BIOSYSTEMS ENGINEERING 192 (2020) 245–256

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Cow identification based on fusion of deep parts features

Hengqi Hu [a], *Baisheng Dai* [a,b,**], *Weizheng Shen* [a,*], *Xiaoli Wei* [a], *Jian Sun* [a], *Runze Li* [a], *Yonggen Zhang* [c]

[a] School of Electrical Engineering and Information, Northeast Agricultural University, Harbin, 150030, China
[b] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
[c] School of Animal Science and Technology, Northeast Agricultural University, Harbin, 150030, China

ARTICLE INFO

Article history:
Received 29 March 2019
Received in revised form
21 January 2020
Accepted 5 February 2020
Published online 21 February 2020

Keywords:
Cow identification
Computer vision
Convolutional neural networks
Deep parts feature fusion
Precision livestock farming

Livestock identification is of great significance for achieving precision livestock farming as it is a prerequisite of modern livestock management and automatic behaviour analysis. With respect to cow identification, methods based on computer vision have been widely considered due to their advantage of non-contact and practicality. In this paper, a novel non-contact cow identification method based on fusion of deep parts features is proposed. First, a set of side-view images of cows were captured, and then the YOLO object detection model was applied to locate the cow object in each original image, which was then divided into three parts, head, trunk and legs, by a part segmentation algorithm using frame differencing and segmentation span analysis. Then, three independent convolutional neural networks (CNNs) were trained to extract deep features from these three parts, and a feature fusion strategy was designed to fuse the features, i.e., deep parts feature fusion. Finally, a support vector machine (SVM) classifier trained by the fused features was used to identify each individual cow. The proposed method achieved 98.36% cow identification accuracy on a dataset containing side-view images of 93 cows, which outperformed existing works. Experimental results showed the effectiveness of the proposed cow identification method and the good potential for this method in individual identification of other livestock.

© 2020 IAgrE. Published by Elsevier Ltd. All rights reserved.

1. Introduction

In modern dairy management, accurate and reliable identification of each individual cow is of great significance for precision livestock farming, which can be used in intelligent milking, health monitoring, weighing, automatic behaviour analysis, etc (Kumar et al., 2018). Therefore, individual identification of dairy cows is the basis for other applications. In this paper, we proposed an identification method of the side of the cow, which can be applied to some fields of individualised behaviour detection and intelligent analysis of dairy cows, such as lameness detection, drinking behaviour analysis,

* *Corresponding author.*
** *Corresponding author.* School of Electrical Engineering and Information, Northeast Agricultural University, Harbin, 150030, China.
E-mail addresses: bsdai@neau.edu.cn (B. Dai), cswzshen@sina.com (W. Shen).
https://doi.org/10.1016/j.biosystemseng.2020.02.001
1537-5110/© 2020 IAgrE. Published by Elsevier Ltd. All rights reserved.

Nomenclature

A_{det_i}	Area of det_i
A_t	Threshold of area
COW	Image set to be identified
det_i	The i^{th} object detection result
DPFF	Deep Parts Feature Fusion
FPS	Frames per second
$intvl_i$	The i^{th} intervals
$label_{det_i}$	Object name of det_i
L_{intvl_i}	Lower boundary height of $intvl_i$
LT	Lower boundary height of TS
M_r	Maximum value of all span ranges
M_u	Maximum value of all upper boundary heights
R_{intvl_i}	Span range of $intvl_i$
TS	Set of intervals containing the trunk part
U_{intvl_i}	Upper boundary height of $intvl_i$
w_r	Weight for span range of the intervals containing the trunk part
w_u	Weight for the upper boundary of trunk

linear appraisal, mastitis detection and individual localisation. Furthermore, it can provide a reference for cow identification based on other views and for application to other fields of dairy cow husbandry.

Currently, sensor-based systems are widely used in farms for livestock identification or localisation and tracking, e.g., passive radio-frequency identification (RFID) ear tags (DeVries, Von Keyserlingk, Weary, & Beauchemin, 2003; Voulodimos, Patrikakis, Sideridis, Ntafis, & Xylouri, 2010), active RFID ear tags (Barbari, Conti, & Simonini, 2009; Ipema, Van De Ven, & Hogewerf, 2013; S M C; Porto, Arcidiacono, Giummarra, Anguzza, & Cascone, 2014) and other wireless technologies such as radar and wireless local area networks (WLANs). However, ear tags are intrusive and tend to be lost and damaged over time (Kumar et al., 2018). In addition, the reading/writing distance of the passive ear tag is limited (Voulodimos et al.). The active tag has a greater reading/writing distance, but its capability depends greatly on battery power. The local position measurement (LPM) system based on radar technology (Gygax, Neisen, & Bollhalder, 2007) is too expensive to be applied extensively in production. The WLAN tracking system (Huhtala, Suhonen, Mäkelä, Hakojärvi, & Ahokas, 2007) has a reasonable price, but these localisation and tracking systems require cows to be able to wear a transponder. Generally, sensor-based systems have high accuracy in identification and a wide boundary of applicability, but the systems are contact-type, and ear tags and wearable devices are easily lost or damaged over time. In contrast, computer vision-based systems are low cost and non-contact type, which can reduce management costs of farm and stress reactions of livestock. Therefore, it is essential to explore the individual identification of dairy cows based on computer vision-based systems.

Recently, computer vision has been widely used in wildlife science and animal biometrics due to its automation degree, non-contact, and practicality (Kühl & Burghardt, 2013; Weinstein, 2018). In addition to individual identification, researchers have made significant contributions to other typical topics such as object detection (Chen et al., 2014; Jiang et al., 2019), measurement of functional traits (Guo et al., 2017; Nir, Parmet, Werner, Adin, & Halachmi, 2018; Pezzuolo, Guarino, Sartori, & Marinello, 2018; Salau, Haas, Junge, & Thaller, 2017), 3D reconstruction of movement and morphology (Guo et al., 2017; Haggag, Abobakr, Hossny, & Nahavandi, 2016; Lavy et al., 2015), body condition estimation (Alvarez et al., 2018; Spoliansky, Edan, Parmet, & Halachmi, 2016; Yukun et al., 2019), and automatic detection of dairy cow behaviours, such as lying recognition (Porto, Arcidiacono, Anguzza, & Cascone, 2013), feeding and standing behaviour detection (Porto, Arcidiacono, Anguzza, & Cascone, 2015), lameness detection (Jabbar, Hansen, Smith, & Smith, 2017), and oestrus and mating detection (Tsai & Huang, 2014). However, in most of these contributions, the individual identification of cows was neglected. Specifically, although reported automatic methods for detecting dairy cow behaviours have good performance and applicability, more accurate and individualised behaviour detection and analysis could be achieved if individual cows could be identified. Therefore, identification is the basis of artificial intelligence technology applied in precision livestock farming. Our proposed method not only can improve identification performance in practice, but can also be directly combined with related works such as lameness detection, drinking behaviour analysis, linear appraisal, mastitis detection and individual localisation. Furthermore, it can provide a reference for cow identification based on other views, and be applied to the abovementioned topics, and other fields of dairy cow husbandry, to implement precision livestock farming.

