

Re-identifying beef cattle using improved AlignedReID++

YING Xiaoyi¹, ZHAO Jizheng², YANG Lingling³, ZHOU Xinyi³, WANG Lei³, GAO Yannian³, ZAN Linsen⁴,
YANG Wucai⁴, LIU Han³, SONG Huaibo^{3*}

(1. College of Information Engineering, Northwest A & F University, Yangling 712100, China; 2. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710000, China; 3. College of Mechanical and Electronic Engineering, Northwest A & F University, Yangling 712100, China; 4. College of Animal Science and Technology, Northwest A & F University, Yangling 712100, China)

Abstract: Accurate and continuous identification of individual cattle is crucial to precision farming in recent years. It is also the prerequisite to monitor the individual feed intake and feeding time of beef cattle at medium to long distances over different cameras. However, beef cattle can tend to frequently move and change their feeding position during feeding. Furthermore, the great variations in their head direction and complex environments (light, occlusion, and background) can also lead to some difficulties in the recognition, particularly for the bio-similarities among individual cattle. Among them, AlignedReID++ model is characterized by both global and local information for image matching. In particular, the dynamically matching local information (DMLI) algorithm has been introduced into the local branch to automatically align the horizontal local information. In this research, the AlignedReID++ model was utilized and improved to achieve the better performance in cattle re-identification (ReID). Initially, triplet attention (TA) modules were integrated into the BottleNecks of ResNet50 Backbone. The feature extraction was then enhanced through cross-dimensional interactions with the minimal computational overhead. Since the TA modules in AlignedReID++ baseline model increased the model size and floating point operations (FLOPs) by 0.005 M and 0.05 G, the rank-1 accuracy and mean average precision (mAP) were improved by 1.0 percentage points and 2.94 percentage points, respectively. Specifically, the rank-1 accuracies were outperformed by 0.86 percentage points and 0.12 percentage points, respectively, compared with the convolution block attention module (CBAM) and efficient channel attention (ECA) modules, although 0.94 percentage points were lower than that of squeeze-and-excitation (SE) modules. The mAP metric values were exceeded by 0.22, 0.86 and 0.12 percentage points, respectively, compared with the SE, CBAM, and ECA modules. Additionally, the Cross-Entropy Loss function was replaced with the CosFace Loss function in the global branch of baseline model. CosFace Loss and Hard Triplet Loss were jointly employed to train the baseline model for the better identification on the similar individuals. AlignedReID++ with CosFace Loss was outperformed the baseline model by 0.24 and 0.92 percentage points in the rank-1 accuracy and mAP, respectively, whereas, AlignedReID++ with ArcFace Loss was exceeded by 0.36 and 0.56 percentage points, respectively. The improved model with the TA modules and CosFace Loss was achieved in a rank-1 accuracy of 94.42%, rank-5 accuracy of 98.78%, rank-10 accuracy of 99.34%, mAP of 63.90%, FLOPs of 5.45 G, frames per second (FPS) of 5.64, and model size of 23.78 M. The rank-1 accuracies were exceeded by 1.84, 4.72, 0.76 and 5.36 percentage points, respectively, compared with the baseline model, part-based convolutional baseline (PCB), multiple granularity network (MGN), and relation-aware global attention (RGA), while the mAP metrics were surpassed 6.42, 5.86, 4.30 and 7.38 percentage points, respectively. Meanwhile, the rank-1 accuracy was 0.98 percentage points lower than TransReID, but the mAP metric was exceeded by 3.90 percentage points. Moreover, the FLOPs of improved model were only 0.05 G larger than that of baseline model, while smaller than those of PCB, MGN, RGA, and TransReID by 0.68, 6.51, 25.4, and 16.55 G, respectively. The model size of improved model was 23.78 M, which was smaller than those of the baseline model, PCB, MGN, RGA, and TransReID by 0.03, 2.33, 45.06, 14.53 and 62.85 M, respectively. The inference speed of improved model on a CPU was lower than those of PCB, MGN, and baseline model, but higher than TransReID and RGA. The t-SNE feature embedding visualization demonstrated that the global and local features were achieve in the better intra-class compactness and inter-class

Received date: 2024-04-30 Revised date: 2024-07-12

Fund project: National Key Research and Development Program (2023YFD1301801); National Natural Science Foundation of China (32272931); Shaanxi Province Agricultural Key Core Technology Project (2024NYGG005); Shaanxi Province Key R&D Program (2024NC-ZDCYL-05-12)

Biography: YING Xiaoyi, Research interest is image processing. Email: yingxiaoyi@nwfau.edu.cn

*Corresponding author: SONG Huaibo, PhD, Professor, Research interest is image processing and pattern recognition. Email: songhuaibo@nwfau.edu.cn

variability. Therefore, the improved model can be expected to effectively re-identify the beef cattle in natural environments of breeding farm, in order to monitor the individual feed intake and feeding time.

Keywords: method; identify; beef cattle; precision livestock; re-identification; AlignedReID++; deep learning

doi: [10.11975/j.issn.1002-6819.202404229](https://doi.org/10.11975/j.issn.1002-6819.202404229)

CLC number: TP391.4

Documents code: A

Article ID: 1002-6819(2024)-18-0132-15

YING Xiaoyi, ZHAO Jizheng, YANG Lingling, et al. Re-identifying beef cattle using improved AlignedReID++[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(18): 132-146. (in Chinese with English abstract) doi: [10.11975/j.issn.1002-6819.202404229](https://doi.org/10.11975/j.issn.1002-6819.202404229) <http://www.tcsae.org>

应潇溢, 赵继政, 杨玲玲, 等. 基于改进 AlignedReID++的肉牛个体识别方法[J]. 农业工程学报, 2024, 40(18): 132-146. doi: [10.11975/j.issn.1002-6819.202404229](https://doi.org/10.11975/j.issn.1002-6819.202404229) <http://www.tcsae.org>

0 Introduction

Efficient and accurate identification of individual cattle plays a crucial role in advancing precision farming^[1]. It serves as a prerequisite for automatically monitoring animal behaviors and acquiring health information^[2]. With regard to beef cattle, individual feeding time and feed intake are of particular interest for diagnosing health problems, feed efficiency, and growth monitoring^[3-4]. Thus, it is necessary to detect and identify beef cattle in livestock farming environments for automatic feeding records.

For individual cattle identification, traditional methods primarily involve ear tagging^[5], branding^[6], and implantable microchip techniques^[7]. These methods are labor-intensive and prone to inducing stress reactions in cattle, compromising animal welfare. On the other hand, computer vision technology is used for animal monitoring and behavior tracking^[8]. The traditional identification approach based on computer vision usually involves feature extraction algorithms (weber local descriptor, WLD, Histogram of oriented gradient, HOG, etc.). After feature extraction and dimensionality reduction, traditional classifiers (adaBoost, support vector machine, SVM, etc.) are employed for individual cattle identification^[9-11]. However, they often rely on pre-processing efforts to enhance recognition accuracy and generally exhibit poor robustness to factors like lighting and occlusion.

Recently, cattle face/head recognition methods based on deep learning can autonomously learn semantic features in images and be more robust than traditional methods^[12-15]. Wang et al.^[16] employ pre-trained network parameters from the VGGFace dataset to initialize the VGG-16 convolutional neural network (CNN). They employ transfer learning technique and achieve an accuracy of 93% on a small-sample cattle face dataset. CHEN et al.^[17] utilize few-shot learning to train and optimize a lightweight cattle face recognition model, achieving an accuracy of 90% on the test set. WENG et al.^[18] propose a cattle face recognition model based on a dual-branch CNN. The recognition accuracy for unobstructed and relatively diverse poses of Simmental beef cattle and

Holstein dairy cattle faces exceeds 99%. XU et al.^[19] aim to lighten the cattle face detection model by replacing the Backbone network of RetinaFace with MobileNet, acquiring rapid detection and localization of cattle faces. Additionally, ArcFace Loss^[20] is applied for cattle face identification. The proposed CattleFace model achieve a recognition accuracy of 91.3% with speed reaching 24 frames per second (FPS). These methods mentioned above demonstrate efficient recognition of cattle face images based on deep learning.

During the process of cattle feeding, beef cattle face images possess high similarity. Cattle continuously rises and bows its head and changes head orientation. The visual features of cattle faces alter significantly with location and orientation. Changes in illumination and background further increase the difficulty of identification. Cattle move and change places frequently to feed. When their face images are captured by different cameras, their identification becomes more difficult. However, the identifications in such scenarios are also particularly important for continuous monitoring of their feed behavior. A similar situation in computer vision is pedestrian re-identification (ReID). In the literature on pedestrian ReID, many methods have been proposed to address the challenges of recognizing the same individual image across different cameras, times, or scene^[21]. SUN et al.^[22] propose the PCB method. Global feature of a pedestrian image is further sliced horizontally, and multiple horizontal local feature blocks are pooled. Then, the features are downsampled by the small-size convolution operation to generate multiple feature vectors. The classification losses of multiple horizontal feature vectors are computed for model training, and these vectors are fused for similarity computation of image pairs during the test period. WANG et al.^[23] propose MGN, which utilizes a multi-branch method with one global branch and two local branches. Specifically, the two local branches use different horizontal slicing strategies respectively to obtain local feature representations with different granularities to help the model learn more robust feature representations. LUO et al.^[24] propose the AlignedReID++ method, which employs a multi-branch

approach including global and local branches. They introduce the dynamic matching local information method in the local branch to automatically align the horizontal slice information. Further, AlignedReID++ method employs Hard Triplet Loss^[25] and Cross-Entropy Loss to achieve better pedestrian ReID performance. These methods above provide efficient solutions to pedestrian identification across cameras in natural scenes.

