2020 16th IEEE International Conference on Automation Science and Engineering (CASE)
August 20-21, 2020, Online Zoom Meeting

BiLSTM-based Individual Cattle Identification for Automated Precision Livestock Farming

Yongliang Qiao, Daobilige Su, He Kong, Salah Sukkarieh, Sabrina Lomax and Cameron Clark

***Abstract*— Individual cattle identification plays an important role for automation in precision livestock management. Existing methods for cattle identification require radio frequency and visual ear tags, all of which are prone to loss or damage. In this work, we propose a deep learning-based framework to identify beef cattle using image sequences, unifying merits of both Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) network methods. A CNN (Inception-V3) was used to extract features from a video dataset taken of the rear-view of cattle, after which extracted features were fed to a BiLSTM layer to capture spatial-temporal information enabling the identification of each individual animal. A total of 363 rear-view videos of 50 cattle were collected for our dataset. The proposed method achieved 91% identification accuracy using a 30-frame video length, improving that of Inception-V3 use or LSTM. Additionally, increasing video sequence length to 30-frames enhanced identification performance. Our approach can use spatial-temporal features to identify cattle, and enables automated identification for precision livestock farming.**

I. Introduction

In precision livestock management, individual cattle identification is required for automated analysis of animal activities and productivity [1], [2]. With accurate individual cattle information, the welfare and productivity of each animal can be evaluated and optimized to improve productivity [3], [4]. Existing cattle identification methods mainly adopt on-animal sensors such as ear-tags, collar, and radio frequency identification modules, which incur a cost and can be lost [6]–[8].

With the development of visual sensors and image processing technologies, vision-based cattle identification as a non-contact approach is gaining interest [9]. For example, biometric and visual features, extracted from images of cattle muzzles, faces, coats, retinas, and irises, have been shown to be helpful for identifying cattle due to their uniqueness and immutability [10]. Similarly, Kusakunniran et al. proposed an automatic cattle identification approach by fusing visual features extracted from muzzle images [11], while Andrew et al. (2016) utilized the cattle coat patterns for identification [6]. In addition, Okura et al. (2019) unified gait and texture features for cow identification based on 3D video analysis using RGB-D cameras [12]. However, this work manually selected and extracted features, which is impacted by illumination or camera viewpoints [13].

In recent years, deep learning approaches with powerful feature extraction and image representation capabilities have been widely used in visual recognition and image segmentation [5], [15]. As a result, deep learning-based cattle identification approaches are being explored [16]. For example, Andrew et al. (2017) used deep neural architectures to exploit unique coat markings for Holstein Friesian cattle identification [17]. Moreover, Shen et al. (2019) proposed a CNN based identification method for dairy cattle, achieving 97% accuracy from 105 side-view images [18]. In addition, Zin et al. (2018) trained a deep learning model based on back images to identify cows [9].

Most deep learning-based approaches rely on extracted CNN features from images, ignoring important temporal information. As cattle videos contain both spatial and temporal information, here we use video data for cattle identification using both CNNs [14] and BiLSTM [20] networks to extract spatial information. To reveal the hidden patterns and features in time-space data [19], [21], respectively, we propose to bring together these methods to increase cattle identification accuracy, as shown in Fig.1.

Here we (1) provide a BiLSTM based deep learning approach for cattle identification, capable of automatically learning spatial-temporal features; (2) determine the impact of video sequence length on identification accuracy; (3) compare Inception-V3, LSTM approaches with the BiLSTM deep learning approach.

II. Related Work

Four main cattle body areas are often used for visual identification namely the muzzle, face, back and torso. For example, Gaber et al. (2016) used the Weber Local Descriptor to extract robust features from cattle muzzle print images for cattle identification [22]. Cai et al. (2013) presented a facial representation model of cattle based on local binary pattern texture features [23]. The work in [24] also relied on extracted cattle facial features.