Reported studies on cow identification use mainly four body parts: muzzle, face, back and trunk, as shown in Fig. 1(a)–(d). Gaber, Tharwat, Hassanien, and Snasel (2016) extracted features of the muzzle print images using the Weber local descriptor (WLD) algorithm, and then used the AdaBoost classifier to achieve identification. Kumar et al. (2018) used the muzzle part to identify individual cattle. They used CNNs and deep belief networks (DBNs) to extract features of muzzle point images, and their method achieved good performance on identification. However, it is hard to capture muzzle images due to head movement and the small size of the muzzle. Therefore, it is not realistic to utilise the muzzle part for identification in actual production environments. Kumar, Tiwari, and Singh (2016) also used facial images to identify individual cattle. In this method, the AdaBoost face detection algorithm (Viola & Jones, 2004) was used to detect cattle faces, and SVM and incremental support vector machine (ISVM) were trained for identification. However, uneven lighting conditions, head movements, and unconstrained environment introduced complexities that restricted the identification performance of the method. Li, Ji, Wang, Sun, and Yang (2017) proposed using SVM and other methods to classify the tail-head image of the cow. In this method, Zernike moments were used to describe the shape characteristics of the white pattern in the tail-head image. However, it is difficult to identify cows with a mainly black colour using this method. (Zin et al., 2018) extracted the cow back images by applying frame differencing and trained a CNN to classify them. Zhao and He (2015) trained a CNN to classify trunk images, using frame differencing to obtain the

Fig. 1 – Some parts of interest for cow identification: (a) muzzle, (b) face, (c) back, (d) trunk, and (e) overall cow object.

rough contour of the cow, and then analysed the segmentation span of binary images to extract the trunk part.

However, except for the trunk part, the overall side-view image of a cow (Fig. 1(e)) includes the head and legs, which also contain discriminative texture and contour features that are ignored by the work of Zhao and He (2015), who extracted deep CNN features only from the trunk part of the cow. Furthermore, considering that the side of a cow is more stable and easier to detect than the muzzle and face, and that it is hardly used for individual identification, so in our previous work (Shen et al., 2019), the side-view images of 105 cows were used to train a CNN. However, when performing identification using the cow object images, each part of the cow object cannot be weighted, and each part also has a certain capacity to identify different cows. In this paper, to better exploit the deep CNN features of each part of the cow object and to better explore the impact of each part and the multipart combinations on cow identification, we first fine-tuned an object detection model to locate the cow object in images containing the side-view of a cow. Next, we improved a trunk extraction algorithm to segment the detected cow object into three parts, namely the head, trunk, and legs, and trained three independent CNNs to extract deep features from these three parts. Then, we designed strategies to fuse these features and analysed the contribution of each part and the multipart combinations to cow identification and their performance. Finally, we designed a model called deep parts feature fusion (DPFF) to achieve cow identification. In addition, we not only analysed the identification performance of three feature fusion strategies but also verified that the identification performance of the proposed model of fusing the deep CNN features from each part of the side of the cow is higher than that of directly using the overall cow object, a single part, or the combinations of any two parts.

In general, our proposed method identifies individual cows by fusing the deep CNN features. It mainly contains four modules: 1) object detection; 2) part segmentation; 3) feature extraction and fusion; 4) classification. The core contributions of this work are summarised as follows.

- A novel deep parts feature fusion framework is proposed in this work, with which more discriminative features can be extracted for cow identification.
- The contribution of each part of a cow and their combinations to individual identification are analysed, based on which an optimal combination can be determined.
- The proposed identification method is evaluated on a private dataset containing side-view images of 93 cows and achieves an outperformance over existing works.

The remainder of the paper is organised as follows. We first introduce our experimental materials and derive our methods in section 2. The experimental results, related analysis, and discussions are presented in section 3. Finally, the paper is concluded in section 4.

2. Materials and methods

2.1. Barn under study

Image acquisition was performed at Kangda Cow Farm, located in Shuangcheng, Harbin, China. The dairy cows on this farm (n = 226) belonged to the Holstein-Friesian breed, and they were housed in free-stall barns with sand bedding and equipped with self-locking neck clips at the feed line. The barn of approximately 130 m × 30 m was equipped with a fan and spray to reduce the temperature around the cows. The basal diet was mainly composed of alfalfa hay, maize silage and loose-meal concentrate based on bean pulp, DDGS, maize processing by-product, and cottonseed.

2.2. Image acquisition

The images in our dataset were captured in the milking parlour, and all lactating cows were milked twice a day (5:00–8:30 and 16:30–20:00). Figure 2 shows the layout of the milking parlour. The milking parlour had a length of 20.2 m and width of 7.5 m, and was composed of a 2 × 12 side-by-side parallel milking area and a return alley. The milking area was 17.8 m in length and 1.6 m in width, and the return alley was 7.5 m in length and 2.4 m in width. As shown in Fig. 2, after milking, the cows took turns through the return alley, and an ASUS Xtion2 camera that fixed in 'Δ' was used to capture the side-view image sequences of cows. We captured the image sequences of both sides by rotating the camera, the horizontally placed camera was used to capture cows out of the right milking area (top channel in Fig. 2), and the tilted camera was used to capture the cows out of the left milking area (bottom channel in Fig. 2). A total of 105 cows were captured, and all cows had been used in our previous work (Shen et al., 2019).

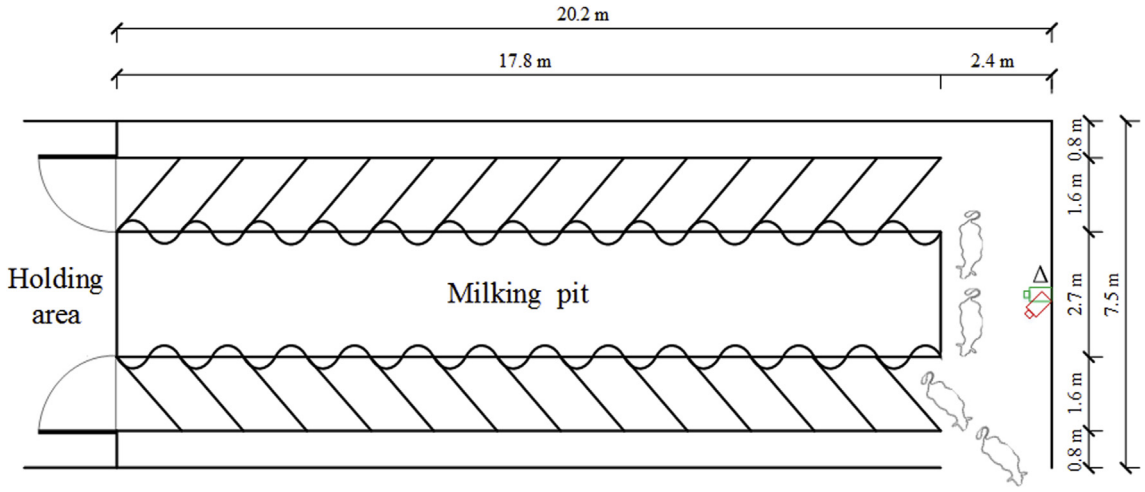

Fig. 2 – Layout of the milking parlour.

