICCV.1999.790410

16. Ma, R., Ali, H., Chung, S., Kim, S. C., Kim, H. A Lightweight Pig Face Recognition Method Based on Automatic Detection and Knowledge Distillation. Applied Sciences, 2024, 14, 259. https://doi.org/10.3390/app14010259

17. Połap, D., Jaszcz, A., Wawrzyniak, N., Zaniewicz, G. Bilinear Pooling with Poisoning Detection Module for Automatic Side Scan Sonar Data Analysis. IEEE Access, 2023, 11, 72477-72484. https://doi.org/10.1109/ACCESS.2023.3295693

18. Shojaeipour, A., Falzon, G., Kwan, P., Hadavi, N., Cowley, F. C., Paul, D. Automated Muzzle Detection and Biometric Identification via Few-Shot Deep Transfer Learning of Mixed Breed Cattle. Agronomy, 2021, 11. https://doi.org/10.3390/agronomy11112365

19. Song, H. B., Ma, B. L., Shang, Y. Y., Wen, Y. C., Zhang, S. J. Detection of Young Apple Fruits Based on YOLOv7-ECA Model. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54, 233-242.

20. Wang, C. Y., Bochkovskiy, A., Liao, H. Y. M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object. 2022. https://doi.org/10.1109/CVPR52729.2023.00721

21. Wang, J. B., Wu, J., Wu, J. W., Wang, J. P., Wang, J. YOLOv7 Optimization Model Based on Attention Mechanism Applied in Dense Scenes. Applied Sciences-Basel, 2023, 13. https://doi.org/10.22541/au.168541924.48454251/v1

22. Wang, Q., Wu, B., Zhu, P. F., Li, P., Zuo, W., Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, 11531-11539. https://doi.org/10.1109/CVPR42600.2020.01155

23. Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing, 2004, 13, 600-612. https://doi.org/10.1109/TIP.2003.819861

24. Weng, Z., Fan, L. Z., Zhang, Y., Zheng, Z. Q., Gong, C. L., Wei, Z. Y. Facial Recognition of Dairy Cattle Based on Improved Convolutional Neural Network. IEICE Transactions on Information and Systems, 2022, E105D(6), 1234-1238. https://doi.org/10.1587/transinf.2022EDP7008

25. Weng, Z., Liu, S., Zheng, Z., Zhang, Y., Gong, C. Cattle Facial Matching Recognition Algorithm Based on Multi-View Feature Fusion. Electronics, 2023, 12, 156. https://doi.org/10.3390/electronics12010156

26. Weng, Z., Meng, F. S., Liu, S. Q., Zhang, Y., Zheng, Z. Q., Gong, C. L. Cattle Face Recognition Based on a Two-Branch Convolutional Neural Network. Computers and Electronics in Agriculture, 2022, 196. https://doi.org/10.1016/j.compag.2022.106871

27. Whittier, J. C., Shadduck, J. A., Golden, B. L., Cox, S. Secure Identification, Source Verification of Livestock - The Value of Retinal Images and GPS. 2003. https://doi.org/10.3920/9789086865154_026

28. Woo, S., Park, J., Lee, J.-Y., Kweon, I.-S. CBAM: Convolutional Block Attention Module. 2018, arXiv:1807.06521. https://doi.org/10.1007/978-3-030-01234-2_1

29. Wu, J., Wu, L., Liao, C. H., Xu, Y. H. Review of Pig Face Recognition System for Anti-Fraud Livestock Breeding Insurance. Computer Knowledge and Technology, 2020, 16, 175-176.

30. Xia, M., Cai, C. J. Cattle Face Recognition Using Sparse Representation Classifier. International Journal of Electrical Power & Energy Systems Part B: Applications, 2012, 3, 1499-1505.

31. Xie, Q. J., Wu, M. R., Bao, J., Yin, H., Liu, H. G., Li, X., Zheng, P., Liu, W. Y., Chen, G. Individual Pig Face Recognition Combined with Attention Mechanism. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38, 180-188.

32. Xu, B. B., Wang, W. S., Guo, L. F., Chen, G. P. A Review and Future Prospects on Cattle Recognition Based on Noncontact Identification. Journal of Agricultural Science and Technology, 2020, 22, 79-89.

33. Xu, B. B., Wang, W. S., Guo, L. F., Chen, G. P., Li, Y. F., Cao, Z., Wu, S. S. CattleFaceNet: A Cattle Face Identification Approach Based on RetinaFace and ArcFace Loss. Computers and Electronics in Agriculture, 2022, 193. https://doi.org/10.1016/j.compag.2021.106675

34. Xu, B. B., Wang, W. S., Guo, L. F., Chen, G. P., Wang, Y. W., Zhang, W. J., Li, Y. F. Evaluation of Deep Learning for Automatic Multi-View Face Detection in Cattle. Agriculture-Basel, 2021, 11. https://doi.org/10.3390/agriculture11111062

35. Yang, L., Zhang, R.-Y., Li, L., Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. International Conference on Machine Learning, 2021, 11531-11539. https://doi.org/10.1109/CVPR42600.2020.01155

36. Yao, L., Hu, Z., Liu, C., Liu, H., Kuang, Y., Gao, Y. Cow Face Detection and Recognition Based on Automatic Feature Extraction Algorithm. Proceedings of the ACM Turing Celebration Conference - China, 2019, Article 95. https://doi.org/10.1145/3321408.3322628

37. Zhang, Z. G., Zhang, Z. D., Li, J. N., Wang, H. Y., Li, Y. B., Li, D. H. Potato Detection in Complex Environment Based on Improved YOLOv4 Model. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37, 170-178.

38. Zhao, J. M., Jiang, S. Q., Li, Q. Application of Improved K-SVD Algorithm in Cattle Face Recognition. Transducer and Microsystem Technologies, 2021, 40, 158-160.

39. Zhu, A. G., Liu, L. J., Hou, W. X., Sun, H. B., Zheng, N. N. HSC: Leveraging Horizontal Shortcut Connections for Improving Accuracy and Computational Efficiency of Lightweight CNN. Neurocomputing, 2021, 457, 141-154. https://doi.org/10.1016/j.neucom.2021.06.065

40. Zhu, M. L., Zhao, L. L., He, S. J. Research and Realization on Cattle Face Recognition System Model Based on CNN Combined with SVM and ResNet. Journal of Chongqing University of Technology (Natural Science), 2022, 36, 155-161.