An Intelligent Edge-IoT Platform With Deep Learning for Body Condition Scoring of Dairy Cow

Junhao Wang, Baisheng Dai, Yang Li, Yongqiang He, Yukun Sun, and Weizheng Shen

***Abstract*—Body condition score (BCS) of dairy cows is the direct reflection of their nutritional status. The timely estimation of BCS is beneficial to improving dairy cow health, milk production, and reproduction. In this work, we propose an intelligent Edge-IoT platform with deep learning for estimating BCS of dairy cow, by integrating inference capability of deep learning and low latency of edge computing in IoT framework. Through capturing images of dairy cow's back with the RGB-D camera, the inference module deployed in the edge computing device first performs cow detection to localize the separate area of each dairy cow and then performs individual identification and estimating BCS of dairy cows simultaneously. The existing systems are mainly commercial systems, such as DeLaval and HerdVision, they use electronic ear tags with radio-frequency identification sensors for cow identification. Compared to existing systems, in the proposed platform, combined the finetuned YOLOv7 model and avoid repeated inference (ARI) algorithm to detect dairy cow. An EfficientID model combined with metric learning is designed for cow identification, and an EfficientBCS model with coordinate attention (CA) is proposed for estimating BCS. The dairy cow's identity (ID) and BCS are finally transmitted to the cloud analysis center. Experimental results show that the accuracy of estimating BCS reached 85% within 0.5 range error conducted on the test set collected in the dairy farm. The total inference time for one dairy cow is 3.138 s. Results show that the platform can be served as an excellent application of dairy cow body condition scoring.**

***Index Terms*—Application platform, body condition scoring, deep learning, edge computing.**

Manuscript received 13 July 2023; revised 26 September 2023 and 7 December 2023; accepted 19 January 2024. Date of publication 24 January 2024; date of current version 9 May 2024. This work was supported in part by the Key Research and Development Program of Heilongjiang Province under Grant 2022ZX01A24; in part by the National Natural Science Foundation of China under Grant 32072788; in part by the Modern Agricultural Industry Technology Collaborative Innovation Promotion System Construction Project of Heilongjiang Province [the Letter of Department of Agriculture and Rural Affairs of Heilongjiang Province (2021)] under Grant 1492; and in part by the Earmarked Fund for CARS under Grant CARS-36. *(Corresponding authors: Baisheng Dai; Weizheng Shen.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Animal Care Advisory Committee of Northeast Agricultural University.

Junhao Wang, Baisheng Dai, and Weizheng Shen are with the College of Electrical Engineering and Information, Northeast Agricultural University, Harbin 150030, China (e-mail: bsdai@neau.edu.cn; wzshen@neau.edu.cn).

Yang Li and Yukun Sun are with the College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China (e-mail: liyang1405053@neau.edu.cn; sunyukun@neau.edu.cn).

Yongqiang He is with the Milk Source Quality Technology Center, Mengniu Company Ltd., Hohhot 010000, China (e-mail: heyongqiang@mengniu.cn).

Digital Object Identifier 10.1109/JIOT.2024.3357862

I. Introduction

In recent years, deep learning has demonstrated excellent inference capabilities in many applications such as smart livestock farming [1], [2], [3]. The existing deep learning models usually have millions of parameters and rely on powerful computational capability. Edge computing can place the inference close to end devices to meet the high computation requirements of deep learning and reduce data transmission delay. By combining deep learning and edge computing, it is possible to achieve an accurate and real-time inference on specific tasks in an IoT framework, e.g., monitoring the health of dairy cows.

Body condition score (BCS) reflects dairy cow's body fat reserves and it is not related to their body size and body weight [4]. Its estimation is usually based on appearance of tissue coverage on bovine back and pelvic bone protrusions [5]. A 5-point scale is commonly used to estimate the BCS of dairy cow [5]. In increments of 0.25, a BCS score of 1 means the dairy cow is emaciated while a score of 5 means it is obese. Traditionally, BCS has often been estimated manually by experts [6]. However, this process is time-consuming, highly subjective, and could easily cause stress to dairy cows. Therefore, there is an urgent demand for automatic and noncontact estimation methods of BCS.

In order to automatically acquire accurate BCS and make the score correspond to specific dairy cow, two modules, i.e., the object detection and individual identification of dairy cow are prerequisite for body condition scoring. Object detection of dairy cow can separate dairy cows from the background in the image. Individual identification of dairy cows can accurately find out which dairy cow is being scored. In farms, plastic ear tags with numbers as well as electronic ear tags with radio-frequency identification (RFID) sensors are currently the most common methods to distinguish individual dairy cows [7]. However, the reading/writing distance of RFID is limited. Ear tags are intensive and tend to get lost over time. In addition, the management or installation of RFID system needs skilled personnel. They will bring invasive damage to dairy cows. To achieve automatic and accurate estimation of BCS in the platform, it is necessary to develop an efficient object detection method and noncontact individual identification method of dairy cows.

Existing systems are mainly commercial systems, and the literatures mainly focused on methods for BCS estimation. The commercial systems includes DeLaval system [8], [9], HerdVision system [10], etc. Both of them can achieve automatic estimation of BCS in the farm. However, they use ear tags with RFID to identify each cow instead of cameras. Most of the literatures focused on improving the accuracy of estimating BCS based on cloud computing and did not form a complete system. It needs to be supplemented with the necessary modules (cow detection, identification, etc.) in farm deployment, as well as considering the computational and transmission latencies of the proposed method.

This work proposed an intelligent Edge-IoT platform to estimate BCS of dairy cow using deep learning. In the platform, a RGB-D camera is served as the end device to collect dairy cows' back images, and an intelligent embedded device integrated inference modules is used as the edge computing device. The captured images are used in the edge computing device to perform three modules including object detection, individual identification, and body condition scoring of dairy cow. First, using the finetuned YOLOv7 model combined with avoid repeated inference (ARI) algorithm to detect multiple cows in real-time. Second, to avoid repeated training caused by new dairy cows in the farm, a feature extraction and metric learning framework for individual identification of dairy cow is proposed. Third, a deep learning model with spatial attention is improved to estimate BCS of dairy cow. Finally, the dairy cow's ID and BCS inferred by the edge computing device will be transmitted to the cloud analysis center via WIFI, Ethernet, and 4G/5G mobile network. In addition, the proposed platform can provide users with enhanced Quality of Service (QoS). High-latency and computationally intensive tasks in the platform are offloaded to edge servers, and one edge computing device only works in one farm to avoid QoS deterioration caused by user allocation. Particularly, the metric learning can avoid repeated training caused by data changes in deep learning models. It can save time and labor, and maintain QoS during platform operation.

The main contributions of this article are as follows.

1) An intelligent Edge-IoT platform based on deep learning is introduced for estimating dairy cow's BCS.
2) The finetuned lightweight YOLOv7 model combined with the ARI algorithm is designed to detect each of dairy cow.
3) A feature extraction and metric learning framework is proposed for identifying dairy cow.
4) The improved EfficientBCS model with spatial attention is used to estimate BCS.

The remaining sections of this article are organized as follows. Section II reviews the related work of dairy cow's body condition scoring, object detection, and individual identification. Section III presents the proposed platform. Section IV gives the experiments of the platform. Section V provides the conclusions and future perspectives.

II. Related Work

In this section, we provide a brief overview of work related to dairy cow's body condition scoring, object detection, and individual identification.

A. Body Condition Scoring

With the development of computer vision technology, automatic scoring methods of dairy cow body condition have been researched. According to the different ways of feature extraction, these noncontact BCS scoring methods can be divided into manual feature extraction-based and deep learning feature extraction-based. In the first scoring method, the BCS is obtained through regression analysis of metrics, such as convexity and body length of key parts of the dairy cow. Azzaro et al. [11] labeled the dairy cow's back image with manual anatomical points. In this method, linear and polynomial kernel principal component analysis was used to reconstruct the shape of the dairy cow to evaluate the BCS. Halachmi et al. [6] localized the contour of dairy cows in thermal images and extracted features based on deviation of the contour from its fitted parabola to estimate BCS. Spoliansky et al. [12] proposed a depth image-based BCS evaluation method that constructs a regression model after preprocessing. This method used the depth information to reflect the fat reserves. However, this traditional manual feature extraction is highly subjective and not conducive to automated dairy cow body condition scoring.

Due to the excellent automatic inference ability of deep learning, the deep learning feature extraction-based method can achieve efficient BCS estimation with less human intervention. The method based on deep learning only needs to manually select the key area for evaluation, and then send the image into the deep learning model to automatically obtain the score. Liu and Qin [13] collected the posterior and lateral posterior RGB image of dairy cows and used improved VGG-16 with Convolutional Block Attention Module (CBAM) mechanism to estimate BCS. Sun et al. [14] used the deep learning framework to extract features related to body condition for an automatic estimation of BCS, which based on the image consisted of depth, gray, and phase consistency of dairy cow image. Alvarez et al. [15] acquired the dairy cow's back contour profile map, Fourier transform map and depth map fusion from the original depth image and fed it into a convolutional neural network for automated BCS estimation. Zhao et al. [16] constructed 3-D structure feature maps based on convex hull distance of point clouds from the top-view image of cows and realized estimation based on EfficientNet network. Shi et al. [17] proposed an attention-guided PointNet++ network to extract 3-D features from point clouds and estimate BCS automatically. The above methods achieved accurate dairy cow body condition scoring. However, both of them require additional selection of regions of interest such as tail concavity and body contour of dairy cows, which is not efficient enough for estimating BCS in edge environment.

B. Object Detection

The object detection technology in deep learning can quickly and accurately locate the dairy cow, which is easy to implement and has high robustness. Object detection models are mainly divided into one-stage and two-stage models, which the former focuses on accuracy and the latter on speed. Gardenier et al. [18] used the two-stage object detection model Faster R-CNN to detect dairy cow hooves and predict whether it is lame. Yılmaz et al. [19] used the one-stage object detection model YOLOv4 to locate the torso of cows and achieve breed classification. Huang et al. [20] improved the one-stage object detector SSD to locate the key body parts such as tails, pins, and rump of the cow, then realized the estimation of dairy cow's BCS. With the development of one-stage object detectors, the accuracy gradually approaches the two-stage models and they have faster inference speed. One-stage object detectors are gaining more and more attention in dairy smart farming.