Among the 105 cows, 12 cow head parts were obscured by the previous cow during image acquisition. Therefore, the remaining 93 cows were used in this experiment.

Image acquisition was carried out in a relatively natural environment, in which some other objects, such as milk transmission pipes and iron railings, may have caused difficulties in the detection and identification of cows. Therefore, the experimental images can be used to evaluate the performance of the cow identification method objectively, as they reflect many common characteristics of cows' actual breeding environment.

2.3. *Experimental data*

The duration in which the whole cow was in the field of view was approximately 2 s. The image sequences were recorded with a frame rate of approximately 30 FPS at a resolution of 640 (horizontal) × 480 (vertical) pixels. Overlapping frames for some cows in these images were manually removed by post-processing. The remaining valid frames (4353 images) were randomly divided into two parts: one part was used to fine-tune the object detection model, and the other part was input into the model to locate the cow objects, which were manually assigned into corresponding cow categories to form an image set for the following identification. In this paper, we fine-tune YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016) for cow detection. YOLO is a deep learning framework for object detection, pre-trained on MSCOCO (Lin et al., 2014), which is a dataset constructed by Microsoft for research on image classification, object detection, segmentation, etc. This framework can detect the position of all trained objects in the image and it meets the real-time needs of cow identification. YOLO has already be pre-trained with cow object, and in this work, 1000 images of cows were selected to fine-tune it to improve the detection performance. Finally, a total of 958 side-view images of cow objects were obtained, with an average of 10 images per cow. We randomly assigned 70% of each cow's images into the training set (593 images) and the remaining 30% into the validation set (365 images) for cow identification.

2.4. *Individual identification*

Figure 3 shows the flowchart of the proposed cow identification method. First, the input images were sent to YOLO to locate each cow object of interest, which was then divided into three parts: head, trunk, and legs. Then, the segmented parts were sent to three CNNs for feature extraction. The final identity feature was formed after partial features were fused. Finally, an SVM was trained using the fused features to achieve individual identification of the dairy cows. The details are presented as follows.

2.4.1. *Object detection*

In this work, YOLO was applied to locate the cows, and it can also suppress interference from other cows and background noise. Figure 4 shows two examples of object detection results. The images of interest for each cow were assigned into the image set to be identified by Eq. (1).

$$COW = \left\{ det_i \middle| A_{det_i} > A_t, label_{det_i} = cow; i = 1, 2, \ldots n \right\} \quad (1)$$

where COW represents the image set to be identified, A_{det_i} represents the area of the object detection result det_i, and A_t is an area threshold. To select the salient cow of interest in each image, the value of A_t is empirically set to $0.25 \times (640 \times 480)$. $label_{det_i}$ represents the object name of det_i, and n represents the number of all object detection results. Finally, we selected each object detection result whose area was greater than A_t, and the object name was cow. Since the model can also detect a cow object whose head was obscured or beyond the field of view, to verify the multipart feature fusion experiment, the object detection results that do not contain the head part were manually removed by post-processing, and the remaining results were formed into the image set to be identified.

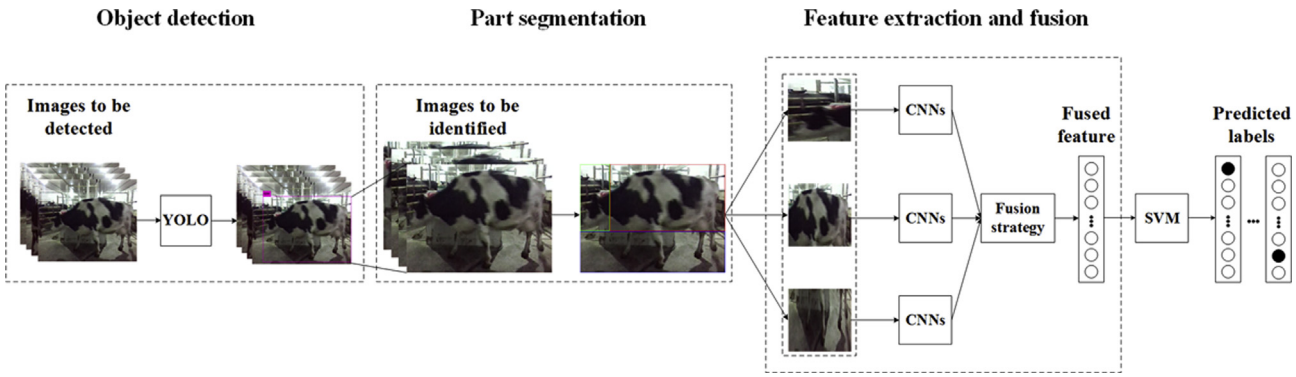

Fig. 3 – Flowchart of the proposed cow identification method.

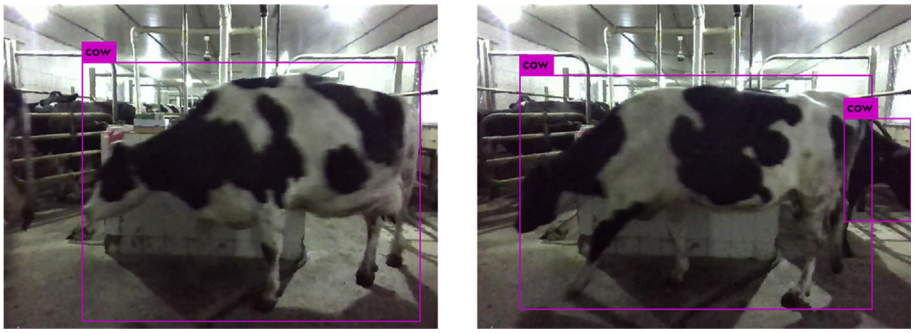

Fig. 4 – Two examples of object detection results.

2.4.2. *Part segmentation*

To extract features for each part separately, the cow object was segmented into three parts: head, trunk, and legs. We first proposed a method to extract the trunk part on the basis of the work of Zhao and He (2015). Then, the head and leg parts were obtained by the geometric relationships.

Specifically, to extract the trunk part, two-frame differencing (performing a difference operation on the grayscale images of two adjacent frames and converting it to a binary image) was first used to obtain a rough contour of the moving cow object. Next, the corresponding border of object detection was used to locate the cow of interest in the interframe difference image, which was then divided into equally spaced vertical intervals. The upper and lower boundaries of the true value of the binary image in each interval were obtained, and the distance between the upper and lower boundary was taken as the span range. A schematic diagram of the interframe difference image is shown in Fig. 5. Here, we define $intvl_i$, U_{intvl_i}, L_{intvl_i}, and R_{intvl_i}, as shown in this figure, where $intvl_i$ represents the i^{th} interval, U_{intvl_i} represents the upper boundary height of $intvl_i$, L_{intvl_i} represents the lower boundary height of $intvl_i$, and R_{intvl_i} represents the span range of $intvl_i$. R_{intvl_i} was calculated by Eq. (2).

$$R_{intvl_i} = U_{intvl_i} - L_{intvl_i} \quad \text{for } i = 1, \ldots, n \tag{2}$$

where n represents the number of intervals and is empirically set to 20.