Recently, deep learning approaches have become popular for visual cattle classification and recognition [9], [18]. After detecting trunks from raw images, Zhao et al. (2015) proposed a CNN network method for cow identification [25]. In addition, Kumar et al. (2018) proposed a CNN based approach by using primary muzzle point image pattern [16]. This approach ignored the head and legs which also include useful information such as contour and gait features. In the

Yongliang Qiao, Daobilige Su, He Kong, and Salah Sukkarieh are with the Australian Centre for Field Robotics, University of Sydney, Australia. E-mails: {yongliang.qiao, daobilige.su, he.kong, salah.sukkarieh}@sydney.edu.au.

Sabrina Lomax and Cameron Clark are with Livestock Production and Welfare Group, School of Life and Environmental Sciences, University of Sydney, Australia. E-mails:{sabrina.lomax, cameron.clark}@sydney.edu.au

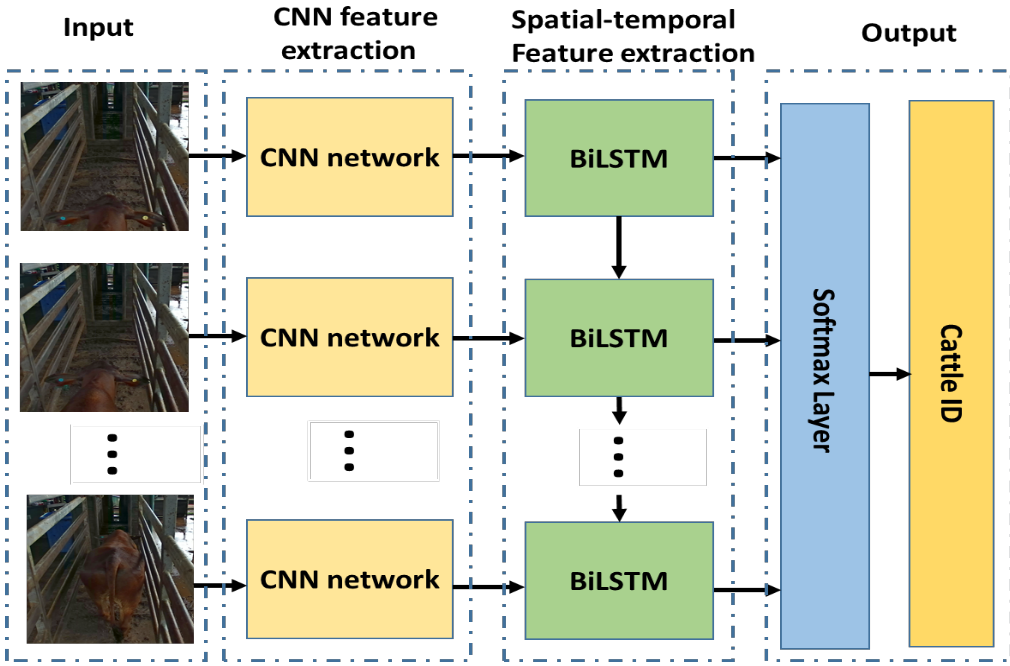


Fig. 1: The overall structure of the proposed cattle identification system.

work of Shen et al. (2019) [18], a CNN network was fine-tuned with side-view images of dairy cows and used for cow identification.

Despite the above progress, existing image frame-based approaches have not utilized behaviour such as posture or gait information. Time-series data such as video and inertial measurement units (IMU) measurements, on the other hand, contain more temporal information. Hence, they have been widely used to monitor cattle behaviour and welfare [26], [27]. For example, Van et al. (2014) developed an automated lameness scoring algorithm based on 3D-video recordings of cow gait [28]. Mcphee et al. (2017) developed a learning-based approach for assessing traits such as rump fat and muscle score [29]. Peng et al. (2019) developed a recurrent neural network (RNN) with a Long Short-Term Memory (LSTM) model to monitor and classify cattle behavior patterns using IMU measurements [30]. Andrew et al. (2017) demonstrated a video processing pipeline for cattle identification and adopted a Long-term Recurrent Convolutional Network to classify cattle videos taken by the unmanned aerial vehicles [17]. This work highlights the paucity of data for beef cattle, potentially due to the difficulty of this task as most beef cattle have uniform coat colours. In our recent work, Qiao et al. (2019) proposed a beef cattle identification framework which uses image sequences unifying the advantages of both CNN and LSTM [4]. In the current paper, we will also focus on beef cattle, and synthesize both CNN and BiLSTM for cattle identification using video datasets, thereby further improving the results of Qiao et al. (2019) [4].

