

Individual Cattle Identification Using a Deep Learning Based Framework[★]

Yongliang Qiao^{*} Daobilige Su^{*} He Kong^{*} Salah Sukkarieh^{*}
Sabrina Lomax^{**} Cameron Clark^{**}

^{*} *Australian Centre for Field Robotics (ACFR), Faculty of Engineering, The University of Sydney, NSW 2006, Australian.*
e-mail: {yongliang.qiao, daobilige.su, he.kong, salah.sukkarieh}@sydney.edu.au.

^{**} *Livestock Production and Welfare Group, School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, NSW 2006, Australia.*
e-mail: {sabrina.lomax, cameron.clark}@sydney.edu.au

Abstract: Individual cattle identification is required for precision livestock farming. Current methods for individual cattle identification requires either visual, or unique radio frequency, ear tags. We propose a deep learning based framework to identify beef cattle using image sequences unifying the advantages of both CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory) network methods. A CNN network was used (Inception-V3) to extract features from a rear-view cattle video dataset and these extracted features were then used to train an LSTM model to capture temporal information and identify each individual animal. A total of 516 rear-view videos of 41 cattle at three time points separated by one month were collected. Our method achieved an accuracy of 88% and 91% for 15-frame and 20-frame video length, respectively. Our approach outperformed the framework that only uses CNN (identification accuracy 57%). Our framework will now be further improved using additional data before integrating the system into on-farm management processes.

© 2019, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Cattle identification, deep learning, LSTM, CNN, precision livestock farming.

1. INTRODUCTION

Cattle identification, which refers to the process of accurately recognizing individual cattle, plays an important role in automatic behavior analysis, weighing, health monitoring and welfare evaluation in precision livestock farming (Banhazi et al., 2012; Berckmans, 2014; McCabe et al., 2019). Once an animal is identified the growth, body weight can be tracked across time to achieve desirable outputs (He et al., 2016; Halachmi and Guarino, 2016). In our recent work (Qiao et al., 2019b), we have proposed a cattle segmentation and contour extraction approach based on Mask R-CNN in a complex background with 0.92 mean average precision. The proposed method achieved contour extraction with an average distance error of 33.56 pixel on an enhanced image dataset. However, the proposed full automatic precision cattle management system, which comprises of components such as the former cattle segmentation and contour extraction method, requires on accurate cattle identification.

The first animal identification method was proposed by Ploegaert (1976), where an automatic identification system based on the Pulse Code Modulation technique was

tested on a practical farm. After that, various animal identification approaches were developed (Andrew et al., 2017; Awad, 2016). Popular cattle identification methods such as ear-tag, collar, and Radio Frequency Identification (RFID) technologies need to put certain sensing devices on the body (Shen et al., 2019). In addition, when monitoring a large number of livestock animals in harsh outdoor environments, devices such as ear-tags are prone to loss or duplication, while collar and RFID devices are not only expensive but also easy to be damaged (Li et al., 2017).

Recently, there has been increasing interest in visual biometrics-based cattle identification as cattle have their own external biometrics characteristics (e.g. body contour and coat pattern) which are unique for each individual animal (Andrew et al., 2016; Zhao and He, 2015; Arslan et al., 2014). Since cattle videos are comprised of a large number of sequential images, they contain both spatial and temporal information such as cattle's physiological and behavioral characteristics (e.g. kinematic gait parameters). This information can be useful for cattle identification (Frost et al., 1997). However, one has to find efficient ways to extract these useful information first.

2. RELATED WORKS

In existing visual approaches, four main cattle body areas are often used for identification: muzzle, face, back and trunk. Gaber et al. (2016) used the Weber Local Descriptor

[★] 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.