Considering that the upper boundary of the intervals containing the trunk part is relatively high, the span range of these intervals is relatively large. Therefore, the intervals containing trunk parts were selected by Eq. (3).

$$TS = \left\{ intvl_i \middle| U_{intvl_i} > M_u - w_u M_r, R_{intvl_i} > w_r M_r; i = 1, \ldots, n \right\} \tag{3}$$

where

$$M_u = max\left\{ U_{intvl_i}, \ldots, U_{intvl_n} \right\} \tag{4}$$

$$M_r = max\left\{ R_{intvl_i}, \ldots, R_{intvl_n} \right\} \tag{5}$$

and where TS represents the set of intervals containing the trunk part. To select the appropriate upper boundary and span range, according to the abovementioned characteristics of TS, two weights—w_u and w_r—were set in Eq. (3), w_u is a defined weight for the upper boundary of the trunk, and w_r is a defined weight for the span range of the intervals containing the trunk part. The values of w_u and w_r were set to 0.06 and 0.4, respectively, according to the work of Zhao and He (2015), with which the salient intervals containing trunk parts could be selected. When U_{intvl_i} and R_{intvl_i} of $intvl_i$ satisfied Eq. (3), it indicated that the i^{th} interval contained the trunk part. In the experiment, we found that the last two intervals (corresponding to the cow hip) of some cows did not satisfy Eq. (3) as U_{intvl_i} was slightly lower than the upper boundary height of

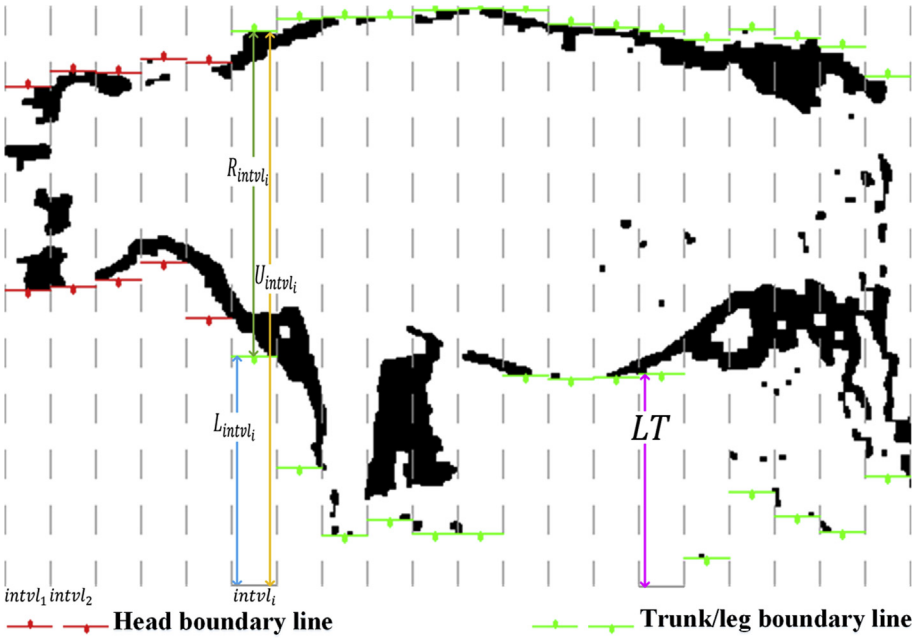

Fig. 5 – Schematic diagram of the interframe difference image.

the trunk part or R_{intvl_i} was relatively small. However, because the interframe difference image was located by the corresponding border of object detection, these two intervals were added to *TS* by default.

After selecting all the intervals containing the trunk part, we next extracted the trunk part by selecting the lower boundary height of the trunk part (*LT*), as shown in Fig. 5. To contain most of the trunk part and avoid the interference of noise and the head part, the second-highest lower boundary height of *TS* was taken as the value of *LT* (the last two intervals that did not satisfy Eq. (3) were automatically ignored), which showed a good effect on part segmentation.

Following the previous step, the head (including the neck part) and leg parts were obtained by the geometric relationships of these three parts in the side-view image. Two examples of the part segmentation results are shown in Fig. 6.

2.4.3. *Deep parts feature fusion*

In the past few years, CNNs have achieved great success in a variety of tasks by learning deep features, especially in visual recognition/classification tasks (He, Zhang, Ren, & Sun, 2016; G.; Huang, Liu, Van Der Maaten, & Weinberger, 2017; Krizhevsky, Sutskever, & Hinton, 2012; Simonyan & Zisserman, 2014; Szegedy et al., 2015). Considering the difficulties of manually defining discriminative features for cows

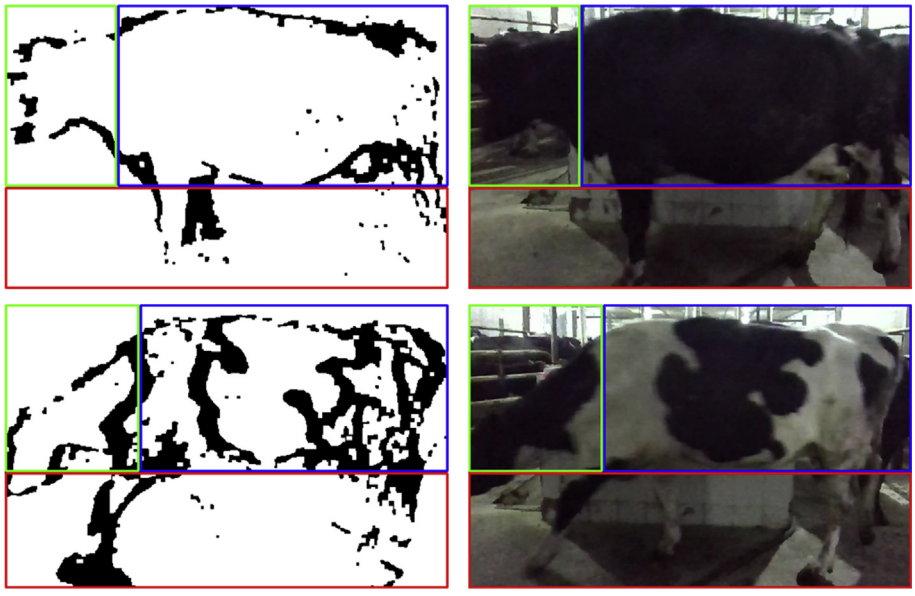

Fig. 6 – Two examples of part segmentation results.

with less texture information using traditional algorithms, CNNs are used to extract deep features of each cow part in this work. Based on the experimental results using different CNNs to identify cows in our previous work (Shen et al., 2019), we chose the network that achieved the highest performance (AlexNet; Krizhevsky et al., 2012), as the base network in this experiment. AlexNet is a typical CNN that had high classification performance in the ImageNet competition. The head, trunk, and leg parts of the cow object in the side-view image contain rich texture and contour features. A cow identification model called deep parts feature fusion (DPFF) was proposed. In this model, three fine-tuned AlexNet models were trained independently to extract deep features of three parts, due to the limitation of our computational resources. An estimated requirement of GPU memory for training three CNNs jointly was suggested to be at least 5 GB. For each CNN in our model, we used the images of the corresponding part as the input of the model, and the weights pre-trained on ImageNet as the initial parameters of the model to fine-tune a new model by updating the parameters of the last two layers of AlexNet. The structure of the DPFF model is shown in Fig. 7. Conv1 to Conv5 are convolutional layers, of which Conv1, Conv2, and Conv5 include a pooling layer, and FC6 to FC8 are fully connected layers. Each convolutional layer and the first two fully connected layers are followed by ReLU nonlinearity. We first trained three independent CNNs—using the training set of three parts to fine-tune the AlexNet model—to extract the deep feature (the feature of the softmax layer) from these three parts. Then, the features were fused by a weighted summation strategy to generate the final features of the cow identity. Finally, an SVM classifier was trained to identify individual cows.