III. The Proposed Framework

As illustrated in Fig. 1, the proposed approach consists of two main steps: 1) a CNN based feature extraction step; and 2) a spatial-temporal feature extraction step based on BiLSTM. More specifically, at each time step (image frame), a set of CNN features, which describe the visual content information of cattle, are firstly extracted from video data. Based on the extracted features, the BiLSTM layer is applied to capture spatial-temporal features between time steps. As BiLSTM has two memory functions, it is efficient in learning temporal features by considering the order dependencies between sequence elements. Hence, the spatial-temporal information and the motion pattern of the cattle can be efficiently captured. Finally, the spatial-temporal features are fed into softmax to output cattle ID number.

A. CNN-based Feature Extraction

An Inception-V3 model was used to transfer learning [31]. This model comprises building blocks including several layers of convolutions, average pooling, max pooling, concats, dropouts, and fully connected layers. For a given image l at time T and ConvNet learned parameter w, the CNN feature was extracted by:

$$X(t) = f_n(\ldots f_2(f_1(l; w_1); w_2)\ldots), w_n) \tag{1}$$

where, $f_1, \cdots, f_n$ are the corresponding layer functions. Each layer output was a CNN feature for image representation with the low layers retaining high spatial resolution whilst the high layers containing more semantic information. To maintain system performance and efficiency, the features extracted from the final pool layer in Inception-V3 were used in our work. The Inception-V3 model with the pre-trained weights on the ImageNet dataset [32] was used to extract cattle features. Thus, each image in the video had a 2048-dimensional feature before being passed to the BiLSTM model.

B. BiLSTM for Spatial-temporal Feature Extraction

To make the most of the video contents, one should consider the visual aspects characterizing the object appearances as well as the motion present within the data. As such, after CNN features are extracted image by image from video, the next step is to extract spatial-temporal features.

LSTM is a popular network for space-time data processing with strong abilities to learn and remember over long sequences of input data [21]. It extends RNN with memory cells. These cells usually have few linear interactions making the information maintaining process easier [27]. Moreover, LSTM makes use of the “gating" concept to update cell states. Each gate is a nonlinear summation unit which controls the operation of the cell memory such as write (input gate), read (output gate) or reset (forget gate) [33].

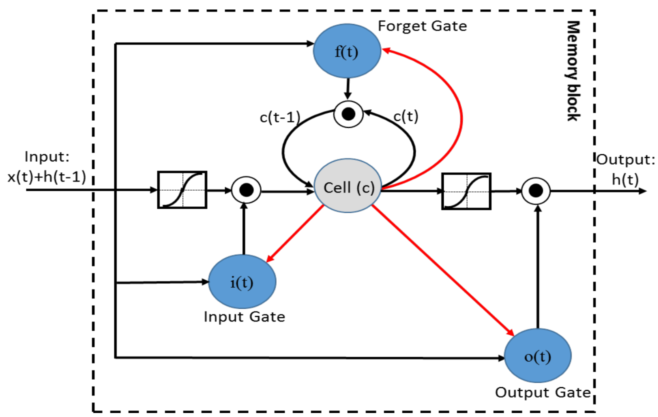


Fig. 2: LSTM cell

As illustrated in Fig. 2, a common LSTM unit consists of a cell, an input gate, an output gate and a forget gate. At every time step t, given the input data x, the computation of the hidden value h_t of an LSTM cell is updated as follows:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{2}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{3}$$

$$c_t = f_t c_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{4}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \tag{5}$$

$$h_t = o_t tanh(c_t)) \tag{6}$$

where, σ is the sigmoid function; tanh represents the hyperbolic tangent activation function; i, f, o and c are the input gate, forget gate, output gate and cell activation vectors, respectively; h is the hidden vector; b denotes bias vectors and matrix W is the connection weight between two units.