C. Individual Identification

The computer vision techniques have been developed significantly in recent years, research on noncontact individual identification methods of dairy cow has emerged. There are currently four areas used for cow identification: 1) muzzle; 2) face; 3) trunk; and 4) back. Kusakunniran et al. [21] used histograms of multichannel local binary patterns to extract features from dairy cow muzzle images and used SVM for individual classification. However, due to the small size of cow's muzzle, it is difficult to detect in practical applications. Kawagoe et al. [22] used YOLO detector to extract the face region image of cow and utilized transfer learning to train a deep neural network to recognize individual cows. It is difficult to achieve stable identification because cows often shake their heads. The majority of dairy cows in farms is Holstein cows, which have different distributions of black and white patterns. Andrew et al. [1] used images of cows from the back view to identify identification of cows-based the InceptionV3-based biometric Long-term Recurrent Convolutional Network. Shen et al. [23] proposed a noncontact dairy cow identification method in the side-view image, which uses the YOLO model to detect dairy cow object and then classify each cow by a convolutional neural network. Xiao et al. [24] utilized improved Mask R-CNN to segment the top-view image of cows and extracted the shape features of the cow's back, then SVM was applied to identify individual cows.

The method of individual identification of dairy cow based on deep learning can achieve automatic and efficient inference. However, when new dairy cows appear in the farm, the deep learning model must to be retrained and it is not suitable for practical application.

III. Proposed Platform

A. Platform Architecture

As shown in Fig. 1, the platform includes end device, edge computing device, and cloud analysis center. The end device is an RGB-D camera, which is used to capture RGB and depth video streams of dairy cows' back. The streams are transmitted to the end computing device via USB. The edge computing device integrated three modules, including object detection, individual identification, and body condition scoring of dairy cows. After the inference of deep learning modules, the ID and BCS are uploaded to the cloud analysis center. The cloud analysis center mainly includes a cloud database and Web pages, which is used to store and visually analyze inference results transmitted by edge computing devices.

B. Edge Computing Device Work Flow

The program on the edge computing device is multithreading, including a main thread and three child threads. The main thread is the interaction thread, which mainly receives video streams from end device and transmits inference results to the cloud platform. Three child threads include an object detection module combined with the finetuned YOLOv7 [25] model and ARI algorithm, an individual identification module combined with the EfficientID model and metric learning, and body condition scoring module based on the EfficientBCS model of dairy cow. The edge computing device and integrated modules are set to start automatically when plugged in.

Particularly, the object detection module is performed on the RGB video stream. For the current frame of the video stream, if the confidence of the output inferred by the YOLOv7 model is greater than 0.5 and the horizontal coordinates in the output are all greater than 0, the edge computing device considers a complete dairy cow captured. Then, the current frame is judged by the ARI algorithm, and only the dairy cows which are not inferred will be performed subsequent modules. The RGB and depth frames would be cropped according to the coordinates of output and resized proportionally to 384×384 for individual identification and estimation of BCS. The RGB frame is extracted by the EfficientID model as a feature vector to identify the dairy cow. The extracted feature vector and the feature vector in the template library are calculated the feature similarity. If the result shows that this dairy cow is in the template library, its ID would be recorded. Otherwise, this feature vector would be registered to the template library and the temporary ID given by the platform would be recorded. The depth frame is first processed with height thresholding and then inferred by the EfficientBCS model. The dairy cow's BCS would be transmitted to the cloud platform along with the ID.

C. Object Detection Module of Dairy Cow

In order to realize real-time detection of multiple dairy cows on the edge computing device, the one-stage YOLOv7 model is used. The dairy cow detection model (YOLOv7) is shown as Fig. 2. The basic block of the model, CBL (Conv, Batch Normalization, Leaky ReLU), consists of 3×3 convolution, batch normalization, and Leaky ReLU activation function. ELAN structure and MP structure are added to the backbone and head networks. ELAN is divided into two branches, branch one adjusts the number of channels by 1×1 convolution, branch two adjusts the number of channels first, and then after the stacked 3×3 convolution for feature extraction, finally aggregates with branch one. MP structure uses both MaxPooling and 3×3 convolution with a step size of 2 to achieve dimensionality reduction.

Existing cow detection methods [26], [27], [28] can detect specific parts of cows in an experimental environment and perform subsequent processing. In the deployment environment, dairy cows usually tend to appear in groups. For the same dairy cow, only one inference for identification and BCS is required. However, the object detector cannot distinguish whether this dairy cow has been inferred or not. It will

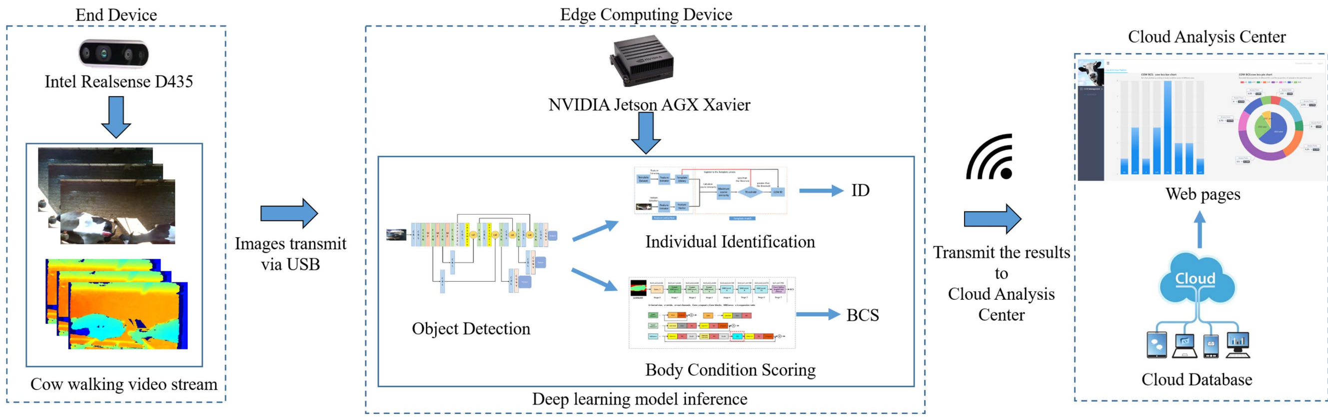


Fig. 1. Architecture of the platform.

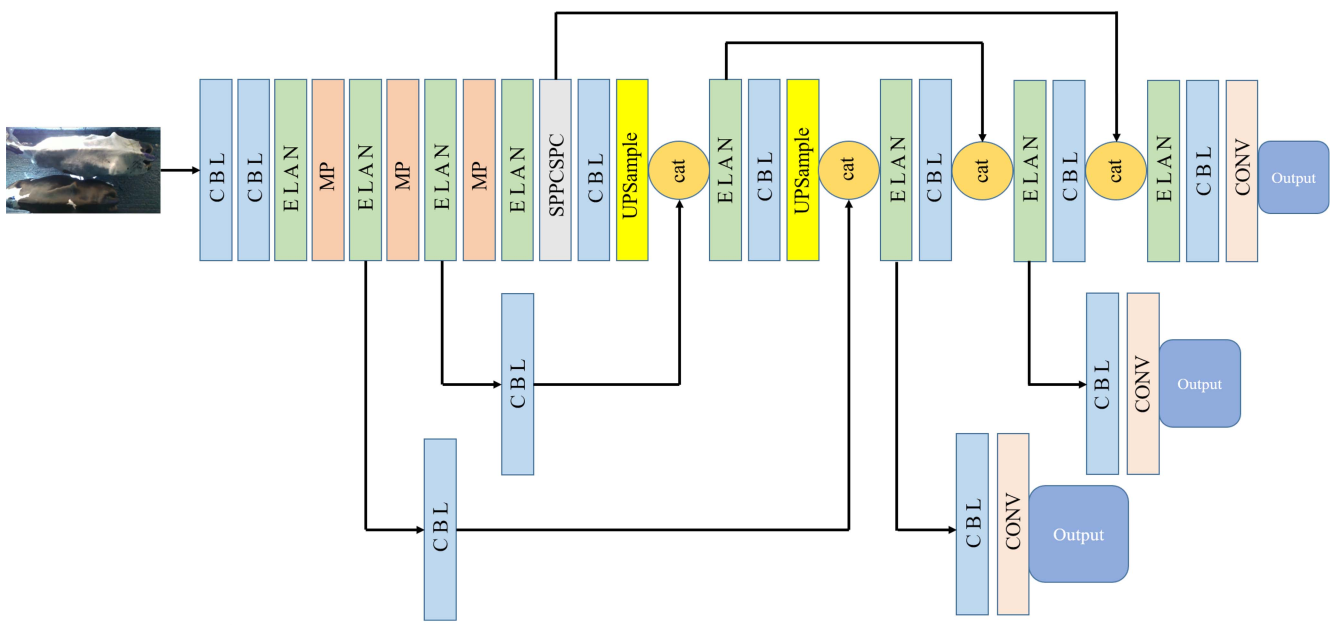


Fig. 2. Dairy cow detection model structure.

continuously detect the coordinates of this cow and feed them to the subsequent modules for repeated inference, which would severely waste computing resources. Therefore, we proposed the ARI algorithm to distinguish the temporary identity of the dairy cow during detection to ensure that each cow will be inferred by the subsequent models but only once.