To avoid overfitting, the dataset containing the head, trunk, and leg parts was augmented by horizontal flipping and random cropping in our experiments. The original images of each part of the cow in the dataset were input into AlexNet to train the model. The prediction was generated by the last fully connected layer after multiple convolution and pooling operations had been performed. There are 93 cows in our dataset, so the last fully connected layer of AlexNet was stacked by 93 neurons, and the output was fed to a 93-way softmax, which gave a distribution over 93 class labels.

After calculating the error between the prediction and the actual label, the stochastic gradient descent was used to update the weights and minimise the loss function. The final model was obtained when the model converged.

Next, the original images of each part of the cow in the dataset were input into the model for feature extraction (extracting the feature of the softmax layer), and the features were fused by a weighted summation strategy to generate the final features of the cow identity.

Finally, the fused features in the training set were used to train an SVM classifier. Then, the fused features in the validation set were input into this SVM model to obtain the final identification result.

3. Experimental results and analysis

3.1. *Identification results and analysis*

In this experiment, we trained three CNNs for the head, trunk, and leg parts and used the training set to complete the fine-tuning of the AlexNet model to achieve the feature extraction of the three parts. We set the parameters of three networks simultaneously, the batch size was set to 78, the dropout rate to 0.5, and the initial learning rate to 0.01. When the model loss function fluctuated slightly, the learning rate was set to 1/10 of the previous one. The kernel function of the SVM model was the radial basis function (RBF). The penalty coefficient C and gamma were optimal empirical values. In this work, we used accuracy as an evaluation metric, i.e., the percentage of correct predictions for the validation set. The experiments were performed on a hardware configuration with an Intel Core i7-7800 × 3.5 GHz CPU, 16 GB of memory, and an NVIDIA GeForce GTX 1060 GPU with 3 GB memory.

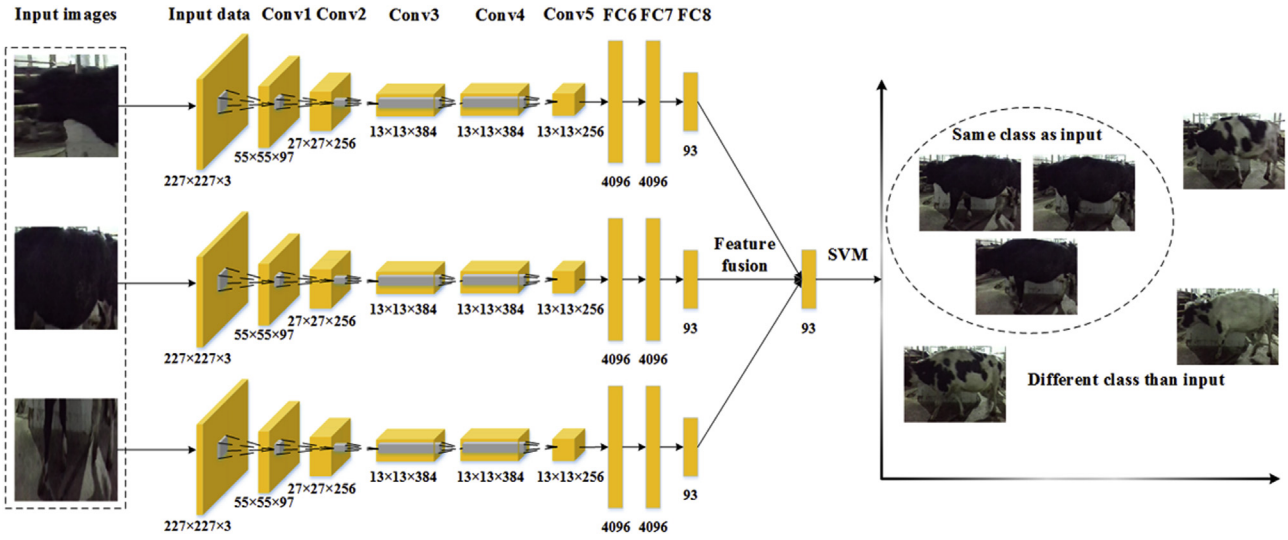

Fig. 7 – Structure of the Deep Parts Feature Fusion (DPFF) model.

Table 1 – Identification performance comparison of DPFF, related methods and existing works.

Method	Accuracy (%)	Object categories	Region of interest(ROI)
Gaber et al. (2016)	Approximately 99.00	31	Muzzle
Kumar et al. (2018)	98.99	500	Muzzle
Kumar et al. (2016)	95.87	300	Face
W. Li et al. (2017)	99.70	10	Tail-head
CNN for back part (Zin et al., 2018)	97.01	45	Back
Image comparison (Zhao & He, 2015)	32.95	30	Trunk
SIFT (Zhao & He, 2015)	37.89	30	Trunk
CNN for trunk part (Zhao & He, 2015)	90.55	30	Trunk
CNN for trunk part on our dataset	69.86	93	Trunk
CNN for cow object	96.65	105	Cow-side
CNN for cow object	97.78	93	Cow-side
Deep Parts Feature Fusion (DPFF)	98.36	93	Each part of the cow-side

Table 1 shows the identification performance comparison of our DPFF, related methods and existing works on the same evaluation metric.

In addition to the trunk that is closer to the ROI of this work, other ROIs such as muzzle, face, tail-head and back have been used in related works. As shown in Table 1, our method achieved a performance superior to all except the method based on the muzzle and tail-head region of cow. The work of Gaber et al. (2016) and Kumar et al. (2018) based on muzzle region obtained an identification accuracy of 99.00% and 98.99% respectively. However, muzzle image acquisition is difficult, and cow's muzzle is easily contaminated considering its small area. The work of Li et al. (2017) based on tail-head region achieved an accuracy of 99.70%, but this was only conducted over 10 cows. In contrast, the proposed DPFF used the side-view of cow, which can provide more discriminable information than the region of muzzle and tail-head, and achieved a competitive accuracy of 98.36% over 93 cows.

Zhao and He (2015) used two traditional identification methods—image comparison and SIFT—to compare with their proposed method (CNN for the trunk part). Comparing images is sensitive to light and image deformation because it is a pixel gray-value-based method. For SIFT, our preliminary experiments showed that the matched features of the method were mainly distributed in the background area of the side-view image of the cow. Moreover, the SIFT-based cow identification method mainly depends on corner detection and local feature description, so the accuracy was limited for cows with less texture. Therefore, the accuracy of these two traditional identification methods was low, as shown in Table 1.