Recently, some researchers have extended ideas within the ReID field to solve the problem of cattle recognition. ANDREW et al.^[26] utilize deep metric learning to train the model using the half population of experimental individuals and re-identify the unseen cattle individuals, achieving an accuracy of 93.8%. CHEN et al.^[27] propose the global and part network (GPN). It consists of three branches by using the multiple feature map of layers of ResNet50, allowing adaptive exploitation of the global and local information. It enables the acquisition of more distinctive feature representations for cow ReID. By adding the Spatial Transformer Modules to the part branch to improve the original GPN, the performance of the refined model achieves 98.0% and 91.0% in rank-1 and mean average precision (mAP), respectively. WANG et al.^[28] use ShuffleNet v2 as a feature extraction network for cow ReID. They utilize the BNNeck structure to combine the Triplet Loss^[29] and Cross-Entropy Loss for better convergence of the model, with an accuracy of 82.93% on the training set. Further, they also conduct with other newly photographed individual cows as a test set without re-training the network, and the rank-1 and mAP metrics of this approach are 94.12% and 73.2%, respectively. These methods mentioned above are experimented with Holstein dairy cattle and shed light on the possibility of individual cow identification. However, few researches have been conducted on individual identification of beef cattle during their feeding period. In such cases, cattle head images are highly similar and it is difficult to identify the images captured by medium to long-range cameras with relatively low resolution across cameras.

In this research, the AlignedReID++ method was employed to gain the capability in cross-camera scenarios for cattle ReID. Specially, the ResNet50^[30] Backbone of AlignedReID++ was enhanced by incorporating the triplet attention (TA) module^[31]. In addition, the cross-entropy loss function of the baseline model was replaced with the CosFace Loss function^[32]. CosFace Loss function was combined with the Hard Triplet Loss function to jointly train the model better. The improved method exhibits great potentiality in practical applications of beef cattle feed behavior monitoring.

1 Materials and methods

1.1 Materials

1.1.1 Video capture

Experimental data were collected from the beef cattle breeding center of Northwest A&F University in Yangling District, Xianyang City, Shaanxi Province. The experimental subjects consisted of thirty four beef cattle housed in the cattle barn. The video capture area encompassed six feeding troughs within the cattle barn. The video capture system was manually set up, as depicted in Figure 1a. The cameras were currently positioned approximately 6 m in front of each feeding trough at approximately 3 m above the ground. TL-IPC683-AEZ was adopted as the data acquisition equipment, Kingston Canvas Select Plus MicroSD memory card was used as the video information container, and JDRead Portable WiFi provided wireless network support for the camera. The distance between any two adjacent cameras was approximately 4 m.



a. Experimental area



b. Example of video data screen

Fig.1 Video acquisition system

Figure 1b displayed the video data screen captured by the system. Nearly one month of color video data in different periods (morning, noon, and evening) was collected by six surveillance cameras that recorded video data at a frame rate of 25 FPS. The resolution of image acquisition was 3 840×2 160 pixels.

1.1.2 Dataset preparation

The following steps were undertaken to create the beef cattle ReID dataset. At first, the YOLOv7^[33] model was trained specifically for beef cattle head detection and got excellent test performance. For 12 days of video data (a total of 131 video clips ranging from 5-20 min in length), YOLOv7 performed target detection on keyframes at 8s

intervals, and cattle head images were extracted from each keyframe and resized to 256×256 pixels. These images were initially named by the format of the person ReID dataset (Market1501^[34] and CUHK03^[35]), such as ‘xxxx_c6s4_00000038_04.jpg’, facilitating subsequent manual identity annotation. The extracted images were saved to the local folders belonging to respective segments. Next, the videos before processing and a cattle head identity document were referred to manually label each cattle head image’s identity. For instance, ‘93_c6s4_00000038_04.jpg’, where ‘93’ represented the cattle’s identity, ‘c6’ indicated the sixth camera, ‘s4’ denoted the fourth segment of the camera’s videos, ‘00000038’ indicated the frame number within the video segment, and ‘04’ signified the detection order of the object in the current frame. At last, the SSIM algorithm and manual inspection were used for data cleaning. Then, similar images of each individual under the same camera were eliminated. This operation aimed to obtain cattle head images depicting various poses and lighting conditions (including blurry and dim images on rainy days), as well as images partially obstructed by metal bars or fodder. Additionally, distorted images resulting from limited viewpoints or cropping transformations were removed.

A total of 10 454 cattle head images were obtained. The dataset was further divided into the train set and test set, and the test set included query and gallery sets. The training set consisted of 5 373 head images from 17 cattle, with an average of nearly 300 images per cattle. The test set consisted of additional 17 cattle; the query set had 543 images, and the gallery set had 4 538 images. Every cattle had nearly 30-40 images in the query set and about 200-300 in the gallery set. More detailed information about dataset distribution was shown in Figure 2.

Because of occlusion, illumination, and cattle’s various postures, there were large intra-class and small inter-class variations among beef cattle groups in natural breeding environments. In Figure 3, some samples exhibited large intra-class variations among the images of the same identity. In Row one, occlusion led to the partial obscuring of crucial features on a cattle’s head; in the scene of this research, it was caused mainly by the metal bars or fodder and further caused the loss of discriminative features. In Row two, lighting condition variation led to alterations in the skin color of different cattle head regions. The shadows caused by weak light might result in the loss of details or the blurring of certain parts (especially about the cattle with deep color), consequently reducing the image quality. In Row three, various postures (side-face pose, head up and bow down posture, etc.) imposed partial occlusion of facial features,

rendering key characteristics challenging to discern. In Row four, background information variation further added some unnecessary interference information.

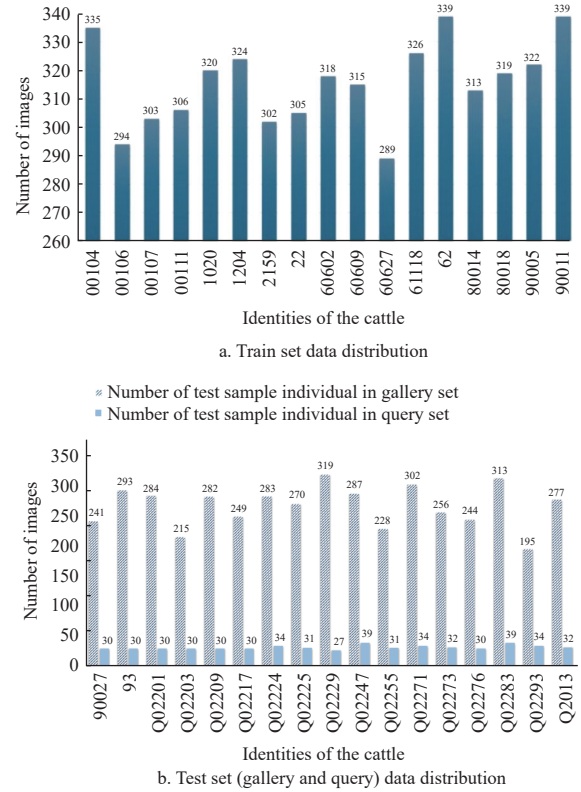


Fig.2 Re-Identification (ReID) dataset distribution

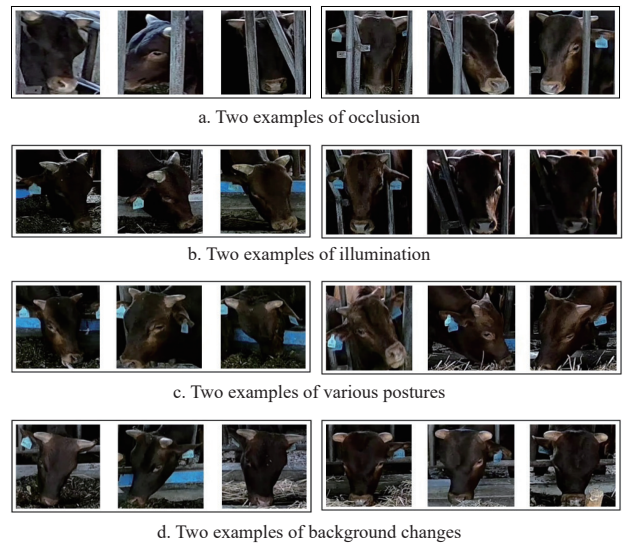


Fig.3 Large intra-class variance caused by occlusion, illumination, various postures and background changes

In Figure 4, subtle distinctions arose among different individuals were observed due to the elevated biological similarity of beef cattle heads. Lighting change caused skin color alteration, which reduced inter-class variance. Further, inter-class variance was influenced by diverse poses or perspectives, such as side profiles and the posture of lowering the head for feeding.

1.2 Beef cattle ReID with AlignedReID++

1.2.1 Beef cattle ReID model framework based on AlignedReID++

The ReID network architecture based on AlignedReID++ was illustrated in Figure 5. The process began with the ResNet50 backbone network. The size of input tensor was $64 \times 3 \times 256 \times 256$, where 64 was the batch size, 3 was the channel number, and 256×256 was the spatial size of images. The ResNet50 backbone network would yield a feature map ($64 \times 2048 \times 256 \times 256$). Subsequently, there were two branches: the global and the local branches. The global branch directly worked on the feature map to explore the global information of images with conducting global average pooling and then transferred it into global feature vectors ($64 \times 2048 \times 1$). At the same time, distance measurement was performed among the global features of sixty-four images in feature embedding space to compare the similarity of images.