Considering that cattle often move back and forth, to overcome the shortcoming of single LSTM cell that can only capture previous information, BiLSTM based cattle identification is proposed to utilize future information. The

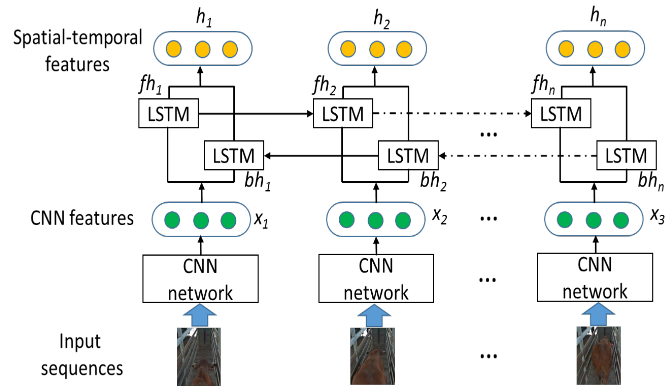


Fig. 3: The architecture of the proposed BiLSTM for cattle spatial-temporal feature extraction.

proposed BiLSTM model, as illustrated in Fig. 3, consists of two independent LSTMs, which can sum up information from forward and backward directions of a sentence, and merge the information coming from the two directions.

More specifically, at each time t, given a video frame length n, with the input CNN features denoted as $[x_1, x_2, \cdots, x_n]$, the forward LSTM computes the hidden vector fh_t based on the previous hidden vector fh_{t-1} and the input CNN feature x_t. The backward LSTM computes the hidden vector bh_t based on the opposite previous hidden vector bh_{t-1} and the input CNN feature x_t. Finally, the forward hidden vector fh_t and the backward hidden vector bh_t are concatenated into the final hidden vector h_t:

$$fh_t = \sigma(W_{xi}x_t + W_{hi}fh_{t-1} + +b_i) \tag{7}$$

$$bh_t = \sigma(W_{xi}x_t + W_{hi}bh_{t-1} + +b_i) \tag{8}$$

$$h_t = W_{yh}fh_t + W_{yh}bh_t \tag{9}$$

BiLSTM, with the extracted CNN features as its input is capable to learn and model each cattle’s unique temporal characteristics. In our work, x_t is a 2048-dimensional CNN feature (see Section III. A), and the network has 2 BiLSTM layer with 2048 cells, a 4096 projection layer for dimensionality reduction and a final softmax output layer.

C. Cattle Identification

The spatio-temporal features generated by BilSTM layer represents each cattle video, and they are fed to a softmax classifier for the final cattle identification. For the cattle video, the probability value of the cattle IDs is obtained by the Softmax layer:

$$P_i = \frac{e^{y_t^i}}{\sum_{j=1}^{n} e^{y_t^i}} \tag{10}$$

where, y_t is model output, index i corresponding to the largest probability value P_i is the final identification result. We take the output with the maximum value (class confidences, the value is between 0 and 1) as the cattle ID. If it matches the ground truth, then it will be regarded as a true result. Otherwise, it is a false result.

IV. Experimental Setup

A. Data Acquisition

Cattle data was collected at a feedlot in southwest Queensland, Australia. The data acquisition system, as shown in Fig. 4, was placed nearby the cattle crush during data collection. This system comprised two stereo ZED cameras (with 110° horizontal field of view), a GPU-equipped embedded PC (Neousys Nuvo 6108GC) and a high-speed volume data storage disk. In our experiment, the left image of the rear view ZED camera was used; the image resolution was set to 1920×1080. A high frame acquisition rate (30fps) was adopted to reduce the influences of motion blur during the herding process from the pen to the crush.

The experiment was conducted on 20 March, 30 April and 30 May 2018 at three different times whilst the same cattle were at feedlot. The cattle were moved to the crush from feed pens for weighing. The live weight of cattle varied between 330 to 550 kg as body size and mass changed across the feed period. Standard deviation of weight in the training dataset was 45 kg.

Fig. 4: Experimental setup at feedlot showing rear view camera (clamped to overhead bar) and side view camera (on tripod).