CNNs can be used to extract deep features of cows, so Zhao and He (2015) used a modified LeNet5 (LeCun, Bottou, Bengio, & Haffner, 1998) to classify grayscale images of the trunk part of the cow. The method outperformed the image comparison and SIFT by 57.60% and 52.66%, respectively. We also implemented their model on our trunk image dataset without additional optimisation, with an accuracy of 69.86%. Possible reasons for the lower performance are that there were more cow categories in our dataset, and the main colour of many cows was black. When the trunk part is close to pure black, it is difficult to extract effective identity features, as in the work of Zhao and He (2015), so discriminative features such as the texture and contour of the cow need to be better extracted.

Considering that the cow object in the side-view image contains more contour features than the trunk part, the information on the head and legs cannot be ignored when the trunk part has no texture or less texture. Therefore, we used a fine-tuned AlexNet model to identify the cow object with these parts on two datasets containing 105 and 93 cows (Shen et al., 2019), and achieved an accuracy of 96.65% and 97.78%, respectively, which outperformed the work of Zhao and He (2015). The dataset of 93 cows was included in the dataset of 105 cows. The accuracy of this method on the dataset of 93 cows was higher than the accuracy on the dataset of 105 cows. One possible reason is that 12 extra cows lost their head part, as we mentioned in section 2.1, so the dataset of 93 cows had more discriminative information. However, not only the cow object but also each part of the side of the cow also contains rich texture and contour features with high discrimination, which can also be used to extract deep CNN features.

Therefore, the DPFF model fusing the deep CNN features from each part, proposed to address the abovementioned shortcomings of those methods, achieved an accuracy of 98.36%, which was the best performance of these methods.

From the experimental results in Table 1, it can be concluded that the method using the overall cow object is better than using the trunk part, and the method of fusing the deep CNN features from each part is better than the former two methods.

3.2. *Feature maps and analysis*

To verify whether our CNNs effectively extracted cow features, we output the overlaying images of four input images with their corresponding feature maps of the first convolutional layer. Some examples are shown in Fig. 8.

As seen in the 32 overlaying images, the model effectively extracted the features of each part of the cow, such as texture and contour. It also demonstrated the good filtering and smoothing effects of the convolution operation, which reduced noise and enhanced signal characteristics. In addition, input images were mapped into different gray spaces by the convolution operation, which resisted external interference under uneven lighting conditions. Therefore, the convolution operation improved the robustness of the network, especially in an unconstrained environment.

We also output the overlaying image of one cow object image with its feature maps of the last convolutional layer. Three examples are shown in Fig. 9. As seen in this figure, the head, trunk, and legs were activated, which demonstrated

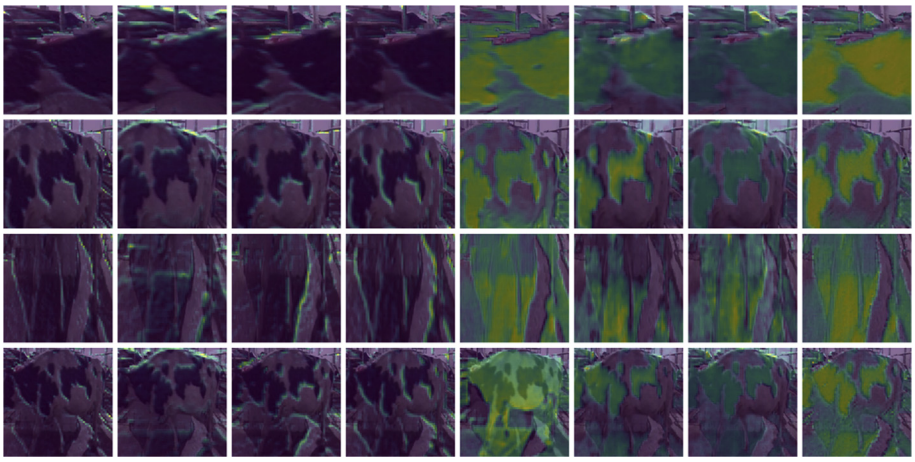

Fig. 8 – Overlaying images of four input images with their corresponding feature maps of the first convolutional layer. The first row to the fourth row are the overlaying images of the head, trunk, legs, and cow object.

Fig. 9 – Overlaying images of one cow object image and its corresponding feature maps of the last convolutional layer.

that each part of the side of the cow can be used to learn discriminative features for cow identification by means of CNNs.

3.3. *Part feature fusion analysis*

To analyse the contribution of each part and the multipart combinations to cow identification and their performance, we identified all possible combinations of cow head, trunk, and legs. We adopted three identification methods: conducting an explicit feature matching step by calculating feature similarity to achieve identification (hereafter referred to as DPFF-matching); directly using eigenvalues of the fused feature of the softmax layer to achieve identification (hereafter referred to as DPFF-softmax); and the proposed SVM-based method (DPFF). For feature fusion, we adopted a strategy of weighting and summing the corresponding feature values of the softmax layer (hereafter referred to as Weighted sum), and the weight coefficients $w_i (i = 1, 2, 3)$ of three parts were all possible combinations with a step size of 0.1. Table 2 shows the identification performance and the best CPU time of different combinations of parts within the three identification methods mentioned above.

In this paper, DPFF-matching was used to validate the deep learning method commonly used in the human biometrics literature, and the reciprocal of the commonly used Euclidean distance was used as similarity. In Table 2, "CPU time" indicates the optimal CPU time for the corresponding identification method. As shown in this table, DPFF with the fusion of three deep parts features achieved the best accuracy. Although DPFF-softmax achieved superior performance with single part feature, its performance was inferior to DPFF with the deep parts fusion scheme. DPFF-matching had a similar performance to DPFF-softmax, but it was also the most time consuming method. This is because when using DPFF-matching for testing, all the images in the database needed to be retrieved. In the case where the number of cows is fixed, this waste of time is not necessary. DPFF-matching can be

Table 2 – Identification performance of different combinations of parts within different methods.

Method	Accuracy (%)							CPU time (s)
	Head	Trunk	Legs	Head + Trunk	Head + Legs	Trunk + Legs	Head + Trunk + Legs	
DPFF-matching	79.70	94.79	84.11	97.26	90.14	96.99	97.81	1.687
DPFF-softmax	80.51	97.44	86.67	96.99	88.77	96.99	97.81	0.031
DPFF	72.88	93.98	84.66	96.71	90.14	97.26	98.36	0.094

Table 3 – Identification performance of the three fusion strategies.

Fusion strategy	Accuracy (%)		
	DPFF-matching	DPFF-softmax	DPFF
Sum	98.09	97.53	97.53
Max	97.81	96.71	97.81
Weighted sum	97.81	97.81	98.36

regarded as a kind of one-shot method (Li, Fergus, & Perona, 2006). Training three CNNs jointly for DPFF-matching or introducing a Siamese network (Taigman, Yang, Ranzato, & Wolf, 2014) for an end-to-end learning can be explored to further improve the identification accuracy, and the estimated GPU memory required by these two schemes should be at least 5 GB.The experimental results of Table 2 proved that DPFF method has the advantages of high accuracy and high efficiency.