The algorithm is as shown in Algorithm 1. Input is the coordinate of the object detection output box and the distance threshold is set to 50. List COW is used to store the dairy cow information that has been detected. Each element *COWn* stores two fields, where $n = 0,1,\ldots, N$, which are the dairy cow's temporary ID k (counting from 0) and the center coordinates of the object detection output box. The number of object detection output box is q, where $q = 0,1,\ldots,Q$. All of the inferred coordinates of boxes stored in the list LOC. When an image is detected, the LOC is traversed. If COW is empty, the temporary ID k and the coordinates LOCq are appended to the COW. Otherwise, traverse the COW and calculate the spatial distance between LOCq and each of the existing coordinates of the COW. The spatial distance equation is shown in (1) below

$$\text{Distance}=\sqrt{(\text{LOCq}_x - \text{COW}^1_{nx})^2 + (\text{LOCq}_y - \text{COW}^1_{ny})^2}. \quad (1)$$

If the spatial distance between *COWn* and LOCq is less than the threshold, the subscript n and spatial distance are stored in the list *COW_NUM*. If there is more than one result below the distance threshold, it means that the dairy cow in the current frame is close to the position of several dairy cows in the previous frame, and the dairy cow's ID cannot be effectively distinguished by the threshold alone. If *COW_NUM* is empty, it means that there are no dairy cows that less than the distance threshold in the previous frame, and the dairy cow in the current frame is new. Otherwise, the elements in *COW_NUM* are sorted by the second dimension, and the dairy cow with the smallest distance is considered to be the same dairy cow as the current one. The center coordinates of this dairy cow in COW are updated. The temporary ID of the current dairy cow

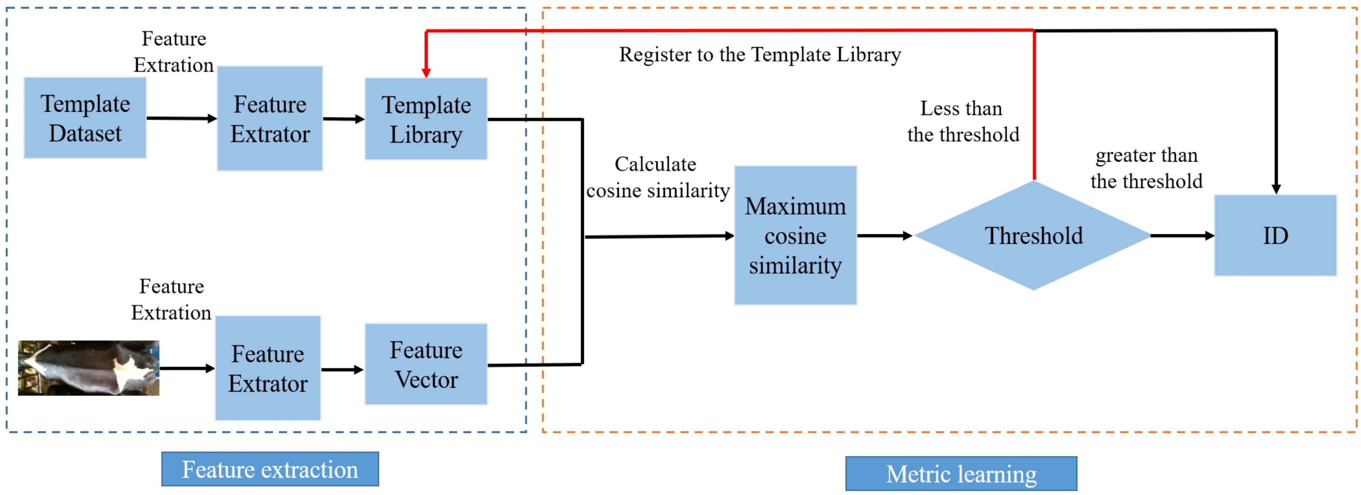


Fig. 3. Individual Identification framework of dairy cow.

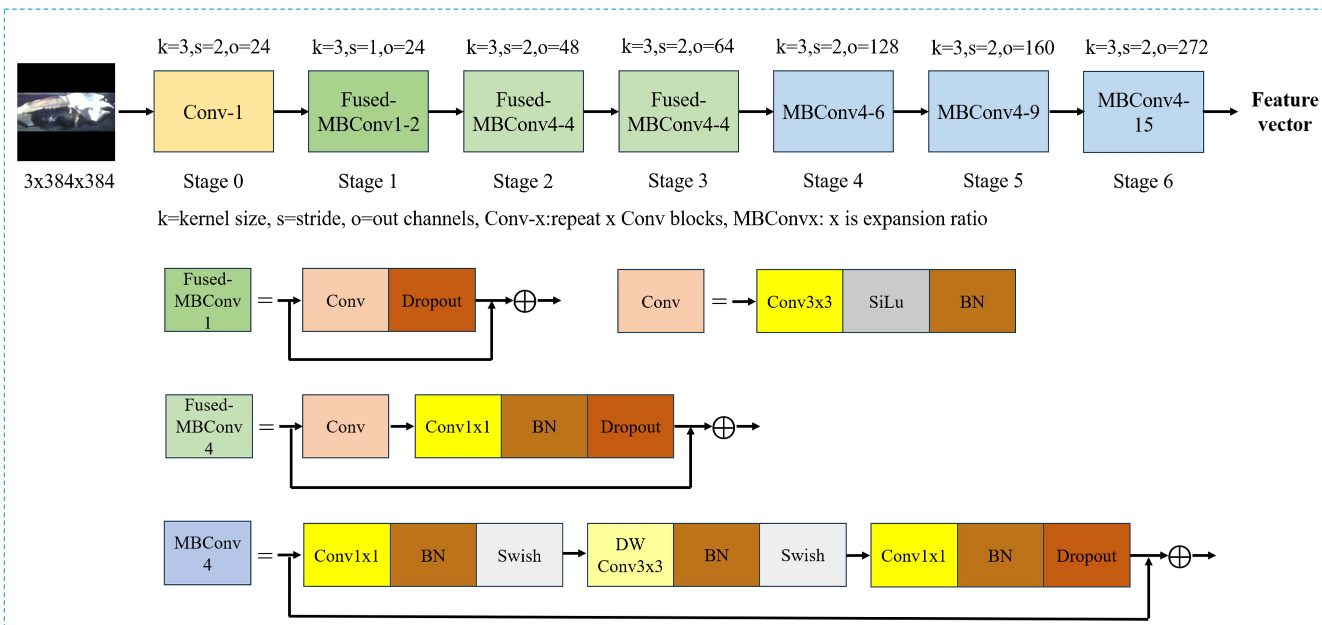


Fig. 4. EfficientID model structure.

is queried in the *COW_NAME*. If it already exists, it means that this dairy cow has been inferred by the rest modules. Otherwise, the rest modules are executed and the temporary ID is stored in *COW_NAME*.

D. Individual Identification Framework of Dairy Cow

Most of the existing methods regard the individual identification of dairy cow as a classification task. However, if there is a new dairy cow to be identified the deep learning model must to be retrained. This problem will bring a large training cost, and it is not conducive to the actual deployment. To solve it, a framework was proposed for individual identification of dairy cow based on the EfficientID model combined with metric learning. Fig. 3 shows the proposed framework, which is mainly divided into two steps: 1) feature extraction and 2) metric learning.

The feature extractor is first trained, and then the images in the template library are extracted by the feature extractor and stored as feature vectors in the template library. The images to be detected are extracted by the feature extractor and then the feature similarity is calculated with the feature vectors in the template library. If the largest value is higher than the threshold, the detected dairy cow is considered to belong to one of the dairy cows in the template library. Otherwise, this dairy cow is given a new ID and the feature vector is registered in the template library.

1) Feature Extractor: An EfficientID model is proposed based on the EfficientNetV2 [29] as shown in Fig. 4. The squeeze-and excitation (SE) channel attention [30] in EfficientNetV2 will downscale and then upscale the feature vector, and this operation will significantly reduce the inference speed and increase the parameters. Therefore, EfficientID

Algorithm 1 ARI Algorithm

Input: Output coordinates for object detection $LOC \in R^{1xQ}$
Output: Boolean type False, False means no follow-up detection

```
Initialize k = 0,Mindistance=50, List COW_NAME∈R^{2xM},List COW∈R^{2xN}
for q in {1,2, ...Q} do
  Initialize List COW_NUM∈R^{2xW}
  if (length of COW equal with 0) then
    Append k and LOCq to COW
    k = k+1
  else
    for n = 1, 2, ...N do
      Calculate the distance LOCq and the coordinate of COWn by (1).
      if (distance less than Mindistance) then
        Append n and distance to COW_NUM
      end if
    end for
    if (length of COW_NUM equal with 0) then
      Append k and LOCq to COW
      k = k+1
    else
      Sort by the second dimension of COW_NUM in ascending order
      Update the COWn with LOCq
    end if
  end if
  if (k not in COW_NAME) then
    Append k to COW_NAME
    return False
  end if
end for
```

removes the SE channel attention and keeps the other structures of the EfficientNetV2.

To enable the proposed model to achieve faster inference and occupy less memory, the model is pruned. Using the L1norm-based channel pruning technique [31], the trained EfficientID model is first pruned and then fine-tuned so that the model gradually recovers to close to the original accuracy.

The basic component blocks of EfficientID are MBConv and Fused-MBConv. MBConv first performs 1×1 convolution, batch normalization, Swish activation function, and then uses a convolution kernel of three depth-separable convolutions to reduce computational complexity. Fused-MBConv includes 3×3 convolutions and uses the SiLu activation function. The convolution type (MBConv and Fused-MBConv), the number of layers, convolution kernel size, and expansion factor for each layer are obtained from neural architecture search (NAS) technology. The model extracts the feature vectors output from Stage 6 for next processing.

2) Metric Learning: The feature vectors of the image to be detected are compared with the feature vectors in the template library one by one to calculate the feature similarity. The cosine similarity is the cosine of the angle between two n-dimensional vectors in the same vector space. It is equal to the product of the dot product (vector product) of the two vectors divided by the product of the length (or size) of the two vectors. The closer the cosine value is to 1, the closer the angle is to 0 degrees, indicating the more similar the two vectors are. The closer the cosine value is to 0, the closer the angle is to 90 degrees, indicating the less similar the two vectors are. The equation for calculating the cosine similarity is as follows:

$$\cos(\theta)_j = \frac{\sum_{i=1}^{n}\left(T(I)_i \times T\left(F_j\right)_i\right)}{\sqrt{\sum_{i=1}^{n}\left(T(I)_i\right)^2} \times \sqrt{\sum_{i=1}^{n}\left(T\left(F_j\right)_i\right)^2}} \tag{2}$$

where $i = 1, \ldots, n$, $j = 1, \ldots, N$. T represents the feature extraction function, I is the unknown dairy cow feature vector to be detected, F_j represents each known cow feature vector in the template library, and i represents the dimensionality of the feature vector. $\cos(\theta)_j$ represents the cosine similarity calculated from the dairy cow feature vector to be tested and the feature vector in the template library.