Given the large size (i.e. 1920×1080) of the original images, in order to improve system efficiency, we extracted and saved the central part of the original image as Region of Interest (ROI). Thus a size of 401× 506 image was obtained after image ROI extraction. In our experiment, a total of 363 individual videos from 50 cattle were used, with each video containing 40-frame long spatio-temporal streams, with frame size of 401×506. For training, the identity of individual cattle were manually recorded according to their ear tags. Illumination conditions, animal' posture changes and the complex background (including the cattle crush and ground) were commercial system challenges that needed to be overcome.

B. Network Training

Keras [34] was used to construct the cattle identification model. Details of hardware information for the current experiment are provided in Table I.

TABLE I: The experimental hardware

Hardware	Type
CPU	Intel Xeon E5-2630 @ 2.20 GHz×20
Memory	32GB
GPU	GeForce GTX 1080 Ti
Hard disk	1 TB

288 videos of 50 cattle were used for training while the remaining 75 videos of the same 50 cattle were used for testing. In the proposed BiLSTM based cattle identification approach, the final pool layer of Inception-V3 was used to extract CNN features. For each image, 2048 dimensional CNN features were obtained. The CNN features for each frame in the videos were then used to train the cattle identification model. For network training, initial learning rate was 0.00005, learning decay factor was 10^{-6}, and loss function was “categorical crossentropy".

V. Results and Discussion

A. Comparison of Different Methods

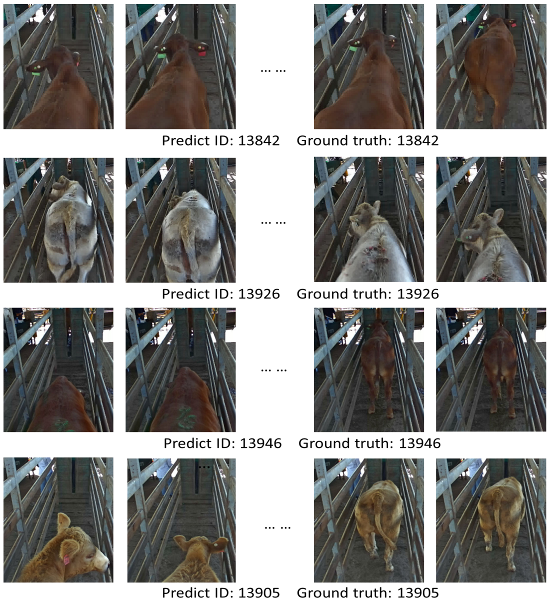
We compared the proposed BiLSTM based cattle identification approach with the Inception-V3 and LSTM approaches. In the Inception-V3 method, we froze the parameters of the first 172 layers and only retrained the last two inception blocks. In LSTM, the network had a LSTM layer with 2048 cells, a 4096 projection layer for dimensionality reduction, a dropout layer (dropout rate is 0.5), and a final softmax output layer. The number of output nodes in the last layer was equal to the number of cattle (50 in our dataset). For comparison, all methods were trained and tested using the same dataset.

The accuracy and processing time of Inception-V3, LSTM against the proposed BiLSTM based approach are compared in Table II (sequence length of 30 frames). The BiLSTM based cattle identification method achieved 91% accuracy for 50 cattle, outperforming LSTM (89%) and Inception-V3 (81%). Differences between methods can be attributed to the BiLSTM model containing a forward LSTM and backward LSTM neural network model, capturing more spatial-temporal information for the identification of cattle.

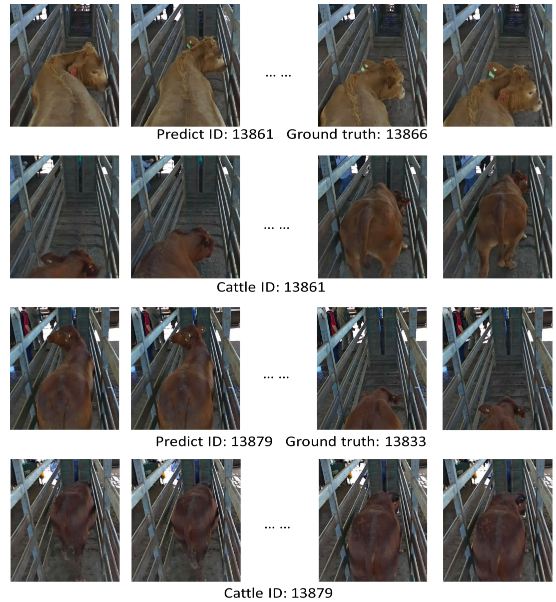
TABLE II: Accuracy comparison between different cattle identification methods