For a single part, from the experimental results in the single part of Table 2, it can be seen that the head and legs also have the capacity to identify different cows, in addition to the trunk part, and the accuracy of single part identification using the DPFF-softmax method was higher than DPFF-matching and DPFF.

For multiple parts, the relevant weights of the multipart combinations in Table 2 were the optimal empirical parameters. The weight coefficient ratio with the combination of head + trunk in the case of using DPFF-matching, DPFF-softmax and DPFF models was 2:8/4:6, 2:8/3:7/4:6 and 2:8, respectively. Correspondingly, the weight coefficient ratio with the combination of head + legs was 4:6/5:5, 4:6/5:5 and 4:6, with the combination of trunk + legs was 7:3, 7:3 and 7:3/6:4, and with the combination of head + trunk + legs was 2:7:1/2:6:2/2:5:3, 2:6:2 and 2:5:3. From the experimental results in the multipart of Table 2, it can be seen that the multipart identification using the DPFF model was generally better than DPFF-matching and DPFF-softmax. The DPFF model achieved 0%, 0.27%, 0.55% and 1.39%, 0.27%, and 0.55% improvements in performance compared to DPFF-matching and DPFF-softmax with the combination of head + legs, trunk + legs, and head + trunk + legs, respectively.

From the experimental results and the weight coefficient ratios of the multipart combinations, it can be found that the trunk part had the highest contribution to the individual identification of cows, followed by the legs and finally the head. This matched our expectations. First, the trunk was the largest part, and the texture information was the most abundant. Second, by analysing the image dataset, we found that the leg part was more discriminative than the head part, despite the relatively close effective areas of the two parts. One possible reason for this is that the texture information of the head part in the side-view image was insufficient, and most cows' heads were close to pure black.

In the end, we found that the identification method of fusing the deep CNN features from head, trunk, and leg parts achieved the highest accuracy of 98.36%, which proved that the individual identification of fusing the deep CNN features from each part was better than that of directly using a single part or the combinations of any two parts.

3.4. *Feature fusion strategy and analysis*

When performing deep parts feature fusion, in addition to using Weighted sum, we also tried the other two fusion strategies: summing (hereinafter referred to as Sum) and maximising (hereinafter referred to as Max) the corresponding feature values. The identification performance of the three fusion strategies is shown in Table 3.

The DPFF method with the Weighted sum strategy achieved the highest accuracy. When using different fusion strategies, the classification using the SVM model was

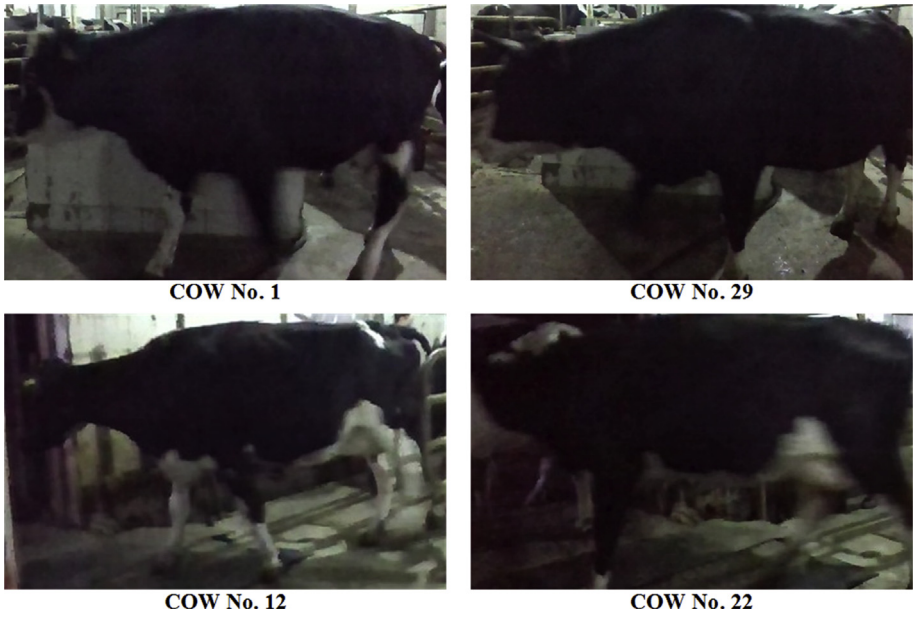

Fig. 10 – Error examples. The left column is the validation sample, and the right column is the corresponding misidentified cow.

generally better than directly using eigenvalues of the fused feature of the softmax layer. Among the three fusion strategies, the performance of the Weighted sum was generally better than the other two fusion strategies.

3.5. *Error analysis and future improvements*

In the validation set, 6 samples were predicted incorrectly. Figure 10 shows two typical error examples: COW No. 1 was misidentified as No. 29, and COW No. 12 was misidentified as No. 22.

By analysing the errors, we found that there are two main types of cow in the error samples: the cow with a mainly black colour and the cow with small individual differences from the misidentified cow. For the cow with a mainly black colour, error easily occurs because it is difficult to extract effective identity features of this type of cow. Deep edge features could be integrated into the identification process. Alternatively, markers could be added to similar black cows to improve the difference. For cows with small individual differences, a fine-grained classification method could be introduced to identify this type of cow.

Compared to our previous work (Shen et al., 2019), we have not only increased the accuracy from 96.65% to 98.36% in this work but also solved the error of blurry images. This is because we used feature fusion to extract the discriminative features.

4. Conclusion

In this paper, a non-contact cow identification method based on fusion of deep parts features has been proposed. YOLO was first applied to extract the cow object in the side-view image, and the cow's head, trunk, and leg parts were extracted by a part segmentation algorithm using frame differencing and segmentation span analysis. Then, three independent fine-tuned AlexNet models were used to extract the deep features of these three parts, and the features were fused by a weighted summation strategy. Finally, an SVM classifier was trained to classify the cow images. A dataset containing side-view images of 93 different cows was created in our experiment, and our DPFF model achieved 98.36% cow identification accuracy, which outperformed existing works. We verified that the performance of DPFF is better than the deep learning methods using the overall cow object, a single part of a cow or the combinations of any two parts. This study showed that our proposed method can improve identification performance and practicality. Furthermore, the method can be directly combined with related works such as lameness detection, drinking behaviour analysis, linear appraisal, mastitis detection and individual localisation. It can also provide a reference for cow identification based on other views, and be applied to other fields of dairy cow husbandry, to implement precision livestock farming. In summary, our proposed method can help farmers achieve many benefits, including: cost savings, labour savings, better cow health monitoring, more accurate and individualised behaviour detection and analysis, etc.

Declaration of Competing Interest

The authors declare that they have no conflict of interest in relation to the work in this article.

Acknowledgments

This work was partly supported by the National Key Research and Development Program of China (grant number 2016YFD0700204), the Academic Backbone Project of Northeast Agricultural University of China (grant number 17XG20), the National Natural Science Foundation of China (grant number 31902210), the Natural Science Foundation of Heilongjiang Province of China (grant number QC2018074), the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province of China (grant number UNPYSCT-2018142), the Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs of China (grant number 2018AIOT-02), and the China Agriculture Research System (grant number CARS-36).