To further explore the local spatial information of images for better beef cattle ReID, the Dynamically Matching Local Information (DMLI) algorithm was employed for local distance computation between cattle heads in the local branch. With the feature map ($64 \times 2048 \times 256 \times 256$) after the feature extraction network, horizontal average pooling was performed on the feature

map ($64 \times 2048 \times 8 \times 1$). Then its dimensionality in the channel dimension were reduced ($64 \times 128 \times 8 \times 1$) and it was reshaped to get the horizontal feature vector ($64 \times 8 \times 128 \times 1$). The local distance between two images was computed by Eq. (1).

$$d_{i,j} = \frac{e^{\|l_A^i - l_B^j\|_2} - 1}{e^{\|l_A^i - l_B^j\|_2} + 1}, i, j \in 1, 2, 3, \dots, 8 \quad (1)$$

where, the local features of image A and B were denoted as $l_A = \{l_A^1, l_A^2, l_A^3, \dots, l_A^8\}$ and $l_B = \{l_B^1, l_B^2, l_B^3, \dots, l_B^8\}$. The distance of l_A and l_B was normalized to $[0, 1)$. $d_{i,j}$ represented the distance between the i^{th} horizontal slice of image A and the j^{th} horizontal slice of image B .

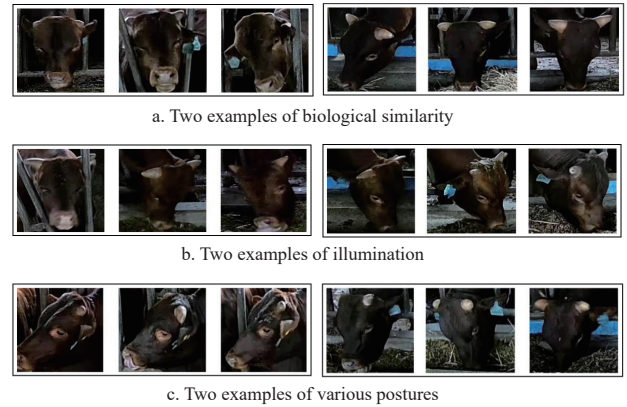
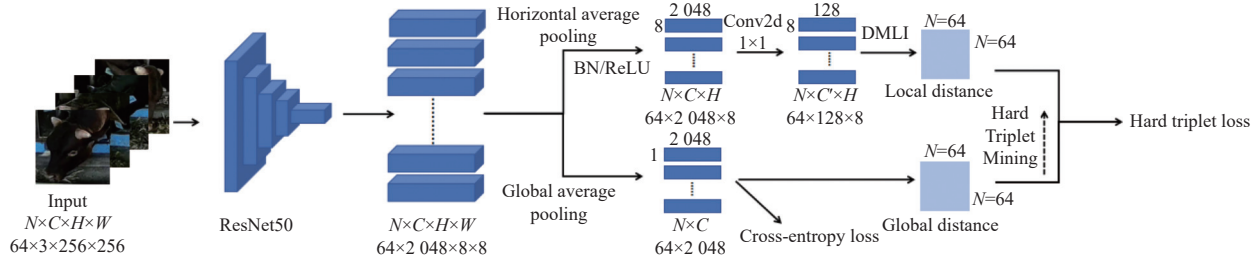


Fig.4 Small inter-class variance caused by biological similarity, illumination, and various postures



Note: $N \times C \times H \times W$ is the size of input tensor, where N is the batch size, C is the channel number, H is the height and W is the width; ResNet50 denotes the feature extraction network; Horizontal Average Pooling represents the average pooling operation in the horizontal direction of the feature map; Global Average Pooling is to compress the spatial dimensions (width and height) of each feature map to produce a fixed-length feature vector; BN denotes two-dimensional batch normalization; ReLU denotes the activation function; Conv2d denotes two-dimensional convolution; DMLI denotes the Dynamically Matching Local Information algorithm; Global Distance denotes the global feature distance between images; Local Distance denotes the local feature distance between images. Hard Triplet Mining denotes the Hard Triplets selections for images; Cross-Entropy Loss and Hard Triplet Loss are loss functions in the training of baseline model.

Fig.5 AlignedReID++ baseline (ResNet50 as feature extraction network)

The shortest path length from (1, 1) to (8, 8) in the distance matrix of local regions determined the ultimate local distance between two images. The points in the matrix were calculated by Eq. (1). A more intuitive illustration of the local distance between image pairs computed by DMLI algorithm was shown in Figure 6.

The shortest distance was calculated through DMLI algorithm, which was shown in Eq. (2).

$$S_{i,j} = \begin{cases} d_{i,j}, & i = 1, j = 1 \\ S_{i-1,j} + d_{i,j}, & i \neq 1, j = 1 \\ S_{i,j-1} + d_{i,j}, & i = 1, j \neq 1 \\ \min(S_{i-1,j}, S_{i,j-1}) + d_{i,j}, & i \neq 1, j \neq 1 \end{cases} \quad (2)$$

where, i represents the i^{th} horizontal slice of image A and j represents the j^{th} horizontal slice of image B . $d_{i,j}$ represents the distance between the i^{th} horizontal slice of image A and the j^{th} horizontal slice of image B . $S_{i,j}$ represents the shortest path length from (1, 1) to (i, j) on the distance matrix, which consists of the distances of the horizontal slices of the two images. After horizontal feature pooling of the feature maps of two images, these two images are obtained eight horizontal feature slices each. When $i=j=8$, $S_{8,8}$ is the ultimate local distance between the two images.

The sum of the global and local distance between images was used during inference stage, which was shown in

Eq. (3).

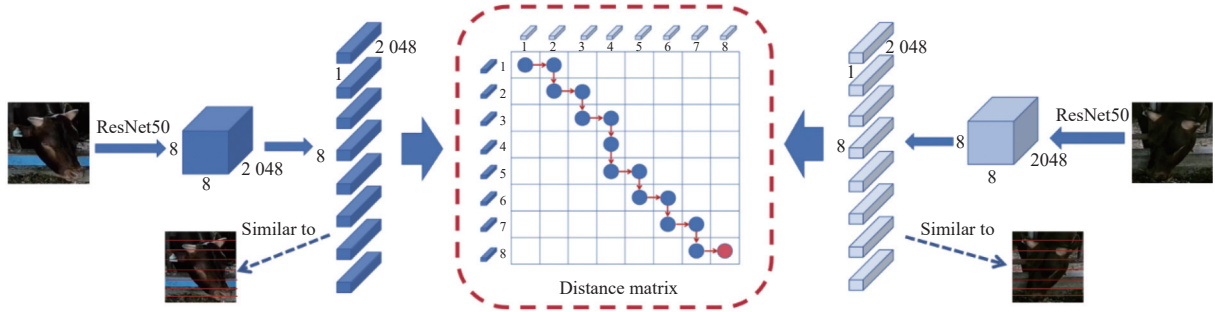
$$d(A, B) = d_g(A, B) + d_l(A, B) \quad (3)$$

where, $d_g(A, B)$ denotes the global distance between image A and B . $d_l(A, B)$ denotes the local distance between image A and B . $d(A, B)$ denotes the total distance between image A and B .

In this research, the total loss of the baseline model was denoted as Eq. (4). The ResNet50 feature extraction network and global branch were trained with Cross-Entropy Loss and

Hard Triplet Loss. $Loss_{CE}$ and $Loss_{HT}^g$ represented the Cross-Entropy Loss and Hard Triplet Loss in the global branch, respectively. At the same time, the Hard Triplets in the local branch were determined by the global distance. $Loss_{HT}^l$ denoted the Hard Triplet Loss in the local branch. And these three parts of the loss were combined to train the AlignedReID++ baseline model together.

$$Loss_{total} = Loss_{CE} + Loss_{HT}^g + Loss_{HT}^l \quad (4)$$



Note: The two images are fed into the feature extraction network to get two different global features. After horizontal average pooling in the local branch of AlignedReID++, these two images are obtained eight horizontal feature slices each. Between the different horizontal feature slices, the feature distances of the horizontal slices are measured using the Euclidean distance metric and the DMLI algorithm is operated in the distance matrix to align the local features (the red arrow path in the figure shows the process of local feature alignment).

Fig.6 DMLI algorithm (ResNet50 as feature extraction network)

1.3 Improvements of beef cattle ReID model based on AlignedReID++ framework

1.3.1 Utilizing Resnet50 with attention mechanism to extract beef cattle head features

The feature extraction network is crucial to the extraction of discriminative, consistent, and high-level feature representations to enable accurate individual ReID. Attention mechanisms play an essential role in ReID tasks by allowing the model to focus on specific regions of an image that are more effective in distinguishing individuals. Moreover, they are adapted well to multiple scale and posture variations of individuals. In this study, TA modules were integrated into the feature extraction network of AlignedReID++ to enhance feature representations among cattle individuals.