Methods	Accuracy (%)	Time (s)
Inception-V3 (Single frame)	81	10.8
LSTM	89	13.1
BiLSTM	**91**	22.3

True and false cattle identification examples of the proposed BiLSTM are provided in Fig. 5. Overall performance of the proposed method is favourable with the exception of cattle that were partly covered with mud. In Fig. 5 (a), cattle with different coat colours (e.g. white, brown, yellow) were successfully identified as well as cattle with similar coat colours (e.g. cattle ID 13842 and 13946). Thus, the proposed



(a) True identification examples



(b) False identification examples and their ground truth videos.

Fig. 5: Examples of BiLSTM based cattle identification results. The video length is 30 frames. Note: for each video, only first two and the last two frames are displayed; some cattle were walking back and forth during data acquisition. For each false identification, its corresponding ground truth video is also displayed.

method can take advantage of information memory ability to learn spatial-temporal features for cattle identification rather than only using visual information.

For the false identification shown in Fig. 5 (b), these cattle were standing static or made very little movement in the videos. In this situation, the cattle identification accuracy was low due a paucity of temporal information. In general, more motion patterns in the videos provide richer spatial-temporal features to improve identification accuracy.

B. Performance Comparison for Different Video Length

The impact of different video lengths (i.e. 10, 20, 30 and 40 frame length) with respect to cattle identification accuracy are provided in Fig. 6. Overall performance of BilSTM was better than that of LSTM. The greatest accuracy of BiLSTM was 91% which outperforms LSTM (89%). Additionally, in the stage of 0 to 30 frames, accuracy of both LSTM and BiLSTM improved with increased sequence length as more useful spatial-temporal features could be extracted. However, increasing the duration of video past 30 seconds appears unnecessary.

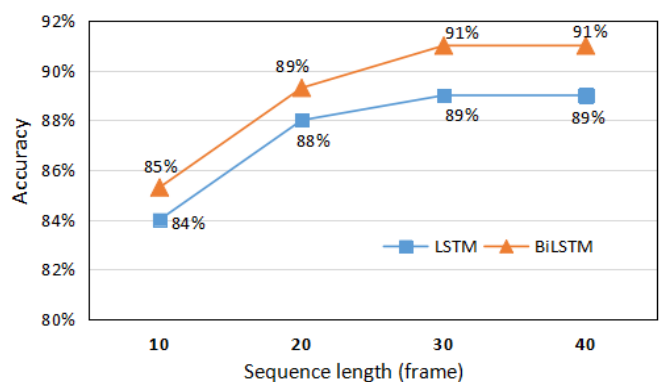


Fig. 6: Cattle identification Accuracy w.r.t. sequence length.

VI. Conclusion and Future Work

A BiLSTM based cattle identification approach using video data is proposed consisting of two essential parts; CNN feature extraction and spatial-temporal feature extraction. Firstly, the visual features were extracted using the Inception-V3 CNN network. Then BiLSTM layer was applied to extract spatial-temporal features for the final cattle identification. achieving 91% identification accuracy using 30-frame length videos which is better than that of using only LSTM (89%) or Inception-V3 (81%). Using video lengths above 30 seconds appears unnecessary.

Although the proposed approach is favourable for cattle identification, performance is poor for cattle with low amounts of motion. In future work, we will focus on further improving cattle identification accuracy using a larger and more complex video dataset.

Acknowledgments

The authors acknowledge the support of the Meat & Livestock Australia Donor Company through the project: Objective, robust, real-time animal welfare measures for the Australian red meat industry (P.PSH.0819). The authors also express their gratitude to Khalid Rafique, Javier Martinez, Amanda Doughty, Ashraful Islam, and Mike Reynolds for their help in experiment organization and data collection.