E. Body Condition Scoring Model of Dairy Cow

The RGB-D camera captures depth information from the dairy cow's back, effectively showing the undulation of the dairy cow's back. The undulation information reflects the fat reserves of key parts of the back, which in turn reflects the body condition of the dairy cow. To reduce noise interference in the depth image, a height thresholding strategy is used to traverse the depth image pixel points for processing. The normal body height of lactating dairy cows is approximately between 110–140 cm [32], so gray values greater than 1600 and less than 900 were assigned 0 values.

The EfficientBCS model is designed to estimate BCS of dairy cow which treats the dairy cow body condition scoring as a classification task. Instead of manually selecting regions of interest for scoring, the platform uses an attention-guided approach based on spatial coordinates to instruct the model to focus more on attention regions in the depth images. Coordinate attention (CA) [33] is a novel spatial attention mechanism for mobile networks that implements the embedding of location information into channel attention, thus unifying channel and spatial attention without incurring significant computational overhead.

As shown in Fig. 5, CA aggregates the input feature maps along the vertical and horizontal directions into two separate feature maps with orientation attributes using two global averaging pooling operations, respectively. The feature maps with embedded orientation-specific information are then encoded into two attention maps, each of which captures the long-range correlation of the input feature maps along a spatial direction. Thus, the spatial location information can be stored in the generated attention maps. The two attention maps are then applied to the input feature maps by vector multiplication to emphasize the spatial information representation of interest.

As shown in Fig. 6, the input of the EfficientBCS model is a single channel depth image with an image size of $1\times384\times384$. The main modules are basically the same as the EfficientID model. In Stage 7, after 1×1 convolution for channel switching

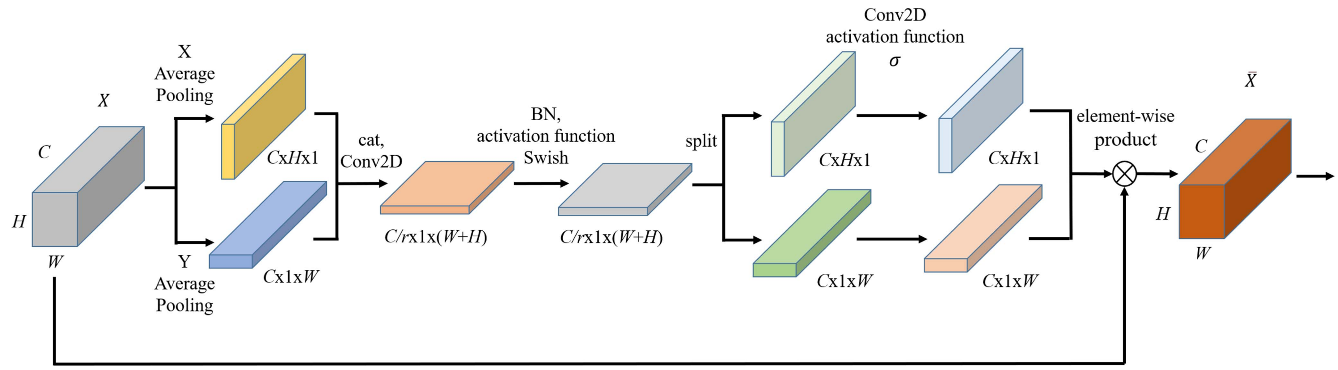


Fig. 5. CA.

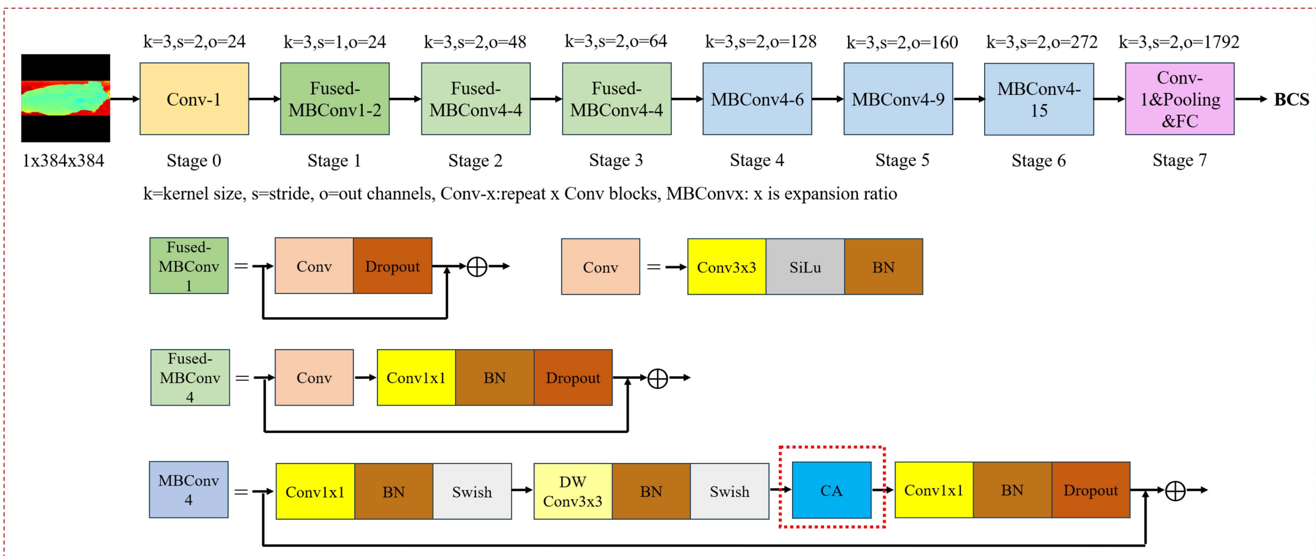


Fig. 6. EfficientBCS network structure. Model input is a single-channel depth image, which is displayed using a pseudo-color image.

and global average pooling for dimensionality reduction, the feature vector is expanded to a 1-D vector by a fully connected layer. Finally, the softmax outputs the BCS.

CA is added to the MBConv module in the deeper layers of the network because the shallow feature layers contain a lot of noisy information.

The EfficientBCS model is not pruned because the estimation of BCS is more difficult than the individual identification. Moreover, the bottleneck of the platform's inference speed is in the individual identification module, there is no need to sacrifice the accuracy of the body condition scoring to achieve higher inference speed.

F. Cloud Analysis Center

In order to allow farmers to make different nutritional adjustments to corresponding dairy cows according to their BCS inferred from the edge computing device, a cloud analysis center was developed. It can visualize real-time inference results, and save data to a cloud database. It is a cross-platform application, compatible with mainstream operating systems, such as MS Windows, Mac OS. The cloud analysis center is built upon SpringBoot framework, which interacts with the database through MyBatis protocol. The front-end Web pages is built by HTML, CSS, and Javascript. The inference results from the edge computing device is decomposed into three data fields, Cow ID, Cow BCS, and Upload Time, and then stored in a MySQL database. The Web pages used Echarts and Bootstrap to visually analyze the results.

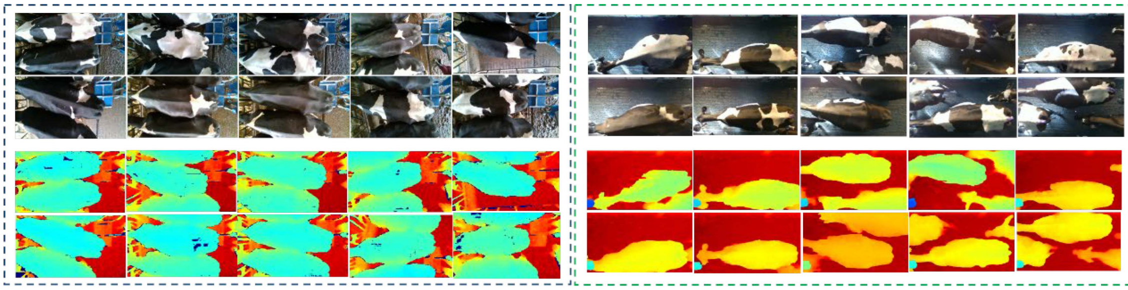
IV. Experiments

A. Data Set

The data sets of the platform include data set A and data set B, in which data set A is used to train and test the deep learning model and data set B is used to evaluate the performance of platform.

Data set A was collected at the Nestlé Dairy Cow Breeding Training Center in Shuangcheng, Harbin, Heilongjiang Province, China. The camera was placed on a fixed aluminum tube 3.2 m above the floor in the barn, and the resolution of the RGB and depth video streams is 1280×720 at a frame rate of 30. 5675 RGB and depth images were collected from 182 cows.

Data set B was acquired from Shengkang Herding in Lindian County, Daqing City, Heilongjiang Province, China. The acquisition camera was located above the floor of the milking parlor passage at a distance of 2.8 m. The resolution of the RGB and depth video streams is 848×480 and the frame rate is 60. 76 dairy cows were captured and a total of 10 802



(a) (b)

Fig. 7. Data sets of the platform. (a) Data set A (used for model training and testing). (b) Data set B (used for platform testing).

TABLE I
EXPERIMENTAL DATA DISTRIBUTION OF THE PLATFORM (DATA IN BRACKETS WILL BE MERGED)

	Train set	Test set	TL set	TH set	TT set	Dataset	Total
Object detection	4540	1135	-	-	-	A	182 cows, 5675 images
Feature extraction	2480	620	-	-	-	A	100 cows, 3100 images
Metric learning	-	-	1550 (+1440)	2990 (-2990)	2685 (+1550)	A	182 cows, 5675 images
Body condition scoring	627	184	-	-	-	A	116 cows, 811 images
Platform test	-	3825	-	-	-	A, B	208 cows, 3825 images

RGB and depth images were obtained by resolving the video streams.