TA module utilizes the three-branch structure to capture cross-dimensional interactions via rotation operation and residual transformation. It effectively encodes information across channel and spatial dimensions with minimal computational overhead. Further, Z -Pool is proposed to reduce the input feature's zeroth dimension to two by concatenating the average pooled feature and max pooled feature on the zeroth dimension. The Z -Pool was shown in Eq. (5).

$$Z-Pool(x) = [MaxPool_{0-d}(x), AvgPool_{0-d}(x)] \quad (5)$$

where, x represents the input feature after rotation operation

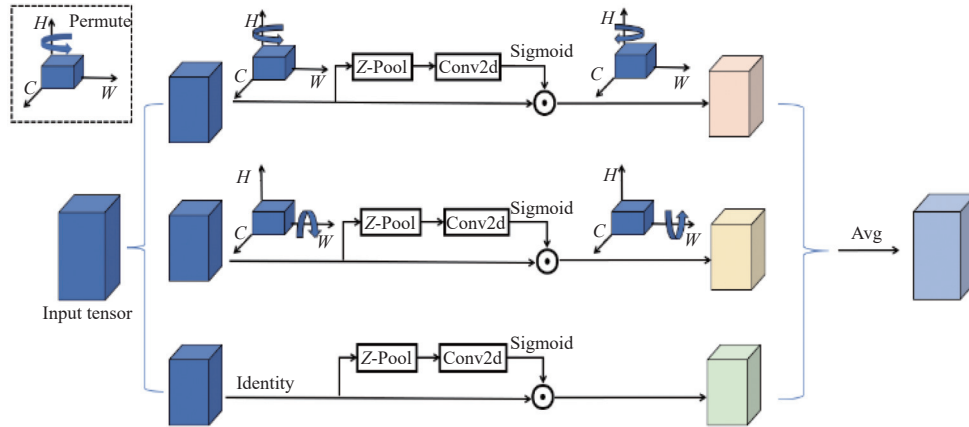
and $0-d$ means the zeroth dimension. $MaxPool_{0-d}(x)$ represents max pooling of x in the zeroth dimension, and $AvgPool_{0-d}(x)$ represents average pooling of x in the zeroth dimension.

The structure of the TA module was depicted in Figure 7a and Figure 7b. The first branch is tasked with calculating attention weights along the channel dimension C and the spatial dimension W . The input feature x is initially permuted from $C \times H \times W$ to $H \times C \times W$ and undergo Z -Pool applied along the H dimension. After this, a 7×7 convolution operation is performed subsequently. Ultimately, spatial attention weights are generated by applying a Sigmoid activation function. The spatial attention weights multiplied x as the residual transformation to get the new feature map. Then, it is restored to the $C \times H \times W$ dimension to gain the attended feature x_1 for element-wise addition. The second branch is employed for the computation of the attention weights across the channel dimension C and the spatial dimension H . Furthermore, x is initially permuted from $C \times H \times W$ to $W \times H \times C$, and the Z -Pool is applied along the W dimension. The subsequent operations are similar to the first branch, generating attended feature x_2 . Moreover, in the third branch, x does not require the rotation operation, and Z -Pool is applied along the C dimension. It is followed by the same operations as the first and second branches to get attended feature x_3 . With the output features from the three branches

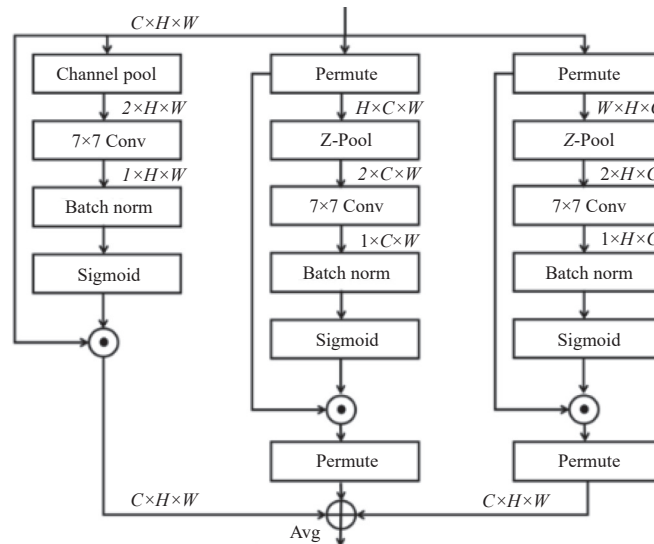
summed, the averaged feature can be computed to capture discriminative feature representations of the image at minimal computational overhead.

The original ResNet50 BottleNeck architecture was

shown in Figure 8a. To facilitate ResNet50 backbone network in capturing cross-dimensional interactions for more robust feature representations, TA module was inserted into every BottleNeck of the ResNet50, shown in Figure 8b.



a. TA module schematic diagram.



b. TA module architecture diagram.

Note: Permute denotes the dimensional transformation; Identity denotes the input tensor without dimensional transformation; Z-Pool denotes the pooling method in Eq. (5); Conv2d denotes two-dimensional convolution; Sigmoid denotes the activation function; \odot denotes broadcast element wise multiplication; Avg denotes the average of features of three branches. $C \times H \times W$ is the size of input tensor; Channel Pool denotes the pooling operations on channels of feature maps; Permute denotes the dimensional transformation; Z-Pool denotes the pooling method in Eq. (5); 7×7 Conv denotes the two-dimensional convolution with a 7×7 kernel; Batch Norm denotes two-dimensional batch normalization; Sigmoid denotes the activation function; \odot denotes broadcast element wise multiplication; \oplus denotes broadcast element-wise addition; Avg denotes the average of features of three branches.

Fig.7 Triplet attention (TA) module

Furthermore, other attention modules, including Squeeze-and-Excitation (SE)^[36] Module, Convolution Block Attention Module (CBAM)^[37], and Efficient Channel Attention (ECA)^[38] Module, were employed to AlignedReID++ with the same operation in Figure 8b to obtain various ResNet50 variants with different attention mechanisms. The performance of AlignedReID++ baseline models added with different attention mechanism mentioned above were evaluated in this research.

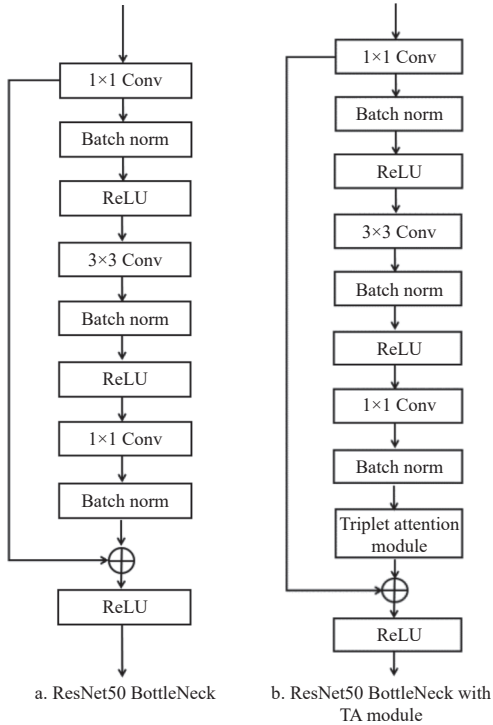
1.3.2 Combining CosFace Loss with Hard Triplet Loss for beef cattle ReID

Angular Margin Loss is proposed in the field of face

recognition. To cope with the recognition of large-scale facial images with high similarity, intra-class compactness and inter-class variability between images should be strengthened, which is also the goal of ReID tasks. The Angular Margin Loss function is designed to assess the angle relationships between feature vectors. It enforces angular boundaries and facilitates the closer proximity of similar feature vectors, which aids in capturing similarity more effectively. A margin is added to the Softmax Loss for the desire to learn discriminative features efficiently. As a typical Angular Margin Loss, CosFace Loss function based on Softmax Loss can be written as Eq. (6).

$$L_{CosFace} = -\frac{1}{N} \sum_{i=1}^N \lg \frac{e^{s(\cos\theta_{y_i} - m)}}{e^{s(\cos\theta_{y_i} - m)} + \sum_{j=1, j \neq y_i}^n e^{s\cos\theta_j}} \quad (6)$$

where, $L_{CosFace}$ represents the CosFace Loss Function, N is the batch size, and n is the number of categories of training samples. y_i is the index of the category of i^{th} training sample. θ_j is the angle between the feature vector of i^{th} training sample and the j^{th} classifier weight vector. m is the margin for the adjustment for the similarity between feature vectors of the same class. s is the scaling factor for control of the length of the feature vector. m and s are all scalar hyperparameters.



Note: 1×1 Conv denotes the two-dimensional convolution with a 1×1 kernel; Batch Norm denotes two-dimensional batch normalization; ReLU denotes the activation function; 3×3 Conv denotes the two-dimensional convolution with a 3×3 kernel; \oplus denotes broadcast element-wise addition; Triplet Attention Module denotes the module in Fig. 7.

Fig.8 ResNet50 BottleNeck architecture diagram

In this research, the last linear layer in the global branch of the AlignedReID++ baseline model was dropped out and the feature vectors before this layer were used for ID Loss computation. Then the CosFace Loss function was utilized to work on these feature vectors to calculate the ID Loss. And it was cooperated with Hard Triplet Loss in two branches to jointly train the AlignedReID++ model. In the selection of better ID Loss to combine with Hard Triplet Loss, the ArcFace Loss (another type of Angular Margin Loss) was utilized to perform the same operation as the CosFace Loss to train the model for comparison.

1.3.3 AlignedReID++(TA+CosFace) model

In this research, with the improvement of TA module and the CosFace Loss function in Section 1.4.1 and Section

1.4.2, the ultimate improved model (AlignedReID++(TA+CosFace)) was carried out and applied for beef cattle ReID.

Furthermore, the representative ReID models, including PCB, MGN, RGA model^[39], TransReID^[40], and AlignedReID++, were compared with the refined method in this study. Even further, the effects of TA module and CosFace Loss on the improved model were analyzed, respectively.

1.4 Beef cattle ReID evaluation metrics

In this study, the performance of various beef cattle ReID methods was evaluated using five metrics: cumulative matching characteristic (CMC, Rank-k) curves, mAP, floating point operations (FLOPs), FPS, and model size. When calculating the CMC and mAP metrics, if a gallery image had the same identity with the query image and both of them were captured from a same camera, the gallery image was excluded from the retrieval list of the query image.