References

[1] T. M. Banhazi, H. Lehr, J. L. Back, H. Crabtree, P. Schofield, M. Tscharke and D. Berckmans, Precision livestock farming: an international review of scientific and commercial aspects, *Int. Journal of Agricultural and Biological Eng.*, Vol. 5, No. 3, pp. 1–9, 2012.

[2] D. Berckmans, Receding horizon estimation and control with structured noise blocking for mobile robot slip compensation Precision livestock farming technologies for welfare management in intensive livestock systems, *Scientific and Technical Review of the Office International des Epizooties*, Vol. 33, No. 1, pp. 189–196, 2014.

[3] B. K. McCabe, S. Tait, S. Marchuk, S. Lee, D. L. Antille, , J. Eberhard and C. P Baillie, Identification and analysis of organic waste produced from Australian intensive livestock industries: opportunities for aggregation and treatment via co-digestion, *Proc. of Annual Int. Meeting American Society of Agricultural and Biological Engineers*, pp. 1–17, 2019.

[4] Y. Qiao, D. Su, H. Kong, S. Sukkarieh, S. Lomax, and C. Clark, Individual cattle identification using a deep learning based framework, *Proc. of the 6th IFAC AgriControl*, pp. 318–323, 2019.

[5] Y. Qiao, D. Su, H. Kong, S. Sukkarieh, S. Lomax, and C. Clark, Data Augmentation for Deep Learning based Cattle Segmentation in Precision Livestock Farming, *Proc. of the 16th IEEE International Conference on Automation Science and Engineering*, to appear, 2020.

[6] W. Andrew, S. Hannuna, , N. Campbell, and T. Burghardt, Automatic individual holstein friesian cattle identification via selective local coat pattern matching in RGB-D imagery, *Proc. of IEEE Int. Conf. on Image Processing*, pp. 484–488, 2016.

[7] D. He, D. Liu, and K. Zhao, Review of perceiving animal information and behavior in precision livestock farming, *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 47, No. 5, pp. 231–244, 2016.

[8] A. C. Arslan, M. Akar, and F. Alagoz, 3D cow identification in cattle farms, *Proc. of the 22nd Signal Processing and Communications Applications Conf.*, pp. 1347–1350, 2014.

[9] T. T. Zhi, C. N. Phyo, P. Tin, H. Hama, and I. Kobayashi, Image technology based cow identification system using deep learning, *Proc. of the Int. MultiConference of Engineers and Computer Scientists*, Vol. 1, pp. 1–4, 2018.

[10] A. I. Awad, From classical methods to animal biometrics: A review on cattle identification and tracking, *Computers and Electronics in Agriculture*, Vol. 123, pp. 423–435, 2016.

[11] W. Kusakunniran, A. Wiratsudakul, , U. Chuachan, S. Kanchanapreechakorn, and T. Imaromkul, Automatic cattle identification based on fusion of texture features extracted from muzzle images, *Proc. of IEEE Int. Conf. on Industrial Tech.*, pp. 1484–1489, 2018.

[12] F. Okura, S. Ikuma, Y. Makihara, D. Muramatsu, K. Nakada, and Y. Yagi, RGB-D video-based individual identification of dairy cows using gait and texture analyses, *Computers and Electronics in Agriculture*, Vol. 165, pp. 1–9, 2019.

[13] K. Wurtz, I. Camerlink, R. B. DâĂŹEath, A. Peña Fernández, T. Norton, J. Steibel, and J. Siegford, Recording behaviour of indoor-housed farm animals automatically using machine vision technology: A systematic review, *PloS one*, Vol. 14, No. 12, pp. 1–35, 2019.

[14] Y. Qiao, C. Cappelle, Y. Ruichek, and T. Yang, ConvNet and LSH-Based visual localization using localized sequence matching, *Sensors*, Vol. 19, No. 11, pp. 24–39, 2019.

[15] Y. Qiao, M. Truman, S. Sukkarieh, Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming, *Computers and Electronics in Agriculture*, Vol. 165, pp. 1–9, 2019.

[16] S. Kumar, A. Pandey, , K. S. R. Satwik, S. Kumar, S. K. Singh, A. K. Singh, and A. Mohan, Deep learning framework for recognition of cattle using muzzle point image pattern, *Measurement*, Vol. 116, pp. 1–17, 2018.