REFERENCES

Alvarez, J. R., Arroqui, M., Mangudo, P., Toloza, J., Jatip, D., Rodríguez, J. M., et al. (2018). Body condition estimation on cows from depth images using Convolutional Neural Networks. Computers and Electronics in Agriculture, 155, 12–22.

Barbari, M., Conti, L., & Simonini, S. (2009). Spatial identification of animals in different breeding systems to monitor behavior. In *Livestock environment VIII, 31 august–4 september 2008, iguassu falls, Brazil* (p. 135). American Society of Agricultural and Biological Engineers.

Chen, X., Mottaghi, R., Liu, X., Fidler, S., Urtasun, R., & Yuille, A. (2014). Detect what you can: Detecting and representing objects using holistic models and body parts. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1971–1978).

DeVries, T. J., Von Keyserlingk, M. A. G., Weary, D. M., & Beauchemin, K. A. (2003). Validation of a system for monitoring feeding behavior of dairy cows. *Journal of Dairy Science, 86*(11), 3571–3574.

Gaber, T., Tharwat, A., Hassanien, A. E., & Snasel, V. (2016). Biometric cattle identification approach based on weber's local descriptor and adaboost classifier. *Computers and Electronics in Agriculture, 122*, 55–66.

Guo, H., Ma, X., Ma, Q., Wang, K., Su, W., & Zhu, D. (2017). LSSA_CAU: An interactive 3d point clouds analysis software for body measurement of livestock with similar forms of cows or pigs. *Computers and Electronics in Agriculture, 138*, 60–68.

Gygax, L., Neisen, G., & Bollhalder, H. (2007). Accuracy and validation of a radar-based automatic local position measurement system for tracking dairy cows in free-stall barns. *Computers and Electronics in Agriculture, 56*(1), 23–33.

Haggag, H., Abobakr, A., Hossny, M., & Nahavandi, S. (2016). Semantic body parts segmentation for quadrupedal animals. In *2016 IEEE international conference on systems, man, and cybernetics (SMC)* (pp. 855–860). IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *CVPR* (Vol. 1, p. 3).
Huhtala, A., Suhonen, K., Mäkelä, P., Hakojärvi, M., & Ahokas, J. (2007). Evaluation of instrumentation for cow positioning and tracking indoors. *Biosystems Engineering, 96*(3), 399–405.
Ipema, A. H., Van De Ven, T., & Hogewerf, P. H. (2013). Validation and application of an indoor localization system for animals. In *Proceedings of 6th European conference on precision livestock farming, leuven, Belgium* (pp. 10–12).
Jabbar, K. A., Hansen, M. F., Smith, M. L., & Smith, L. N. (2017). Early and non-intrusive lameness detection in dairy cows using 3-dimensional video. *Biosystems Engineering, 153*, 63–69.
Jiang, B., Wu, Q., Yin, X., Wu, D., Song, H., & He, D. (2019). FLYOLOv3 deep learning for key parts of dairy cow body detection. *Computers and Electronics in Agriculture, 166*, 104982.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).
Kühl, H. S., & Burghardt, T. (2013). Animal biometrics: Quantifying and detecting phenotypic appearance. *Trends in Ecology & Evolution, 28*(7), 432–441.
Kumar, S., Pandey, A., Satwik, K. S. R., Kumar, S., Singh, S. K., Singh, A. K., et al. (2018). Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement, 116*, 1–17.
Kumar, S., Tiwari, S., & Singh, S. K. (2016). Face recognition of cattle: Can it be done? *Proceedings of the National Academy of Sciences India Section A: Physical Sciences, 86*(2), 137–148.
Lavy, A., Eyal, G., Neal, B., Keren, R., Loya, Y., & Ilan, M. (2015). A quick, easy and non-intrusive method for underwater volume and surface area evaluation of benthic organisms by 3D computer modelling. *Methods in Ecology Evolution, 6*(5), 521–531.
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324.
Li, F., Fergus, R., & Perona, P. (2006). One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(4), 594–611.
Li, W., Ji, Z., Wang, L., Sun, C., & Yang, X. (2017). Automatic individual identification of Holstein dairy cows using tailhead images. *Computers and Electronics in Agriculture, 142*, 622–631.
Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740–755). Springer.
Nir, O., Parmet, Y., Werner, D., Adin, G., & Halachmi, I. (2018). 3D Computer-vision system for automatically estimating heifer height and body mass. *Biosystems Engineering, 173*, 4–10.
Pezzuolo, A., Guarino, M., Sartori, L., & Marinello, F. (2018). A feasibility study on the use of a structured light depth-camera for three-dimensional body measurements of dairy cows in free-stall barns. *Sensors, 18*(2), 673.
Porto, S. M. C., Arcidiacono, C., Anguzza, U., & Cascone, G. (2013). A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns. *Biosystems Engineering, 115*(2), 184–194.
Porto, S. M. C., Arcidiacono, C., Anguzza, U., & Cascone, G. (2015). The automatic detection of dairy cow feeding and standing behaviours in free-stall barns by a computer vision-based system. *Biosystems Engineering, 133*, 46–55.
Porto, S. M. C., Arcidiacono, C., Giummarra, A., Anguzza, U., & Cascone, G. (2014). Localisation and identification performances of a real-time location system based on ultra wide band technology for monitoring and tracking dairy cow behaviour in a semi-open free-stall barn. *Computers and Electronics in Agriculture, 108*, 221–229.
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788).
Salau, J., Haas, J. H., Junge, W., & Thaller, G. (2017). A multi-Kinect cow scanning system: Calculating linear traits from manually marked recordings of Holstein-Friesian dairy cows. *Biosystems Engineering, 157*, 92–98.
Shen, W., Hu, H., Dai, B., Wei, X., Sun, J., Jiang, L., et al. (2019). Individual identification of dairy cows based on convolutional neural networks. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-019-7344-7.
Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. ArXiv Preprint ArXiv:1409.1556.
Spoliansky, R., Edan, Y., Parmet, Y., & Halachmi, I. (2016). Development of automatic body condition scoring using a low-cost 3-dimensional Kinect camera. *Journal of Dairy Science, 99*(9), 7714–7725.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).
Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1701–1708).
Tsai, D.-M., & Huang, C.-Y. (2014). A motion and image analysis method for automatic detection of estrus and mating behavior in cattle. *Computers and Electronics in Agriculture, 104*, 25–31.
Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision, 57*(2), 137–154.
Voulodimos, A. S., Patrikakis, C. Z., Sideridis, A. B., Ntafis, V. A., & Xylouri, E. M. (2010). A complete farm management system based on animal identification using RFID technology. *Computers and Electronics in Agriculture, 70*(2), 380–388.
Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology, 87*(3), 533–545.
Yukun, S., Pengju, H., Yujie, W., Ziqi, C., Yang, L., Baisheng, D., et al. (2019). 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. *Journal of Dairy Science, 102*(11), 10140–10151.
Zhao, K., & He, D. (2015). Recognition of individual dairy cattle based on convolutional neural networks. *Transactions of the Chinese Society of Agricultural Engineering, 31*(5).
Zin, T. T., Phyo, C. N., Tin, P., Hama, H., & Kobayashi, I. (2018). Image technology based cow identification system using deep learning. In *Proceedings of the international MultiConference of engineers and computer scientists* (Vol. 1).