Firstly, define a function Acc_k shown in Eq. (7). The function will return 1 if the top-k gallery samples have the same identity as the query sample in the retrieval list, which is ordered by the similarity rank between the query sample and gallery samples. The similarity is usually based on Euclidean distance or cosine similarity; otherwise, it will return 0.

$$Acc_k = \begin{cases} 1, & \text{if top-k gallery samples} \\ & \text{contain query identity} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

Then, when computing Rank-k shown in Eq. (8), it suffices to compute the average of Acc_k values for all query images, where Q represents the entire set of query images, q denotes a query sample image, and g denotes all images of the gallery.

$$\text{Rank-k} = \frac{1}{\|Q\|} \sum_{q \in Q} Acc_k(q, g) \quad (8)$$

The CMC curve can be drawn with the Rank-k (Rank-1, Rank-5, Rank-10, etc.) calculated.

mAP is proposed to compare the quality between the two rank lists in some cases that CMC curves fail to provide. The average precision (AP) value of the query sample is actually calculated by integrating the precision corresponding to the query sample at different recall rates within the [0,1] interval. mAP is the average of AP values of all images in query set.

1.5 Experimental setting for improved method

In AlignedReID++(TA+CosFace) model, ResNet50 with TA module was employed as the feature extraction network, which was initialized by the weight pre-trained on

ImageNet. The improved model was further trained with CosFace Loss and Hard Triplet Loss for beef cattle ReID in the experiments. The improved model was trained for 400 epochs using the Adam optimizer; the learning rate was initially set to 2×10^{-4} and weight decay was $5e-04$. The step size to decay learning rate was 150 epochs, and the scaling factor of the learning rate was 0.1. To form the batch of one epoch for the training, $P=17$ (the number of cattle IDs in the train set) and $R=16$ (images randomly selected from each class in P classes) were chosen. The batch consisted of 272 images in the training stage; the train and test mini_batch sizes were all set to be 64. Each cattle head image was resized into 256×256 pixels. The data augmentation included Random Horizontal Flip and Random Erasing. The margin for the CosFace Loss and Hard Triplet Loss in global and local branches was 0.3. Additionally, the scalar hyperparameter s in CosFace Loss was initially set to be 64.

1.6 Experimental setting for t-SNE feature embedding visualization

In order to examine whether AlignedReID++ (TA+CosFace) model gained the ability to be able to extract discriminative features on the cattle head image, the t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction method^[41] was used to perform a qualitative analysis in this research. Thirty four individual cattle in the training and test sets were selected with 60 images each. Of these, for one individual, 10 images were captured per camera. For cattle with small activity ranges, such as individuals captured under only three cameras, the images taken by these cameras were used to supplement the total number of images.

1.7 Experimental platform

All experiments were implemented on the Windows 10 System with a 3.20 GHz 12th Gen Intel(R) Core(TM) i9-12900K processor. The GPU was NVIDIA GeForce RTX 3090Ti. The algorithm development platform was Python 3.7, and the deep learning framework was PyTorch 1.7.0.

2 Results and analysis

2.1 Performance comparison of AlignedReID++ models with different attention mechanisms

The ReID results of AlignedReID++ models with different attention mechanism modules were shown in Table 1. It could be noted that these improvement approaches achieved impressive results on CMC, with about 93% Rank-1 accuracy, 98% Rank-5, and 99% Rank-10 accuracy. AlignedReID++ with SE module got Rank-1 of 94.52%, Rank-5 of 98.78%, Rank-10 of 99.26%, and mAP of

60.20%. AlignedReID++ with CBAM module got Rank-1 of 92.94%, Rank-5 of 98.56%, Rank-10 of 99.28%, and mAP of 59.56%. AlignedReID++ with ECA module got Rank-1 of 92.54%, Rank-5 of 98.34%, Rank-10 of 99.19%, and mAP of 60.30%. AlignedReID++ with TA module got Rank-1 of 93.58%, Rank-5 of 98.86%, Rank-10 of 99.52%, and mAP of 60.42%. Only the addition of ECA modules in AlignedReID++ showed a slight drop in CMC compared to the baseline model that a slight decrease of 0.04 percentage points in rank-1 accuracy. However, the modified models with different attention modules showed significant improvements in mAP. The addition of SE, CBAM, ECA, and TA modules in AlignedReID++ improved the mAP compared to the baseline model by 2.72, 2.08, 2.82 and 2.94 percentage points, respectively.

Table 1 Comparisons of testing results between AlignedReID++ models with different ResNet50 variants

Method	Rank-1 /%	Rank-5 /%	Rank-10 /%	mAP /%
AlignedReID++(ResNet50)	92.58	98.40	99.30	57.48
AlignedReID++(ResNet50+SE)	94.52	98.78	99.26	60.20
AlignedReID++(ResNet50+CBAM)	92.94	98.56	99.28	59.56
AlignedReID++(ResNet50+ECA)	92.54	98.34	99.19	60.30
AlignedReID++(ResNet50+TA)	93.58	98.86	99.52	60.42

2.2 Performance comparison of different ID Loss function combined with Hard Triplet Loss function

The experimental results of various loss function combinations applied in AlignedReID++ were shown in Table 2. There were some improvements in the model's performance with the help of both types of Angular Margin Loss. The model whose ID Loss was CosFace Loss obtained the best performance on CMC and mAP metrics, Rank-1 of 92.82%, Rank-5 of 98.78%, Rank-10 of 99.12% and mAP of 58.40%, respectively. With the combination of CosFace Loss and Hard Triplet Loss, the model outperformed the baseline by 0.24 and 0.92 percentage points, on Rank-1 and mAP, respectively. The model whose ID Loss was ArcFace Loss achieved Rank-1 of 92.46%, Rank-5 of 98.52%, Rank-10 of 99.26% and mAP of 57.84%, respectively. The performance was 0.12 percentage points lower than the baseline model on Rank-1 and was 0.36 percentage points higher than the baseline model on mAP.

Table 2 Comparisons of testing results between various loss function combinations (CE, CosFace, and ArcFace represent Cross-Entropy Loss, CosFace Loss, ArcFace Loss that the ID Loss in the global branch. HT_{global} and HT_{local} are Hard Triplet Losses in the global and local branches)

Method	Rank-1 /%	Rank-5 /%	Rank-10 /%	mAP /%
AlignedReID++(CE+ HT_{global} + HT_{local})	92.58	98.40	99.30	57.48
AlignedReID++(CosFace + HT_{global} + HT_{local})	92.82	98.78	99.12	58.40
AlignedReID++(ArcFace + HT_{global} + HT_{local})	92.46	98.52	99.26	57.84

2.3 Comparison of different models based on AlignedReID++ with different modules

According to the results in Table 3, the effectiveness of TA module and CosFace Loss applied to the AlignedReID++ can be evaluated. AlignedReID++ with CosFace Loss outperformed the baseline by 0.24 percentage points and 0.92 percentage points on Rank-1 and mAP, respectively. AlignedReID++ with TA module outperformed the baseline by 1.0 percentage points and 2.94 percentage points on Rank-1 and mAP, respectively. AlignedReID++ with the combination of two extension modules finally turned out to be the improved method called AlignedReID++(TA+CosFace). It achieved Rank-1 of 94.42%, Rank-5 of 98.78%, Rank-10 of 99.12% and mAP of 63.90%. The performance exceeded the baseline model effectively by 1.84 percentage points and 6.42 percentage points on Rank-1 and mAP, respectively. The improved method was superior to the two models with the single improvement module and got the best performance on Rank-1 and mAP. The experimental results strongly confirmed the effectiveness of two improvement modules applied to AlignedReID++ and the superiority of the improved method.

Table 3 Comparisons of testing results between different models based on AlignedReID++ with different modules

Method	Rank-1 /%	Rank-5 /%	Rank-10 /%	mAP /%
AlignedReID++(Baseline)	92.58	98.40	99.30	57.48
AlignedReID++(TA)	93.58	98.86	99.52	60.42
AlignedReID++(CosFace)	92.82	98.78	99.12	58.40
AlignedReID++(TA+CosFace)	94.42	98.78	99.34	63.90

2.4 Performance comparison between various representative models and the improved method

The comparison results of improved model and representative ReID methods were shown in Table 4. The improved model achieved the Rank-1 of 94.42%, Rank-5 of 98.78%, Rank-10 of 99.34%, mAP of 63.90%, FLOPs of 5.45 G, FPS of 5.64, and model size of 23.78 M. AlignedReID++(TA+CosFace) outperformed the baseline by 1.84 and 6.42 percentage points on Rank-1 and mAP, respectively. Its model size was smaller than baseline model

Table 4 Comparisons between different ReID methods and the improved method

Method	Rank-1 /%	Rank-5 /%	Rank-10 /%	mAP /%	FLOPs /G	FPS /(frame·s ⁻¹)	Model size /M
PCB	89.70	96.30	97.84	58.04	6.13	14.59	26.11
MGN	93.66	98.64	99.30	59.60	11.96	7.94	68.84
RGA	89.06	97.20	98.62	56.52	30.85	3.54	38.31
TransReID	95.40	98.70	99.10	60.00	22.00	5.58	86.63
AlignedReID++	92.58	98.40	99.30	57.48	5.40	7.56	23.81
AlignedReID++(TA+CosFace)	94.42	98.78	99.34	63.90	5.45	5.64	23.78