[17] W. Andrew, C. Greatwood, and T. Burghardt, Visual localisation and individual identification of holstein friesian cattle via deep learning, *Proc. of IEEE Int. Conference on Computer Vision*, pp. 22–29, 2017.

[18] W. Shen, H. Hu, B. Dai, X. Wei, J. Sun, L. Jiang, and Y. Sun, Individual identification of dairy cows based on convolutional neural networks, *Multimedia Tools and Applications* , pp. 1–14, 2019.

[19] Y. Liao, S. Kodagoda, Y. Wang, L. Shi, and Y. Liu, Place classification with a graph regularized deep neural network, *IEEE Trans. on Cognitive and Developmental Sys.*, Vol. 9, No. 4, pp. 304–315, 2016.

[20] J. Xie, B. Chen, X. Gu, F. Liang, and X. Xu, Self-attention-based BiLSTM model for short text fine-grained sentiment classification, *IEEE Access*, Vol. 7, pp. 180558–180570, 2019.

[21] F. Karim, S. Majumdar, H. Darabi, and S. Harford, Multivariate lstmfcns for time series classification, *Neural Networks*, Vol. 116, pp. 237–245, 2019.

[22] T. Gaber, A. Tharwat, A. E. Hassanien, and V. Snasel, Biometric cattle identification approach based on weberâĂŹs local descriptor and adaboost classifier, *Computers and Electronics in Agriculture*, Vol. 122, pp. 55–66, 2016.

[23] C. Cai and J. Li, Cattle face recognition using local binary pattern descriptor, *Proc. of Asia-Pacific Signal and Information Processing Association Annual Summit and Conf.*, pp. 1–4, 2013.

[24] S. Kumar, S. K. Singh, R. Singh, and A. K. Singh, Recognition of cattle using face images, *Animal Biometrics Techniques and Applications*, pp. 79–110, Springer, 2017.

[25] K. Zhao and D. He, Recognition of individual dairy cattle based on convolutional neural networks, *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 31, No. 5, pp. 181–187, 2015.

[26] C. Bahlo, P. Dahlhaus, H. Thompson, and M. Trotter, The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review, *Computers and Electronics in Agriculture*, Vol. 156, pp. 459–466, 2019.

[27] F. Ordóñez and D. Roggen, Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition, *Sensors*, Vol. 16, No. 1, pp. 1–15, 2016.

[28] T. Van Hertem, S. Viazzi, M. Steensels, E. Maltz, A. Antler, V. Alchanatis, A. A. Schlageter-Tello, K. Lokhorst, E. C. Romanini, C. Bahr, D. Berckmans, and I. Halachmi, Automatic lameness detection based on consecutive 3D-video recordings, *Biosystems Engineering*, Vol. 119, pp. 108–116, 2014.

[29] M. McPhee, B. Walmsley, B. Skinner, B. Littler, J. Siddell, L. Cafe, J. Wilkins, V. Oddy, and A. Alempijevic, Live animal assessments of rump fat and muscle score in angus cows and steers using 3-dimensional imaging, *Journal of Animal Science*, Vol. 95, No. 4, pp. 1847–1857, 2017.

[30] Y. Peng, N. Kondo, T. Fujiura, T. Suzuki, H. Yoshioka, and E. Itoyama, Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units, *Computers and Electronics in Agriculture*, Vol. 157, pp. 247–253, 2019.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, Rethinking the inception architecture for computer vision, *Proc. of IEEE CVPR*, pp. 2818–2826, 2016.

[32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, ImageNet: A large-scale hierarchical image database, *Proc. of IEEE CVPR*, pp. 248–255, 2009.

[33] K. Itakura, Y. Saito, T. Suzuki, N. Kondo, and F. Hosoi, Classification of soymilk and tofu with diffuse reflection light using a deep learning technique, *AgriEngineering*, Vol. 1, No. 2, pp. 235–245, 2019.

[34] F. Chollet and others, Keras: The python deep learning library, *Astrophysics Source Code Library*, 2018.