All of the dairy cows in the data sets are Holstein cows. The ear tag numbers of the dairy cows were recorded during the two collections to identify the dairy cows, and BCS of dairy cows was estimated by the experts. The data sets are shown in Fig. 7, the first two rows are RGB images, and the last two rows are pseudo-colored images of the depth images for display.

Data set A is divided for different modules as shown in Table I. Data set B was not used when fine-tuning deep learning models, but was tested during platform testing.

1) Object Detection Module: 5675 RGB images of 182 dairy cows were used, and the data set is divided into a train set and a test set in an 8:2 ratio.

2) Individual Identification Module: First, 3100 RGB images of 100 dairy cows were used to train the EfficientID model which is divided into a train set and a test set in an 8:2 ratio. Then, the data sets in this module were divided into three sets: 1) TL set to extract feature vectors to template library; 2) TH set to calculate the threshold; and 3) TT set to evaluate the performance of identifying dairy cows which are not sure if they already exist in the template library. TL set includes 2990 RGB images of 150 dairy cows (among them, the 1440 RGB images of 50 dairy cows are supplemented by the TH set after calculating the threshold). TH set includes 2990 RGB images of 150 dairy cows. TT set includes 2685 RGB images of 132 dairy cows (among them, the 1550 RGB images of 100 dairy cows are supplemented by the TH set after calculating the threshold).

3) Body Condition Scoring Module: A total of 811 depth images were obtained by selecting seven images from each of 116 dairy cows. The images were randomly divided into train and test sets according to the ratio 8:2, with 627 images for training and 184 images for testing.

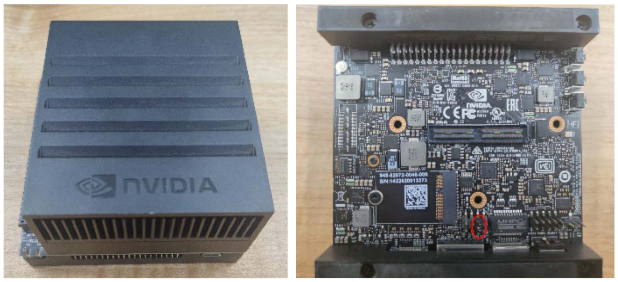
4) Platform Testing: Th 76 dairy cows in the data set B each select ten RGB images and five depth images, and all the images in the TT set, a total of 3825 images of 208 dairy cows.

B. Implementation

The models were trained on an NVIDIA RTX 3090 with 24 GB. Before deep learning model training, normalize the data sets by calculating the mean and variance, and use pretrained models for transfer learning to accelerate model convergence.

The end device is implemented on an Intel RealSense D435 depth camera, whose implementation principle uses structured light solution, suitable for application in outdoor scenes. The resolution of the RGB video stream is up to 1920×1080, and the depth resolution video stream is up to 1280×720, with a maximum frame rate of 90 FPS and a binocular detection range of 0.2–10 m. Edge computing device is implemented on NVIDIA Jetson AGX Xavier, as shown in Fig. 10. Its GPU contains 512 CUDA cores and 64 Tensor cores to accelerate deep learning inference. The NVIDIA Jetson AGX Xavier provides shorted pins 5 and 6 for power-on, and the shorted location is circled in red in Fig. 8.

Both training and testing using the CUDA version 10.2, cuDNN version 8.2.2, Pytorch 1.8 for GPU, and OpenCV 2.1. The cloud analysis center is built by JAVA and the communication protocol between edge computing device and cloud analysis center is Apache HttpClient.



(a) (b)

Fig. 8. End computing device. (a) NVIDIA Jetson AGX Xavier. (b) Shorting pin schematic.

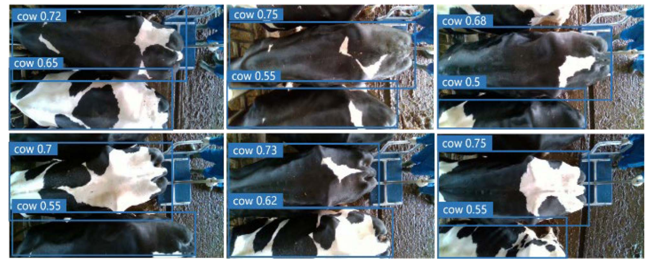


Fig. 9. Detection results of the YOLOv7 model on data set A.

C. Evaluation Metrics

All experiments of the platform used a fivefold cross-validation strategy, which divides all data equally into five data sets. These five data sets are traversed sequentially, and the current data set is used as the validation set, the remaining four are used as the training set. Finally, the average of the five evaluations is used as the final evaluation metric. We use the following evaluation metrics in our experiments.

1) mAP: Mean average precision (mAP) is the average of the average precision (AP) for all sample categories. AP is calculated as the size of the area enclosed by the 2-D curve with precision as the vertical axis and recall as the horizontal coordinate and the area enclosed by the coordinate axis

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

where true positives (TPs) is the number of positive samples predicted by the model with positive true values, false positives (FPs) is the number of positive samples predicted by the model with negative true values, and false negatives (FNs) is the number of negative samples predicted by the model with positive negative true values. Precision is the proportion of all positive samples predicted by the model as shown in (3), and Recall is the proportion of all positive samples predicted by the model out of all positive samples as shown in (4). The mAP@0.5 is the value of mAP when the IOU threshold is higher than 0.5, and the mAP@0.5:0.95 is the mean value of mAP in different situations where the IOU threshold is calculated from 0.5 to 0.95 (step size is 0.05).

2) Accuracy: Accuracy is the ratio of all positive and negative samples to all samples with correct predictions. True negatives (TNs) are the number of negative samples predicted by the model with negative true values. The equation of Accuracy is shown as follows:

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \tag{5}$$

3) ROC Curve: TP rate (TPR) is defined as the ratio of predicted positive cases and actual positive cases to all positive cases (true label is positive), and FP rate (FPR) is defined as the ratio of predicted positive cases and actual negative cases to all negative cases (true label is negative). The receiver operating characteristic (ROC) curve is plotted geometrically with FPR as the horizontal coordinate and TPR as the vertical coordinate. The area enclosed by the ROC curve and the axis is the area under curve (AUC) value. The larger the value, the more accurate the prediction of the model, and the maximum value is 1.

D. Results and Discussion

1) Performance of Object Detection Module: In order to evaluate the performance of the YOLOv7 model in detecting multiple dairy cows on the data set A, and compare it with other existing object detection models, including YOLOv5 model, the two-stage object detection model Faster R-CNN [34], and DETR [35] based on the Transformer. The detection results of YOLOv7 model are as shown in Fig. 9. The YOLOv7 model can effectively detect complete and partially visible dairy cows and provide different confidence. The compared results are shown in Table II. The YOLOv7 model achieves Precision of 98.9%, Recall of 99.4%, and mAP of 98.3% which is the best compared with other models. Since the ELAN structure in the YOLOv7 model can learn both long and short paths at the same time, it still maintains the learning generalization ability of the model in the deep network. The detection speed of the YOLOv7 model is 46fps and the params is only 12.3M. It can easily deployed on the edge computing device to detect multiple dairy cows.

2) Performance of Individual Identification Module: The experiment was performed on data set A to evaluate the impact of model improvement and pruning on model performance, inference time, and params. The results is as shown in Table III. The compared models include EfficientNetV2, EfficientID, as well as 30% and 50% pruning of EfficientID. First, EfficientID is compared with EfficientNetV2, the results show that the SE channel attention is not effective on the dairy cow data set. There is no obvious change in accuracy, but params and inference speed of EfficientID are reduced. Second, the 30% pruning EfficientID lost 0.8% accuracy compared to unpruned EfficientID, but the inference speed and params were further reduced. The decrease of accuracy is almost negligible caused by 30% pruning. Then, in order to evaluate the impact of the amount of pruning on the model, 50% pruning of EfficientID was also experimented. The decrease of accuracy reaches 10% which has seriously reduced the performance of the model. Finally, the 30% pruning EfficientID model has the best effect of feature extraction on data set A.

As shown in Table IV, the 30% pruning of EfficientID was also compared with the classic convolutional neural network VGG-16 [36] and the mobile convolutional neural network MobileNetv2 [37]. Compared to the VGG-16 model, the

TABLE II
COMPARISON OF DIFFERENT OBJECT DETECTORS ON DATA SET A

Model	Precision(%)	Recall(%)	mAP(.5:.95)(%)	FPS($f{\cdot}s^{-1}$)	Params(M)
YOLOv5-s	97.3	99.7	97.1	41	14.4
Faster R-CNN	93.5	92.8	92.1	13	109
DETR	95.3	92.9	93.3	8	159
YOLOv7-tiny	**98.9**	**99.4**	**98.3**	**46**	**12.3**

TABLE III
IMPROVEMENT AND PRUNING EXPERIMENT OF AN EFFICIENTID MODEL ON DATA SET A

Model	Accuracy (%)	Infer-time (ms)	Params (M)
EfficientNetV2	99.8	31	82
EfficientID	99.7	23	67
EfficientID (pruning 30%)	**99.0**	**20**	**32.6**
EfficientID (pruning 50%)	89.6	18	15.4

TABLE IV
PERFORMANCE OF DIFFERENT MODELS AS FEATURE EXTRACTORS ON DATA SET A

Model	Accuracy (%)	Infer-time (ms)	Params (M)
MobileNetv2	97.3	18	14
VGG-16	98.6	135	503
EfficientID (Pruning 30%)	99.0	20	32.6

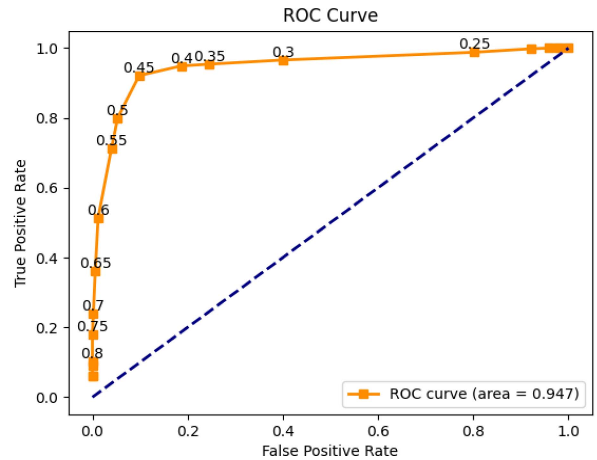


Fig. 10. ROC curve of metric learning.