3 Discussions

This study is dedicated to re-identifying beef cattle heads over medium to long-range distances and across cameras, which is essential for the monitoring of cattle feed intake. The AlignedReID++ model was improved by using

by 0.03 M. PCB achieved the Rank-1 of 89.70%, Rank-5 of 96.30%, Rank-10 of 97.84%, mAP of 58.04%, FLOPs of 6.13 G, FPS of 14.59, and model size of 26.11 M. MGN achieved the Rank-1 of 93.66%, Rank-5 of 98.64%, Rank-10 of 99.30%, mAP of 59.60%, FLOPs of 11.96 G, FPS of 7.94, and model size of 68.84 M. Compared to the models that were based on Local Feature Learning (PCB and MGN), AlignedReID++(TA+CosFace) was also superior to theirs, exceeded by 4.72 and 5.86 percentage points with PCB on Rank-1 and mAP, respectively, and exceeded by 0.76 and 4.30 percentage points with MGN on Rank-1 and mAP, respectively. Its model size was smaller than PCB and MGN by 2.33 and 45.06 M, respectively. The TransReID model achieved the Rank-1 of 95.40%, Rank-5 of 98.70%, Rank-10 of 99.10%, mAP of 60.00%, FLOPs of 22.00 G, FPS of 5.58, and model size of 86.63 M. AlignedReID++(TA+CosFace) outperformed it by 3.90 percentage points on mAP, and although with a slight decrease by 0.98 percentage points on Rank-1. The improved model's size was smaller (more lightweight) than it, about 62.9 M smaller. RGA model achieved the Rank-1 of 89.06%, Rank-5 of 97.20%, Rank-10 of 98.62%, mAP of 56.52%, FLOPs of 30.85 G, FPS of 3.54, and model size of 38.31 M. AlignedReID++(TA+CosFace) outperformed it by 5.36 and 7.38 percentage points on Rank-1 and mAP, respectively. Its model size was smaller than RGA by 14.53 M. Furthermore, AlignedReID++(TA+CosFace) also showed excellent performance on Rank-5 and Rank-10, exceeding these representative methods on the two metrics. Besides, the FLOPs of AlignedReID++(TA+CosFace) were smaller than PCB, MGN, RGA, and TransReID by 0.68, 6.51, 25.4 and 16.55 G, respectively. The FLOPs were 0.05 G bigger than the baseline model while AlignedReID++(TA+CosFace) had the smallest model size among the compared models. The FPS on the CPU of AlignedReID++(TA+CosFace) was lower than PCB, MGN, and its baseline model. It was higher than TransReID and RGA.

the TA module and the CosFace Loss function. With the assistance of the advanced attention mechanism and Angular Margin Loss function, the performance of the improved method on the test set was 94.42% for Rank-1 and 63.90% for mAP. AlignedReID++(TA+CosFace) had less computational

parameters than the baseline model, whereas it increased rank-1 and mAP by 1.84 and 6.42 percentage points compared to the baseline model, respectively. Furthermore, it also exhibited better ReID performance compared to other representative models.

3.1 Reasons for choosing TA module for the improved model

According to Table 1, all the results confirmed that attention modules in the ResNet50 feature network enabled the model to extract crucial features representations within beef cattle head images and further improve the ReID performance. Although the AlignedReID++(ResNet50+TA) was inferior to the AlignedReID++(ResNet50+SE) on Rank-1, the former exceeded the latter on mAP with a smaller addition of model parameters and computational overhead. The addition of TA increased the model size and FLOPs by 0.005 M and 0.05 G, while the addition of SE increased the model size and FLOPs by 2.44 M and 2.53 G. It was also one of the reasons that the ResNet50 with TA modules was chosen as the feature extraction network in the improved method in this research.

3.2 Comparison with other studies on cattle ReID

Simultaneously, the proposed method was compared with other studies on cattle ReID. Table 5 shows that all studies achieved a Rank-1 accuracy of over 90%. The proposed method could reach 94.42% Rank-1 accuracy, which was 0.67 and 0.30 percentage points higher than ANDREW's method^[26] and WANG's method^[28]. And the performance was lower than CHEN's method^[27] by 3.58 percentage points and BAKHSHAYESHI's method^[42] by 0.71 percentage points. With regard to mAP metric, the mAP of the proposed method was lower than CHEN's method^[27] and WANG's method^[28]. On the one hand, pure-coloured beef cows are much harder to distinguish than cows with patterns, and on the other hand, the sample size of this study was truly insufficient. In CHEN et al.'s study, there were 2 000 training sample individuals, allowing the model to learn more robust generic features to identify cows. For the study of ReID, it is seen that the number of training samples plays an important role in the generalization performance of the model.

Table 5 - Comparison of different ReID performance for the proposed method and others

Study	Method	ROI	Category	Performance %/
ANDREW et al. ^[26]	ResNet50	Back	Holstein Cattle	Rank-1: 93.75
CHEN et al. ^[27]	GPN	Head	Holstein Cattle	Rank-1: 98.0 mAP: 91.0
WANG et al. ^[28]	ShuffleNet v2	Body	Holstein Cattle	Rank-1: 94.12 mAP: 73.20
BAKHSHAYESHI et al. ^[42]	SNN	Head	Holstein Cattle	Rank-1: 95.13
Ours	Improved AlignedReID++	Head	Beef Cattle	Rank-1: 94.42 mAP: 63.90

3.3 Loss trend curves in respective branches of various models based on AlignedReID++

The training loss trend curves of different models based on AlignedReID++ in Table 3 were shown in Figure 9.

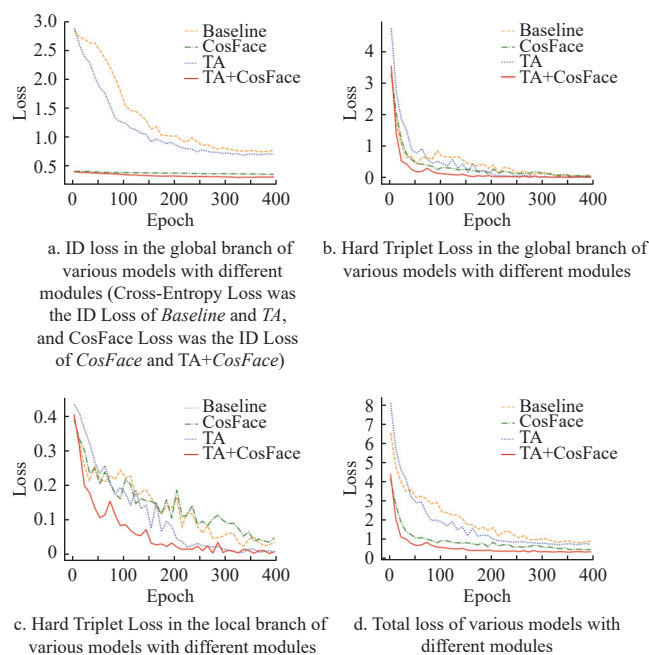


Fig.9 Loss trend curves of the models based on AlignedReID++.

Overall, the loss trend curves in two branches of different models typically showed an initial steep decrease followed by fluctuation within a certain range as the number of epochs increased. According to each loss trend curve, it could be noted that the hybrid loss function combination, namely the combination of Cross-Entropy Loss function and Hard Triplet Loss function or CosFace Loss function and Hard Triplet Loss function, was able to train the hybrid loss of respective models to converge to a stable value. TA+CosFace model obtained a smaller convergence loss value than other models during the training stage. For the TA+CosFace model, the average of the total loss of the first ten epochs was 4.35 (CosFace Loss was 0.42, Hard Triplet Loss in the global branch was 3.53, Hard Triplet Loss in the local branch was 0.40). With the epochs increasing, the hybrid loss decreased rapidly with the help of the CosFace Loss function and Hard Triplet Loss function. It became stable after about 290 epochs. In the end, the hybrid loss of the model converged to about 0.33. The CosFace Loss was about 0.31, and the Hard Triplet Losses in two branches were all around 0.01, approaching zero. The convergence loss result of TA+CosFace was superior to other models based on AlignedReID++, and TA+CosFace realized the optimal training compared with others. Based on the loss trend curves of these models, the effects of two improvement modules on the AlignedReID++(TA+CosFace) model were also validated and the superiority of AlignedReID++

(TA+CosFace) was proved.

3.4 t-SNE feature embedding visualization of individual samples

The t-SNE visualization of the global and local features extracted from the AlignedReID++(TA+CosFace) model was shown in Figure 10. It could be seen that in the training set, the global and local features of the same individuals were tightly clustered, and the different individuals were separated far from each other, which perfectly achieved intra-class compactness and inter-class variability. For the seventeen beef cattle in the test set that did not appear in the training phase, it could be seen that the global and local features of the test individuals could also achieve intra-class compactness. This result indicated that the model had learned how to accurately discern individual cattle with the same identity. Among different beef cattle, most of the individual samples kept certain distances from each other. The visualization results demonstrated that some individuals were suitable for global features for discrimination among them. Between individual samples of ‘93’ and ‘90 027’, the inter-class distances based on global features were larger than those based on local features. The pair of samples was highlighted with a red rectangle in Figure 10a and Figure 10b. While, some individual samples were more appropriate to use local features for discrimination. For example, ‘Q02255’ and ‘Q02217’ showed a more explicit class border in local features. Besides, ‘Q02255’ and ‘Q02217’ showed better intra-class compactness in local features than global features. The pair of samples was highlighted with a blue rectangle in Figure 10c and Figure 10d. For some similar individuals, such as ‘Q02225’, ‘Q02224’, and ‘Q02293’ cattle, highlighted with a yellow rectangle in Figure 10c and Figure 10d, their inter-class differentiation between individuals needed to be further improved.

3.5 Retrieval results comparison between the baseline and the improved model

To further validate the effectiveness of the improved method, the retrieval results of cases with motion blur, with sideways posture, with occlusion, with light change and in weak light condition were explanatorily discussed. Figure 11 showed some results with low AP values. In each subfigure, the first image was the query image. The subsequent five gallery images were the closest matches to the query image in the feature embedding space, and they were captured from different camera perspectives. Among them, the correctly matched gallery images were surrounded by red solid boxes and the incorrectly matched gallery images were surrounded by blue dotted boxes.