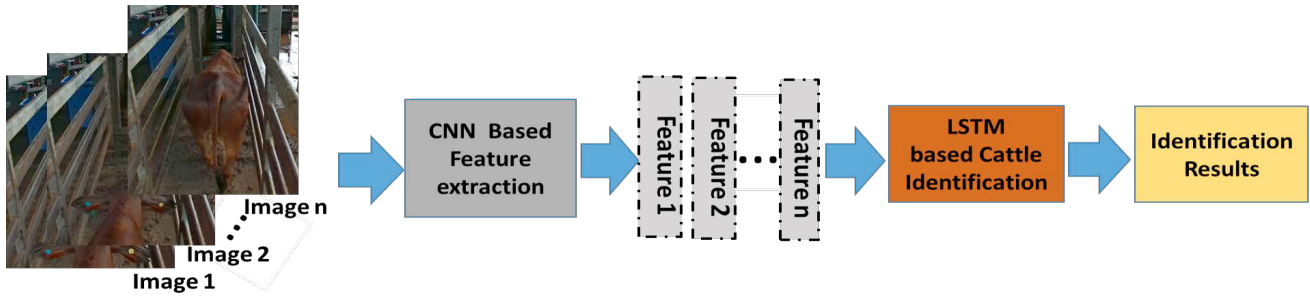


Fig. 1. Proposed CNN and LSTM based beef cattle identification scheme

to extract robust features from cattle muzzle print images for cattle identification. Cai and Li (2013) presented a facial representation model of cattle based on local binary pattern texture features. The work in Kumar et al. (2017) proposed cattle identification method based on extracted cattle facial features.

Recently, deep learning approaches with powerful feature extraction and image representation abilities have been widely used in the fields of visual classification and recognition (Tompson et al., 2014; Shen et al., 2019). After detecting trunks from raw images, Zhao and He (2015) proposed a CNN network method for cow identification. Kumar et al. (2018) proposed a CNN based approach for identification of individual cattle by using primary muzzle point image pattern. Zin et al. (2018) trained a CNN based on back images of cows to identify individual cows. This approach ignored the information of the head and legs which also include useful identification information such as contour and texture features. In Shen et al. (2019), a CNN network was fine-tuned with side-view images of dairy cows and used to perform cow identification. However, previous image frame based approaches have not utilized behaviour such as walking posture or gait information.

Time-series data such as video and inertial measurement units (IMU) measurements which contain more temporal information have been more widely used to monitor cattle behaviours and welfare (Bahlo et al., 2019; Ordóñez and Roggen, 2016). Van Hertem et al. (2014) developed an automated lameness scoring algorithm based on 3D-video recordings of cow gait. McPhee et al. (2017) developed a learning-based approach for assessing traits such as rump fat and muscle score. Their results demonstrate that using curvature feature to represent a cow's body shape is useful in the evaluation of Angus cows and steers. Peng et al. (2019) developed a recurrent neural network (RNN) with an LSTM model to monitor and classify cattle behavior patterns using IMU measurements. Andrew et al. (2017) demonstrated a video processing pipeline for cattle identification and adopted a Long-term Recurrent Convolutional Network to classify the cattle videos taken by the unmanned aerial vehicles. Their tests suggest that video data and standard deep learning components can be used together to enhance identification performance of Friesian dairy cattle with uniform coats.

Despite the above progress in vision and deep learning based cattle identification, most existing methods focus on dairy cattle. There is limited work targeting beef cattle,

potentially due to the difficulty of this task as most beef cattle have uniform coat colours.

3. PROPOSED APPROACH

Here we propose a deep learning based framework for beef cattle identification using video data (Fig. 1). The method leverages the strengths of both CNNs and LSTM networks, which have been shown to be efficient in extracting spatial information (Liao et al., 2016; Qiao et al., 2019a), and modelling the hidden patterns or features in time-space data (Chen et al., 2017; Karim et al., 2019), respectively. More specifically, at each time step (image frame), a set of CNN features, which describe both the visual content and the motion information of cattle, are firstly extracted from video data. Based on the extracted features, the LSTM model will be trained to classify each cattle video sequence considering the temporal evolution of the features for each time step. As LSTM processes one image frame at a time, it can model sequence of elements that are interdependent. We will use both CNN features and temporal information to improve cattle identification accuracy in a feedlot environment.

The outline of the proposed approach is illustrated in Fig. 1. As illustrated in Fig.1, the proposed approach consists of two main steps: 1) CNN based feature extraction; 2) LSTM based cattle identification. The aim is to identify cattle videos that are represented by a sequence of CNN features (one feature per image) corresponding to different cattle.

3.1 CNN based Feature Extraction

The CNN model Inception-V3 is often used for image recognition (Szegedy et al., 2016) and is made up of building blocks including convolutions, average pooling, max pooling, contacts, dropouts, and fully connected layers.