TABLE V
PERFORMANCE OF DAIRY COW BODY CONDITION SCORING ON DIFFERENT MODELS

Model	Accuracy (%)	Infer-time (ms)	Params (M)
EfficientNetV2	42.2	30	83
VGG-16	38.5	123	503
MobileNetV2	37.6	27	14
EfficientBCS	**46.2**	**38**	**85**

EfficientID is slightly behind in accuracy, but better in inference speed and params. This is because the EfficientID model uses depthwise convolutions at different layers to significantly reduce inference time without loss of accuracy. The difference in inference speed between the two models is not significant but the EfficientID is 2.7% more accurate than MobileNetv2. An EfficientID model used NAS search technology to find the best combination of convolution kernel and network layers to achieve high accuracy. Therefore, the 30% pruning of the EfficientID model has the best performance combined inference speed and accuracy in individual identification module to extract feature vectors of dairy cow and easily deployed in the edge computing device.

In order to evaluate the performance of the model when the identity of the dairy cow to be tested is unknown, experiments were also performed on the TT set. The cosine similarity is calculated for the feature vectors between TH set and template library. The calculated cosine similarity value needs to be compared with a threshold to determine whether the dairy cow already exists in the template library. The threshold is set to a variable between 0 and 1, with an initial value of 0.05 and increments of 0.05. Fig. 10 shows the ROC curve of this experiment and the AUC value is 0.947. When the threshold value is 0.45, the TPR is 92.1%, the FPR is 6.8%. If the maximum value of cosine similarity calculated between the dairy cow to be tested and the feature vectors in the template library is greater than 0.45, this dairy cow is considered to exist in one of the template library, otherwise register its feature vector to the template library as a new dairy cow.

3) Performance of Body Condition Scoring Module: Table V shows the results of body condition scoring of dairy cow based on different models at 0 range error. The accuracy of the EfficientBCS model is improved by 4% compared to the EfficientNetV2, because CA spatial attention can effectively guide the network to learn the important spatial regions of estimating BCS compared to SE channel attention. Since CA spatial attention is slightly more complex than SE channel attention, the inference time and params are increased by 8 ms and 2M, respectively. Although the inference time is sightly longer than the EfficientNetV2 model, it is still tens of millisecond level for a single image during inference, which can easily implement the real-time estimation of BCS. Compared with VGG-16, the accuracy of the EfficientBCS model is significantly improved from 38.5% to 46%, the model inference time is reduced by half and the params is less than 1/5 of the 503M of VGG-16. And the largest accuracy improvement is 8.6% compared with MobileNetV2. Since

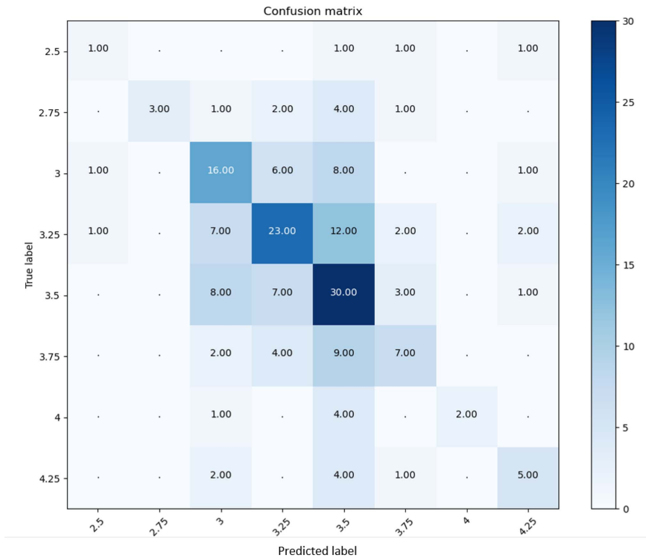


Fig. 11. Confusion matrix for BCS of dairy cows on an EfficientBCS model.

a large number of similarity calculations of feature vectors are required for identifying dairy cows, and the individual identification module and body condition scoring module are performed in parallel, the inference time of estimating BCS will not prolong the inference time of the platform. Based on the above experiments, EfficientBCS is the most effective model for estimating BCS of dairy cow.

Fig. 11 shows the confusion matrix for the BCS of dairy cow based on the EfficientBCS model, and the values on the main diagonal represent accurate predictions. Ideally, the values of the confusion matrix should all be distributed along the main diagonal. Overall, most of the predictions of the confusion matrix are concentrated around the main diagonal, but some results still deviate from the other positions. This is due to the fact that the estimation accuracy of dairy cow's BCS in 0.25 increments is difficult to distinguish even for experts with extensive experience, so the platform uses different levels of deviation indicators to evaluate BCS of dairy cow.

The dairy cow body condition evaluation experiments were performed within 0, 0.25, and 0.5 point deviation, and the evaluation results are shown in Table VI. The average prediction accuracy reached 46% at 0 error, 72% at 0.25 error, and 88% at 0.5 error. Due to the small number of lactating dairy cows with extreme BCSs of 2.5, 2.75, and 4.25 on the farm, the model can learn few samples during training, resulting in lower inference accuracy on these extreme BCSs compared to other BCSs.

4) Comparison With Existing Methods/Systems: For DeLaval system, Mullins et al. [8] mentioned that the correlation between the average categorical automatic and manual BCS was 0.76. The evaluation metrics of DeLaval used a 0 to 5 scale with 0.1 intervals, it is more suitable for regression analysis. However, our and other literature's BCS criterion is based on 0.25 intervals, which tends as a classification problem of discrete values. Therefore, the metrics for regression analysis (R2, RMSE, etc.) are not considered in our work. Table VII shows the results of comparing the platform's individual identification model with existing methods. The EfficientID model has a relatively high accuracy. The comparison of EfficientBCS model with existing literature methods is shown in Table VIII. In different error ranges, our model is slightly inferior to point cloud-based methods, but is close to or better than other methods. However, the point cloud-based methods are more time-consuming. They are difficult to achieve end-to-end execution and have difficulties for deployment in edge environment. In general, the model proposed by the platform meets the needs of farms, but improving the accuracy is one of our key issues to be addressed in the future.

TABLE VI
ACCURACY WITH DIFFERENT RANGE ERRORS FOR BCS ESTIMATION ON DATA SET A

BCS	Accuracy (0 range error)	Accuracy (0.25 range error)	Accuracy (0.5 range error)
2.5	0.25	0.25	0.25
2.75	0.27	0.36	0.55
3	0.5	0.69	0.91
3.25	0.49	0.89	0.94
3.5	0.61	0.82	0.98
3.75	0.32	0.73	0.91
4	0.29	0.29	0.86
4.25	0.42	0.42	0.5
Average	**0.46**	**0.72**	**0.88**

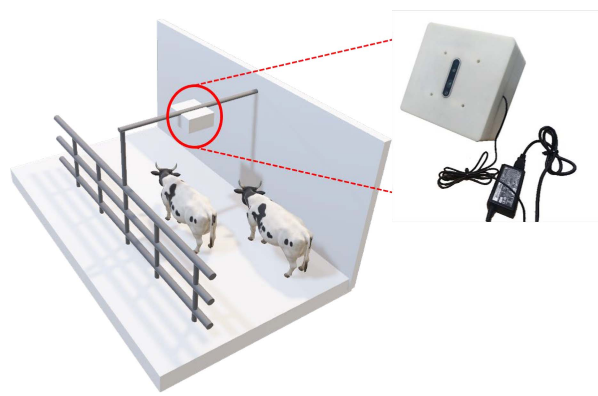


Fig. 12. Platform integrated equipment deployment diagram.

5) Performance of Platform Testing: The platform is to be tested in the milking parlor aisle as shown in Fig. 12. The white equipment is integrated the RGB-D camera and NVIDIA Jetson AGX Xavier, which is powered by an external power supply.

Data set B and TT set of Data set A were tested and the results are shown in Table IX. Since the camera height in the data set A is 3.2 m, and the camera height in the milking parlor environment is 2.8 m, the pixel value of the depth image had to be subtracted by 400 to ensure a uniform height.

On the edge device, the single image inference speed for the cow detection module is 0.037 s, and the detection frame rate can be stabilized at 27 fps when multiple dairy cows

TABLE VII
COMPARISON OF AN INDIVIDUAL IDENTIFICATION MODEL WITH DIFFERENT LITERATURES

	Shen [23]	Xiao [24]	Kawagoe [22]	Ours
acc(%)	96.65	98.67	99.7	99.0
dataset	105 cows 1433 images	48 cows 12000 images	21cows 83170 images	182 cows 5675 images

TABLE VIII
COMPARISON OF BODY CONDITION SCORING WITH DIFFERENT LITERATURES

Error range	Shelley [38]	Shi [17]	Zhao [16]	Ours
0	0.338	0.49	0.45	0.46
0.25	0.714	0.8	0.912	0.72
0.5	0.939	0.96	0.976	0.88

TABLE IX
PERFORMANCE OF PLATFORM TESTING

Task	Accuracy (%)	Infer-time (s)
Object Detection	98.6	0.037
Individual Identification	90.4	2.31
BCS Evaluation (exact prediction)	42	0.061
BCS Evaluation (0.25 range error)	71	0.061
BCS Evaluation (0.5 range error)	85	0.061

are present in the image. To save time for calculating the similarity of feature vectors, two feature vectors were extracted from each dairy cow and stored in the template library. The inference time for identifying one dairy cow is 2.31 s on average, with an accuracy of 90.4%.