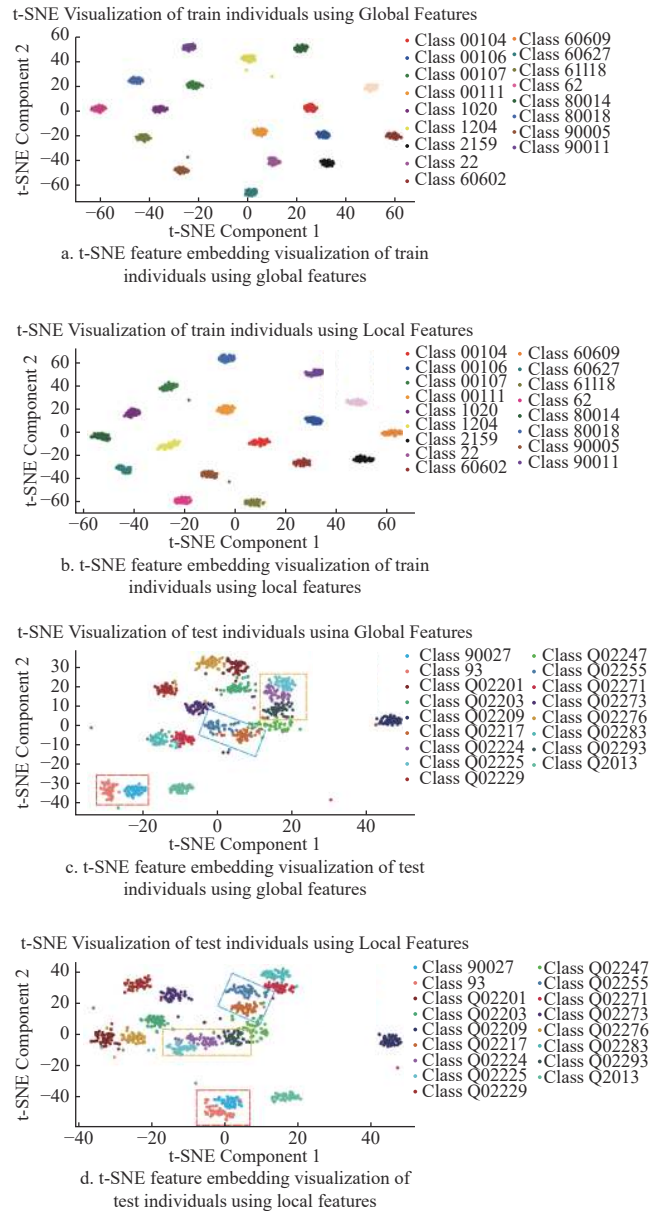
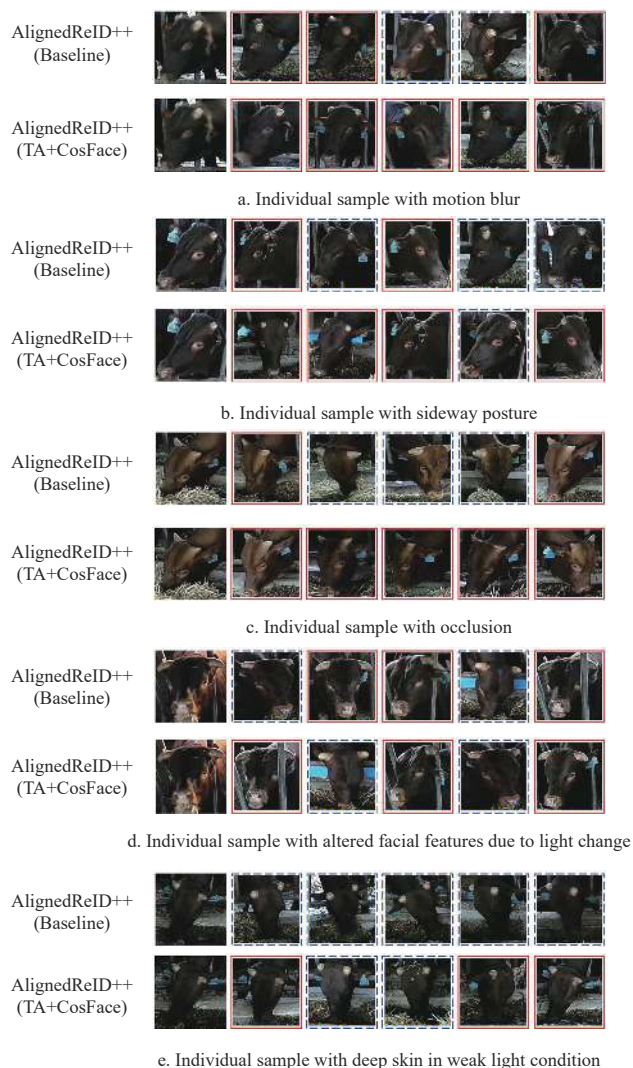


Fig.10 t-SNE feature embedding visualization of global and local features

In Figure 11a, the query image had some motion blur. The AP value of the baseline model was 40.50%, and three images among the top-5 retrieval results correctly matched the query image. The AP value of the improved model was 50.43%, which was 9.93 percentage points higher than the baseline model. All images among the top-5 retrieval results matched the query image correctly. In Figure 11b, the cattle in the query image had a sideways posture. The AP value of the baseline model was 34.49%. Two images among the top-5 retrieval results correctly matched the query image. The AP value of the improved model was 55.32%, which was 20.83 percentage points higher than the baseline model. Four images among the top-5 retrieval results correctly matched the query image.



Note: In each subfigure, the first image was the query image. The subsequent five images were gallery images. Among them, the correctly matched gallery images were surrounded by red solid boxes and the incorrectly matched gallery images were surrounded by blue dotted boxes.

Fig.11 Retrieval results of hard samples in query set with the low Average Precision (AP) value between the baseline model and the improved model

In Figure 11c, the query image was obscured by the feed. The AP value of the baseline model was 24.32%. Two images among the top-5 retrieval results correctly matched the query image. The AP value of the improved model was 54.38%, which was 30.46 percentage points higher than the baseline model. All images among the top-5 retrieval correctly matched the query image. In Figure 11d, there was a query image of facial features altered due to light change. In this case, the AP value of the baseline model was 14.75%. Three images among the top-5 retrieval results correctly matched the query image. The AP value of the improved model was 22.75%, which was 8.00 percentage points than the baseline model. Three images among the top-5 retrieval results correctly matched the query image. In Figure 11e, there was a query image that was in weak light condition and the cattle was dark-skinned. In this case, the AP value of the

baseline model was 8.60%. None of the images among the top-5 retrieval results correctly matched the query image. The AP value of the improved model was 16.08%, which was 7.48 percentage points higher than the baseline model. Three images among the top-5 retrieval results correctly matched the query image. These results above indicated that the improved model had a better ReID performance in the cases of complex variations in images.

3.6 Limitations and Future Work

The study strongly demonstrates that it is feasible to perform medium to long-range and cross-camera identification of beef cattle heads with models transferred from pedestrian ReID to cattle ReID. However, in the current study, the improved model was evaluated on a relatively small number of beef cattle. In future, it is needed to expand the number of sample individuals in the training set to obtain better generalization performance. Second, the beef cattle ReID model proposed was evaluated in this study. It could be combined with target detection and behavioral recognition models to form an integrated solution for individual cattle's feed monitoring in future work. Third, breeding, slaughter, disease, and death of beef cattle can lead to frequent changes in localized herds. When new beef cattle individuals enter the application scenario, it is critical to enable the existing models to achieve fast learning and generalization based on a small amount of new sample data, so that effective ReID of new individuals can be performed timely. In the future, Meta-Learning^[43] or Incremental Learning^[44] will be applied to quickly update the model using a small number of new sample images, enabling the ReID of newly added individuals in the scene.

4 Conclusions

In precision farming, accurate beef cattle identification is an essential prerequisite in monitoring the behaviors of beef cattle, especially for feeding behavior. In this research, the pedestrian re-identification (ReID) technique was utilized for cross-camera retrieval of the cattle heads to verify their identities. Triplet attention (TA) modules were inserted into the BottleNecks of the ResNet50 Backbone of the AlignedReID++ model. CosFace Loss replaced Cross-Entropy Loss to cooperate with Hard Triplet Loss for better model training. The improved model achieved Rank-1 of 94.42%, Rank-5 of 98.78%, Rank-10 of 99.34%, mean average precision (mAP) of 63.90%, floating point operations (FLOPs) of 5.45 G, frames per second (FPS) of 5.64, and model size of 23.78 M. It outperformed its baseline model and the classical ReID methods on rank-1 and mAP. The improved model got the smallest model size between the comparisons and relatively small FLOPs. This research demonstrates the feasibility to utilize ReID method for

continual individual beef cattle recognition across cameras in natural scenes.