For our work, a given image L at time T and ConvNet learned parameter w , the CNN feature was extracted using:

$$X(t) = f_n(\dots f_2(f_1(l; w_1); w_2) \dots), w_n) \quad (1)$$

where f_1, \dots, f_n are the corresponding layer functions. Each layer output is a kind of CNN features for image representation. In general, the low layers retained high spatial resolution whilst the high layers contained 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. Inception-V3 model with the pre-trained weights on the ImageNet

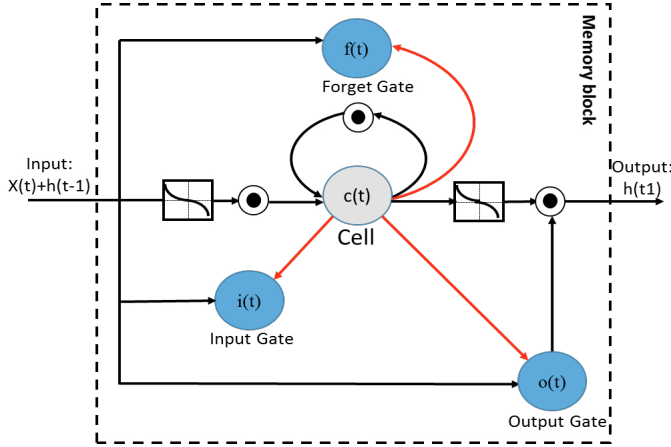


Fig. 2. LSTM cell

dataset (Deng et al., 2009) were used to extract cattle features. Thus, each video image had a 2048-dimensional feature before being passed to the LSTM model.

3.2 Cattle Identification using LSTM

LSTM is a popular network for space-time data processing with strong abilities to learn and remember over long sequences of input data (Karim et al., 2019). It extends RNN (Recurrent Neural Network) with memory cells. These cells usually have few linear interactions making the information maintaining process easier (Ordóñez and Roggen, 2016). 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) (Itakura et al., 2019).

To make the most of the video contents, one should consider the visual aspects and also characterize the object appearances as well as the motion present within the data. As such, after CNN features were extracted image by image from each video, we trained the LSTM model to identify the cattle video sequences.

As illustrated in Fig. 2, a common LSTM unit consist 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 was updated as follows:

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

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

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

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

$$h_t = o_t \tanh(c_t) \quad (6)$$

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.

For cattle identification, LSTM, with the extracted CNN features as its input, learned and modelled each animal's unique temporal characteristics. At each time t , given a video frame length r , the input CNN features were

denoted as $[x_{t-l}, x_t, \dots, x_{t+r}]$. In our work, each frame x_t was a 2048-dimensional CNN feature (see Section 3.1), and the network had 1 LSTM layer with 2048 cells, a 1024 projection layer for dimensionality reduction, a dropout layer (dropout rate is 0.5), and a final softmax output layer. The output with the maximum value (class confidences, the value is between 0 and 1) was used as the cattle identification result. If it matched the ground truth class label, then it was a true result.

4. EXPERIMENTAL SETUP

4.1 Data Acquisition

Our data was collected at an Australian Country Choice feedlot. The data acquisition system, as shown in Fig. 3, was placed nearby the cattle crush during data collection. Two stereo ZED cameras (with 110° horizontal field of view), a GPU-equipped embedded PC (Neosys Nuvo 6108GC) and a high-speed volume data storage disk were used. In our experiment, only 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 influence of motion blur during the herding process from the open-air pen to the restraint device (crush).

The data was collected in 2018 at three different times (induction, middle and end point) on 20 March, 30 April and 30 May, respectively. The cattle were moved to yards from the open-air pens to determine live weight. Cattle live weight varied from 330 kg to 550 kg as body size and mass changed in these three months. The standard deviation of weight in the training dataset was 45 kg. Data was acquired when the cattle were walking along the race (path) from right to left in Fig. 3.



Fig. 3. Area for data acquisition, the race leading to the crush at Brisbane Valley feedlot showing rear view camera clamped to overhead bar and the side view camera on tripod.

Given the 1920×1080 of the original images, in order to improve the system efficiency, we extracted and saved the central part of the original image (Region of Interest, ROI). As illustrated in Fig. 4, a size of 401×506 image was obtained after image ROI extraction. In our experiment, a total of 516 cattle videos (10320 image frames) from 41 cattle were used, with each video containing 20-frame

long spatio-temporal streams and each frame size was 401×506 . For training, the actual identities of cattle were manually recorded according to their ear tags.