The mean inference time for dairy cow body condition scoring is 0.061 s, with accuracies of 42%, 74%, and 87% at 0, 0.25, and 0.5 errors, respectively. The small decrease in accuracy compared to training is due to the uneven distribution of sunlight in the parlor passage, which is susceptible to direct sunlight. When strong sunlight is directed onto the dairy cow's dorsal hide, there is specular reflection when the infrared beam is directed onto the dairy cow's dorsal hide surface, resulting in missing depth values in this area. As shown in Fig. 13, there are depth image of dairy cow's back with slight, moderate, severe missing. Due to the different degrees of missing in the depth images, the accuracy of the platform test is slightly lower than the test accuracy of the experiments.

Fig. 14 shows the process of inferred ID and BCS of dairy cow by edge device, visualized by the OpenCV library. The cow4 and cow5 are the temporary IDs assigned by the ARI algorithm. Fig. 14(a) shows the platform's inference about the individual identification and BCS of cow5 when it is fully present, and Fig. 14(b) shows the inferred ID and BCS of cow5. The object detection module can detect multiple dairy cows and ensure that a dairy cow only performs the individual identification and estimation of BCS once. When the inference is completed, the dairy cow's ID and BCS will be transmitted to the cloud analysis center.

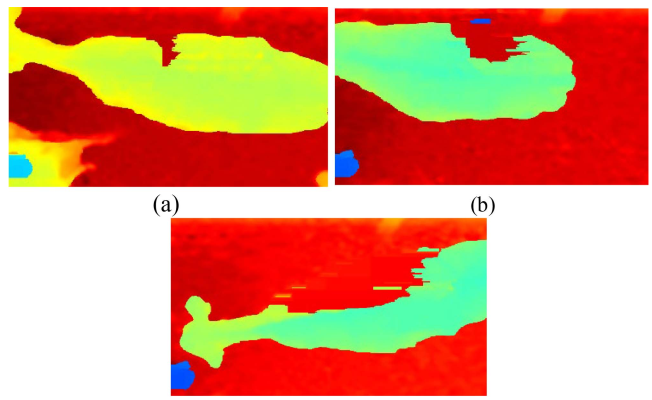

(a) (b) (c)

Fig. 13. Different degrees of depth deficit in the depth images of the dairy cow's back. (a) Slight missing. (b) Moderate missing. (c) Severe missing.

(a) (b)

Fig. 14. Real-time inference with multiple dairy cows. (a) Cow5 will be inferred. (b) inference result of cow5.

The cloud analysis center includes the cloud database and Web pages. Then, the dairy cow information inferred by edge computing device are displayed in Bootstrap, as shown in Fig. 15. Each page displays ten pieces of data and supports searching by dairy cow's ID, which is convenient to find the BCS information and upload time of a specific dairy cow. The information of dairy cows is visualized by Echarts, as shown in Fig. 16. The bar chart visualizes the number of dairy cows with different BCS, while the pie chart reflects the proportion of dairy cows with different BCS among all dairy cows and the number of dairy cows detected in the last three years.

Table X shows the latency of cloud computing and platform (edge computing) when there is a single node. Among them, dairy cow's individual identification and body condition scoring are executed in parallel, and the latency only counts the

Enter the cow identity Search Edit Delete

cowID	COW Identification	COW BCS	Upload Time
1	200100	3.5	2023-03-30-15-38
2	211115	3.25	2023-03-30-10-30
3	200983	2.75	2023-03-30-11-00
4	211175	2.5	2023-03-29-10-54
5	200555	3	2023-03-28-10-54
6	210578	3.75	2023-03-27-10-54
7	211103	4	2023-03-26-10-54
8	211003	4.25	2023-03-25-10-54
9	255469	3.5	2022-12-24-14-32
10	200456	3.5	2022-12-25-14-32

‹ 1 2 3 4 5 › ››

Fig. 15. Display of BCS by Bootstrap.

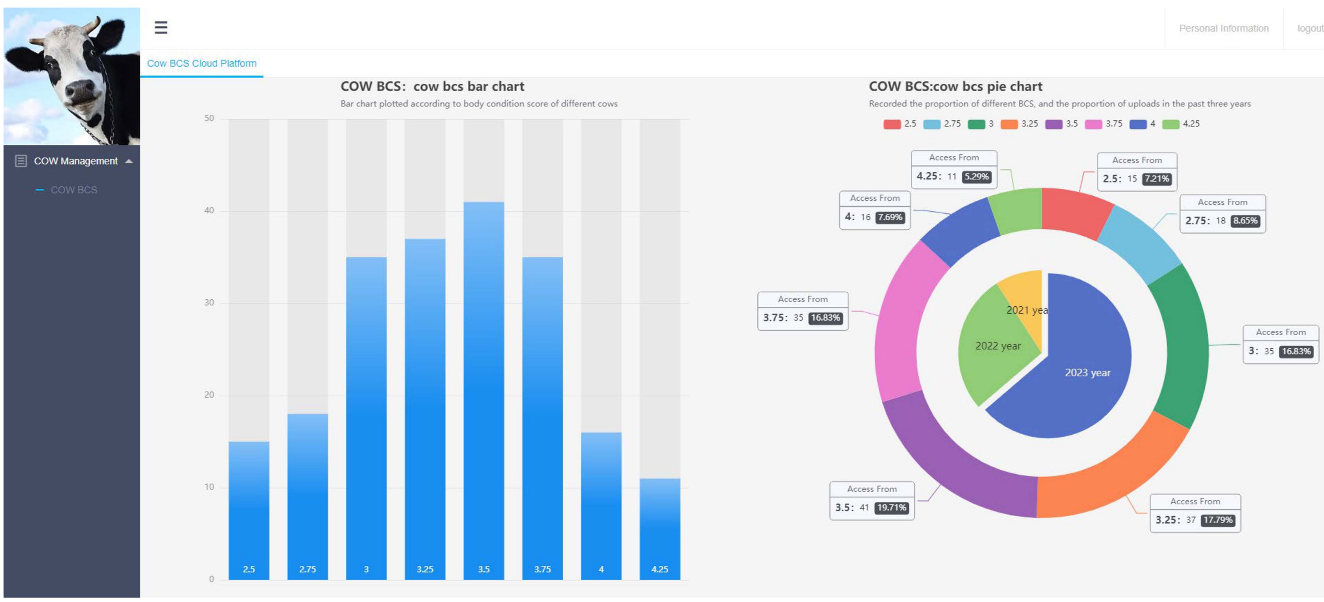


Fig. 16. Visual analysis of BCS by Echarts.

TABLE X
TIME COMPARISON BETWEEN CLOUD COMPUTING AND PLATFORM

	video transmission (one cow)	image preprocessing	models' inference			result uploading	sum
			YOLOv7+ARI	ID	BCS		
cloud computing	6.48	0.72	0.027	1.97	0.038 (ignore)	0.028	9.225
proposed platform	-	0.759	0.037	2.31	0.061 (ignore)	0.032	3.138

larger one. The time components of the platform are mainly image preprocessing, models' inference, and result uploading, and the corresponding times are 0.759, 2.347, and 0.032 s. The computing latency of cloud computing is greater than the platform in Table X. In cloud computing, the RGB-D camera will always capture RGB and depth video streams of cows walking and upload them to the cloud via the network. One cow will appear for at least 3 s with 30 fps, and the transmission delays of one RGB image and one depth image are 0.037 and 0.034 s, respectively. The transmission delay caused by uploading video streams is huge. Therefore, the video streaming scenario is not suitable for cloud computing.

In addition, if the number of edge nodes increases, the transmission bandwidth and computing power requirements of the central server will also increase. In general, it is more advantageous to perform computation at the edge in complex environments.

6) QoS Analysis: We discussed several important metrics of QoS with commercial systems in terms of reliability, network lifetime, cost-effectivity, decisioning delay, and scalability.

1) *Reliability:* The end device uses the same RGB-D camera (Intel RealSense D435) as HerdVision system. A noncontact and more reliable camera system better survives the harsh farm environment compared to wearable devices. In addition, the platform used NVIDIA Jetson AGX Xavier, an industrial-grade edge device with powerful arithmetic, to ensure reliability. The mold is 3-D printed with resin and has good heat and moisture resistance.
2) *Network Lifetime:* Similar to the HerdVision system, the proposed platform is able to connect to the Internet via WIFI, Ethernet, and 4G/5G modules. All devices are connected directly to the power supply to avoid shortening network lifetime due to battery drain. The platform offloads computation to edge servers and only transmits inference results rather than images through the network, thus saving energy and extending network lifetime.
3) *Cost-Effectivity:* The DeLaval system requires installation on DeLaval sort gate or on DeLaval VMS™, which is an additional expense. The proposed platform is entirely based on camera systems and does not require the installation of additional equipment. It can save the cost of hiring a professional scorer and purchasing ear tags.
4) *Decisioning Delay:* The platform deploys lightweight CNN and uses pruning to effectively optimize memory space and inference speed. Table X shows the decision delays of this platform.
5) *Scalability:* More dairy cow health detection tasks can be migrated to the proposed camera system-based edge-IoT platform without additional equipment, such as cow lameness detection, weight estimation, etc.

V. Conclusion

In this work, an intelligent Edge-IoT platform based on deep learning for estimating BCS was realized. The end device is an RGB-D camera which captures the back video stream of dairy cows and transmits the streams to edge computing device. The edge computing device performs three modules of object detection, individual identification and body condition scoring of dairy cows by integrated deep learning models. The inferred ID and BCS of dairy cows are upload to the cloud analysis center for visually analyzing. On the data sets collected from the farm, the accuracy of object detection and individual identification reached 98.6% and 90.4%, respectively, and dairy cow's ID is enabled to be registered in the template library if it is not registered before. The estimation accuracy of the body condition estimation model reached 74% within 0.25 error and 87% within 0.5 error. It takes 3.138s for the complete inference of one dairy cow, which realized real-time inference. In future work, first, depth completion is required for partial missing of dairy cow depth images caused by lighting, etc.. Second, cows that need to be supplemented with extreme BCS in the data sets to solve the issue of low accuracy in extreme BCS cases. Third, consider integrating the better BCS estimation methods (such as point cloud) into the platform and improving their inference speed. Finally, the platform is considered to integrate more intelligent perception tasks for dairy cows.