[参 考 文 献]

- [1] BERCKMANS D. General introduction to precision livestock farming[J]. *Animal Farming*, 2017, 7(1): 6-11.
- [2] CONGDON J, HOSSEINI M, GADING E, et al. The future of artificial intelligence in monitoring animal identification, health, and behaviour[J]. *Animals*, 2022, 12(13): 1711.
- [3] DAVISON C, BOWEN J, MICHIE C, et al. Predicting feed intake using modelling based on feeding behaviour in finishing beef steers[J]. *Animal*, 2021, 15(7): 100231.
- [4] WANG K Y, ZHAO X Y, HE Yong. Review on noninvasive monitoring technology of poultry behavior and physiological information[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2017, 33(20): 197-209. (in Chinese with English abstract)
- [5] JOHNSTON A, EDWARDS D. Welfare implications of identification of cattle by ear tags[J]. *Veterinary Record*, 1996, 138(25): 612-614.
- [6] LAY Jr D, FRIEND T, RANDEL R, et al. Behavioral and physiological effects of freeze or hot-iron branding on crossbred cattle[J]. *Journal of Animal Science*, 1992, 70(2): 330-336.
- [7] SETSER M, CANTOR M, COSTA J. A comprehensive evaluation of microchips to measure temperature in dairy calves[J]. *Journal of Dairy Science*, 2020, 103(10): 9290-9300.
- [8] FUENTES S, VIEJO C, TONGSON E, et al. Animal biometric assessment using non-invasive computer vision and machine learning are good predictors of dairy cows age and welfare: The future of automated veterinary support systems[J]. *Journal of Agriculture and Food Research*, 2022, 10: 100388.
- [9] CAI C, LI J. Cattle face recognition using local binary pattern descriptor[C]// 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Kaohsiung, Taiwan: IEEE, 2013: 1-4.
- [10] KUMAR S, SINGH S. Cattle recognition: A new frontier in visual animal biometrics research[J]. *Proc. Natl. Acad. Sci. , India, Sect. A Phys Sci.* 2020, 90: 689-708.
- [11] KUMAR S, TIWARI S, SINGH S. Face recognition for cattle[C]// 2015 Third International Conference on Image Information Processing (ICIIP). Wagnaghat, India : IEEE, 2015: 65-72.
- [12] HOSSAIN M, KABIR M, ZHENG L, et al. A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions[J]. *Artificial Intelligence in Agriculture*, 2022, 6: 138-155.
- [13] MAHMUD M, ZAHID A, DAS A, et al. A systematic literature review on deep learning applications for precision cattle farming[J]. *Computers and Electronics in Agriculture*, 2021, 187: 106313.
- [14] YANG S Q, LIU Y Q, WANG Z, et al. Improved YOLO v4 model for face recognition of dairy cow by fusing coordinate information[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CASE)*, 2021, 37(15): 129-135. (in Chinese with English abstract)
- [15] XU X S, WANG Y F, HUA Z X, et al. Light-weight recognition network for dairy cows based on the fusion of YOLO v5s and channel pruning algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CASE)*, 2023, 39(15): 153-163. (in Chinese with English abstract)
- [16] WANG H, QIN J, HOU Q, et al. Cattle face recognition method based on parameter transfer and deep learning[J]. *Journal of Physics: Conference Series*, 2020, 1453: 012054.
- [17] CHEN Y, KUAN C, HSU J, et al. Lightweight cow face recognition algorithm based on few-shot learning for edge computing application[C]// 2021 ASABE Annual International Virtual Meeting. Michigan, USA: the American Society of Agricultural and Biological Engineers, 2021: 1.
- [18] WENG Z, MENG F, LIU S, et al. Cattle face recognition based on a two-branch convolutional neural network[J]. *Computers and Electronics in Agriculture*, 2022, 196: 106871.
- [19] XU B, WANG W, GUO L, et al. CattleFaceNet: A cattle face identification approach based on RetinaFace and ArcFace loss[J]. *Computers and Electronics in Agriculture*, 2022, 193: 106675.
- [20] DENG J, GUO J, XUE N, et al. ArcFace: Additive angular margin loss for deep face recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, California, USA: IEEE, 2019: 4690-4699.
- [21] MING Z, ZHU M, WANG X, et al. Deep learning-based person re-identification methods: A survey and outlook of recent works[J]. *Image and Vision Computing*, 2022, 119: 104394.
- [22] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C]// Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany: Springer, 2018: 480-496.
- [23] WANG G, YUAN Y, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification[C]// Proceedings of the 26th ACM International Conference on Multimedia. New York, USA: ACM, 2018: 274-282.
- [24] LUO H, JIANG W, ZHANG X, et al. AlignedReID++: Dynamically matching local information for person re-identification[J]. *Pattern Recognition*, 2019, 94: 53-61.
- [25] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification[EB/OL]. (2017-03-22) [2017-11-21] <https://doi.org/10.48550/arXiv.1703.07737>
- [26] ANDREW W, GAO J, MULLAN S, et al. Visual identification of individual Holstein-Friesian cattle via deep metric learning[J]. *Computers and Electronics in Agriculture*, 2021, 185: 106133.
- [27] CHEN X, YANG T, MAI K, et al. Holstein cattle face re-identification unifying global and part feature deep network with attention mechanism[J]. *Animals*, 2022, 12: 1047.
- [28] WANG Y, XU X, WANG Z, et al. ShuffleNet-Triplet: A lightweight re-identification network for dairy cows in natural scenes[J]. *Computers and Electronics in Agriculture*, 2023, 205: 107632.
- [29] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 815-823.
- [30] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 770-778.

- [31] MISRA D, NALAMADA T, ARASANIPALAI A, et al. Rotate to attend: Convolutional triplet attention module[C]// Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE, 2021: 3139-3148.
- [32] WANG H, WANG Y, ZHOU Z, et al. CosFace: Large margin cosine loss for deep face recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 5265-5274.
- [33] WANG C, BOCHKOVSKIY A, LIAO H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 7464-7475.
- [34] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: A benchmark[C]// Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile: IEEE, 2015: 1116-1124.
- [35] LI W, ZHAO R, XIAO T, et al. DeepReID: Deep filter pairing neural network for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA: IEEE, 2014: 152-159.
- [36] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA: IEEE, 2018: 7132-7141.
- [37] WOO S, PARK J, LEE J, et al. CBAM: Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany: Springer, 2018: 3-19.
- [38] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 11534-11542.
- [39] ZHANG Z, LAN C, ZENG W, et al. Relation-aware global attention for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 3186-3195.
- [40] HE S, LUO H, WANG P, et al. TransReID: Transformer-based object re-identification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada: IEEE, 2021: 15013-15022.
- [41] VAN D, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(86): 2579-2605.
- [42] BAKHSHAYESHI I, ERFANI E, ROSA F, et al. An intelligence cattle reidentification system over transport by siamese neural networks and YOLO[J]. *IEEE Internet of Things Journal*, 2024, 11(2): 2351-2363.
- [43] PU N, LIU Y, CHEN W, et al. Meta reconciliation normalization for lifelong person re-identification[C]// Proceedings of the 30th ACM International Conference on Multimedia. New York, USA: ACM, 2022: 541-549.
- [44] PU N, ZHONG Z, SEBE N, et al. A memorizing and generalizing framework for lifelong person re-identification[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(11): 13567-13585.

基于改进 AlignedReID++的肉牛个体重识别方法

应潇溢¹, 赵继政², 杨玲玲³, 周馨怡³, 王磊³, 高延年³,
咎林森⁴, 杨武才⁴, 刘含³, 宋怀波^{3*}

(1. 西北农林科技大学信息工程学院, 杨凌 712100; 2. 西安理工大学计算机科学与工程学院, 西安 710000; 3. 西北农林科技大学机械与电子工程学院, 杨凌 712100; 4. 西北农林科技大学动物科技学院, 杨凌 712100)

摘要: 准确和持续的牛只个体识别对于精准养殖具有重要意义。中远距离和跨摄像头场景下的牛只个体识别是监测个体采食量和进食时间的基础。肉牛具有在采食过程中频繁移动和改变采食位置的特点, 牛只方向变化频繁, 加之牛只个体的生物相似性以及复杂的环境变化(光线、遮挡和背景), 导致跨摄像头牛只个体识别困难。该研究基于 AlignedReID++模型可充分利用全局信息和局部信息进行高效图像匹配的优点, 并在此基础上进行了改进, 以实现更优的牛只个体重识别效果。在改进模型中, ResNet50 主干网络的 BottleNeck 结构应用了三重注意力机制模块, 以实现在引入少量参数量的情况下, 通过跨维交互, 加强模型对于个体图像的特征提取能力; 基线模型的全局分支中的交叉熵损失函数被替换成 CosFace 损失函数, 并与困难三元组损失函数共同训练改进后的模型, 以提升模型分辨相似个体的能力。改进模型的 rank-1 准确率为 94.42%, 平均精度均值为 63.90%。与基线模型相比, rank-1 准确率提高了 1.84 个百分点, 平均精度均值提高了 6.42 个百分点。与 PCB (part-based convolutional baseline) 相比, 在 rank-1 准确率和平均精度指标上分别超出了 4.72 和 5.86 个百分点。与 MGN (multiple granularity network) 相比, 在 rank-1 准确率和平均精度均值指标上分别超出了 0.76 和 4.30 个百分点。与 TransReID 相比, 在 rank-1 准确率指标上低了 0.98 个百分点, 在平均精度均值指标上超出了 3.90 个百分点。与 RGA (relation-aware global attention) 相比, 在 rank-1 准确率和平均精度均值指标上分别超出了 5.36 和 7.38 个百分点。此外, 改进模型的浮点运算次数为 5.45 G, 仅比基线模型大 0.05 G, 分别比 PCB、MGN、RGA 和 TransReID 小 0.68、6.51、25.4 和 16.55 G。同时, 改进模型的模型大小为 23.78 M, 在对比模型中, 其模型大小是最小的。改进模型在 CPU 上的推理速度为 5.64 帧/s, 低于 PCB、MGN 及其基线模型, 高于 TransReID 和 RGA。t-SNE 特征嵌入可视化结果显示, 改进模型提取的个体样本的全局特征和局部特征可以实现良好的类内紧凑性和类间差异性。本研究结果表明, 所提出的方法能够有效地重识别自然养殖场景下的肉牛个体, 对个体采食量和进食时间的监测具有较好的指导意义。

关键词: 方法; 识别; 肉牛; 精准畜牧; 重识别; AlignedReID++; 深度学习