References

[1] S. Kumar et al., "Deep learning framework for recognition of cattle using muzzle point image pattern," *Measurement*, vol. 116, pp. 1–17, Feb. 2018.

[2] A. Cominotte et al., "Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases," *Livestock Sci.*, vol. 232, Feb. 2020, Art. no. 103904.

[3] W. Andrew, C. Greatwood, and T. Burghardt, "Fusing animal biometrics with autonomous robotics: Drone-based search and individual id of friesian cattle," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops*, 2020, pp. 38–43.

[4] E. E. Wildman, G. M. Jones, P. E. Wagner, R. L. Boman, H. F. Troutt Jr, and T. N. Lesch, "A dairy cow body condition scoring system and its relationship to selected production characteristics," *J. Dairy Sci.*, vol. 65, no. 3, pp. 495–501, 1982.

[5] J. D. Ferguson, D. T. Galligan, and N. Thomsen, "Principal descriptors of body condition score in Holstein cows," *J. Dairy Sci.*, vol. 77, no. 9, pp. 2695–2703, 1994.

[6] I. Halachmi, M. Klopčič, P. Polak, D. J. Roberts, and J. M. Bewley, "Automatic assessment of dairy cattle body condition score using thermal imaging," *Comput. Electron. Agric.*, vol. 99, pp. 35–40, 2013.

[7] L. Ruiz-Garcia and L. Lunadei, "The role of RFID in agriculture: Applications, limitations and challenges," *Comput. Electron. Agric.*, vol. 79, no. 1, pp. 42–50, 2011.

[8] I. L. Mullins, C. M. Truman, M. R. Campler, J. M. Bewley, and J. H. C. Costa, "Validation of a commercial automated body condition scoring system on a commercial dairy farm," *Animals*, vol. 9, no. 6, p. 287, 2019.

[9] C. H. Sandgren and U. Emanuelson, "Consistency of measurements from an automatic body condition scoring camera," in *Precision Dairy Farming*, Leiden, The Netherlands: Brill Publ., 2016, pp. 285–290,

[10] (HerdVision, Lancaster, U.K.). "*Automated Cow Health And Productivity Monitoring*. Accessed: Sep. 10, 2023. [Online]. Available: https://herd.vision/

[11] G. Azzaro et al., "Objective estimation of body condition score by modeling cow body shape from digital images," *J. Dairy Sci.*, vol. 94, no. 4, pp. 2126–2137, 2011.

[12] R. Spoliansky, Y. Edan, Y. Parmet, and I. Halachmi, "Development of automatic body condition scoring using a low-cost 3-dimensional Kinect camera," *J. Dairy Sci.*, vol. 99, no. 9, pp. 7714–7725, 2016.

[13] Y. Liu and J. Qin, "Research and application of dairy cows body condition score based on attention mechanism," in *Proc. IEEE 6th Int. Conf. Cloud Comput. Big Data Anal. (ICCCBDA)*, 2021, pp. 600–606.

[14] Y. Sun et al., "Automatic monitoring system for individual dairy cows based on a deep learning framework that provides identification via body parts and estimation of body condition score," *J. Dairy Sci.*, vol. 102, no. 11, pp. 10140–10151, 2019.

[15] J. Alvarez et al., "Body condition estimation on cows from depth images using convolutional neural networks," *Comput. Electron. Agric.*, vol. 155, pp. 12–22, Dec. 2018.

[16] K. Zhao et al., "Automatic body condition scoring for dairy cows based on efficient net and convex hull features of point clouds," *Comput. Electron. Agric.*, vol. 205, Feb. 2023, Art. no. 107588.

[17] W. Shi, B. Dai, W. Shen, Y. Sun, K. Zhao, and Y. Zhang, "Automatic estimation of dairy cow body condition score based on attention-guided 3D point cloud feature extraction," *Comput. Electron. Agric.*, vol. 206, Mar. 2023, Art. no. 107666.

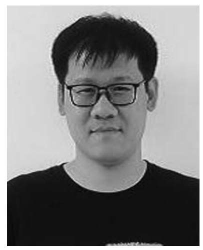
[18] J. Gardenier, J. Underwood, and C. Clark, "Object detection for cattle gait tracking," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2018, pp. 2206–2213.
[19] A. Yılmaz, G. N. Uzun, M. Z. Gürbüz, and O. Kıvrak, "Detection and breed classification of cattle using YOLO v4 algorithm," in *Proc. Int. Conf. Innov. Intell. Syst. Appl. (INISTA)*, 2021, pp. 1–4.
[20] X. Huang, Z. Hu, X. Wang, X. Yang, J. Zhang, and D. Shi, "An improved single shot multibox detector method applied in body condition score for dairy cows," *Animals*, vol. 9, no. 7, p. 470, 2019.
[21] W. Kusakunniran and T. Chaiviroonjaroen, "Automatic cattle identification based on multi-channel LBP on muzzle images," in *Proc. Int. Conf. Sustain. Inf. Eng. Technol. (SIET)*, 2018, pp. 1–5.
[22] Y. Kawagoe, I. Kobayashi, and T. T. Zin, "Facial region analysis for individual identification of cows and feeding time estimation," *Agriculture*, vol. 13, no. 5, p. 1016, 2023.
[23] W. Shen et al., "Individual identification of dairy cows based on convolutional neural networks," *Multimedia Tools Appl.*, vol. 79, pp. 14711–14724, Jun. 2020.
[24] J. Xiao, G. Liu, K. Wang, and Y. Si, "Cow identification in free-stall barns based on an improved Mask R-CNN and an SVM," *Comput. Electron. Agric.*, vol. 194, Mar. 2022, Art. no. 106738.
[25] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 7464–7475.
[26] B. Jiang, Q. Wu, X. Yin, D. Wu, H. Song, and D. He, "FLYOLOv3 deep learning for key parts of dairy cow body detection," *Comput. Electron. Agric.*, vol. 166, Nov. 2019, Art. no. 104982.
[27] K. Zhao, X. Jin, J. Ji, J. Wang, H. Ma, and X. Zhu, "Individual identification of Holstein dairy cows based on detecting and matching feature points in body images," *Biosyst. Eng.*, vol. 181, pp. 128–139, May 2019.
[28] Y. C. Tsai, J. T. Hsu, S. T. Ding, D. J. A. Rustia, and T. T. Lin, "Assessment of dairy cow heat stress by monitoring drinking behaviour using an embedded imaging system," *Biosyst. Eng.*, vol. 199, pp. 97–108, Nov. 2020.
[29] M. Tan, and Q. Le, "Efficientnetv2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10096–10106.
[30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[31] H. Li, A. Kadav, I. Durdanovic, H. Samet, and H. P. Graf, "Pruning filters for efficient convnets," 2016, *arXiv:1608.08710*.
[32] A. J. Heinrichs, and G. L. Hargrove, "Standards of weight and height for Holstein heifers," *J. Dairy Sci.*, vol. 70, no. 3, pp. 653–660, 1987.
[33] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13713–13722.
[34] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.
[35] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," 2020, *arXiv:2010.04159*.
[36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[37] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.
[38] A. N. Shelley, "Incorporating machine vision in precision dairy farming technologies," Ph.D. dissertation, Electr. Comput. Eng., Univ. Kentucky, Lexington, KY, USA, 2016.

Junhao Wang received the B.E. degree in software engineering from the Shandong University of Technology, Zibo, China, in 2021. He is currently pursuing the M.D. degree in computer science and technology with Northeast Agricultural University, Harbin, China.

His current research interest are deep learning and Internet of Things.

Baisheng Dai received the B.Sc. and M.Sc. degrees from Northeast Agricultural University (NEAU), Harbin, China, in 2007 and 2010, respectively, and the Ph.D. degree in computer applied technology from Harbin Institute of Technology, Harbin, China, in 2017.

He worked as a Lecturer with the College of Electrical Engineering and Information, NEAU from 2017 to 2020, where he has been an Associate Professor since 2020. His current research interest include smart livestock farming, digital agriculture, and intelligent visual perception.



Yang Li received the B.A. degree in animal science from Heilongjiang Bayi Agriculture University, Daqing, China, in 2011, and the M.A. degree in animal nutrition and feed science and the Ph.D. degree in animal production from Northeast Agricultural University (NEAU), Harbin, China, in 2014 and 2017, respectively.

He worked as a Lecturer with the College of Animal Science and Technology, NEAU from 2018 to 2019, where he has been an Associate Professor since 2019. His current research interest is ruminant nutrition.

Yongqiang He received the B.D. degree in food quality and safety from Jilin Agricultural University, Changchun, China, in 2014, and the MBA degree in business administration from Wuhan University of Technology, Wuhan, China, in 2015.

He is the General Manager with the Milk Source Quality Technology Center, Mengniu Company Ltd., Hohhot, China.

Yukun Sun received the B.A. degree in animal science and the M.A. degree in animal nutrition and feed science from Jilin Agricultural University, Changchun, China, in 2014 and 2017, respectively, and the Ph.D. degree in animal production from Northeast Agricultural University, Harbin, China, in 2020.

He is working as a Lecturer with the College of Animal Science and Technology, Northeast Agricultural University. His research direction is ruminant nutrition and intelligent pasture management.

Weizheng Shen received the B.D. degree in computer science and technology from Yanshan University, Qinhuangdao, China, in 2000, and the M.D. degree and the Ph.D. degree in agricultural electrification and automation from Northeast Agricultural University, Harbin, China, in 2003 and 2008, respectively.

He is currently a Full Professor with the College of Electrical Engineering and Information, Northeast Agricultural University. His current research interest include smart livestock farming and digital agriculture.