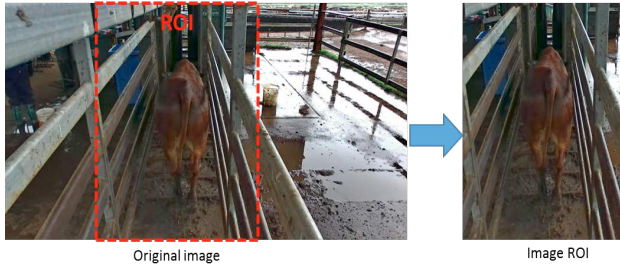


Fig. 4. Image ROI extraction. As the camera position is fixed, a red box with fixed size is used to extract the ROI. Size of extracted cattle image ROI is 401×506 , which makes feature extraction more efficient.

4.2 Proposed Network Training

In our work, Keras (Chollet et al., 2018) with GPU was used to construct the cattle identification model. Details of hardware information in this experiment are provided in Table 1.

Table 1. The experimental hardware

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

In the proposed CNN and LSTM 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. Then all the CNN features for each frame in the videos were used to train the LSTM model. In our experiments, 439 videos of 41 cattle were used as the training dataset while the remaining 77 videos of the same 41 cattle were used as the testing dataset. For LSTM network training, initial learning rate is 0.00005, the learning decay factor was 10^{-6} , and loss function was “categorical crossentropy”.

5. EXPERIMENTAL RESULTS

When the Inception-V3 method (Szegedy et al., 2016) was used alone, a total of 8780 cattle images (from the 439 training videos) were used to fine-tune the network, and 1540 images (from the 77 testing videos) were used as the testing data. Moreover, in the Inception-V3 method, we froze the parameters of the first 172 layers and only retrained the last two inception blocks. The number of output nodes in the last layer was equal to the number of cattle (41 in our data). The duration of the training process was 2 hours with a 0.0005 learning rate.

In order to study the influence of video length for the cattle identification, different video lengths (i.e. 5, 10, 15 and 20 frame length) with their respective accuracies and processing time are presented in Table 2. Using videos of longer length, LSTM can capture more temporal

information and improve accuracy in cattle identification. Cattle identification accuracy of Inception-V3 was lower than that of the CNN and LSTM based approach.

Table 2. Accuracy of cattle identification

Methods	Accuracy (%)	Time (s)
Inception-V3 (single image)	57	10.5
CNN and LSTM (5 frame)	78	11.2
CNN and LSTM (10 frame)	87	11.3
CNN and LSTM (15 frame)	88	12.5
CNN and LSTM (20 frame)	91	12.7



Predict ID: 13946 Ground truth: 13946



Predict ID: 13905 Ground truth: 13905

(a) True identification examples



Predict ID: 13835 Ground truth: 13929



Predict ID: 13946 Ground truth: 13833

(b) False identification examples

Fig. 5. Some examples of CNN and LSTM based cattle identification results. The video length is 20 frames. Due to space limitation, for each video, only first two and the last two frames are displayed. Note also that some cattle were walking back and forth during data acquisition.

Fig. 5 illustrates true and false cattle identification examples of the proposed CNN and LSTM based approach. Overall performance of CNN and LSTM based beef cattle identification was favorable except for few false cases in which cattle were partly covered by mud. We also noted that, in some videos, cattle were standing static or have made little movement. For these cases, the cattle identification accuracy was not as high due to the lack of enough temporal information. In general, with more training data and longer video length, the accuracy of cattle identification can be further improved.

In summary, the experiment results illustrate that the proposed method can extract and learn the extra information (temporal) relevant to individual identification from video data. This is mainly due to the fact that LSTM can learn useful temporal information such as the gait or walking behavior of cattle, which further enhance visual cattle identification performance.

6. CONCLUSION AND FUTURE WORK

Identifying individual cattle is of great significance in computer vision based precision livestock management. In this paper, we proposed a cattle identification method by combining CNN and LSTM using video data. The Inception-V3 CNN network is used to extract visual features. Then the LSTM model is trained to identify individual cattle based on the extracted CNN features. To validate the proposed method, we performed extensive experiments on a dataset of 41 different cattle.

The major contributions of this paper are: (1) a deep learning approach composed of CNN and LSTM networks has been developed, capable of automatically learning feature representations and modeling the temporal dependencies between their activation; (2) a systematic assessment of the proposed framework has been completed with performance validated using different video lengths; experiment results show our method to achieve an accuracy of 88% and 91% for 15-frame and 20-frame video length, respectively; (3) we have compared the proposed method with the CNN approach, and the results show that the proposed approach outperforms the framework that only uses the CNN based method whose identification accuracy is 57%.

For future work, we plan to verify the proposed approach on a larger and more-complex video dataset. The LSTM based approach will also be used to estimate cattle body condition score and live weight.

ACKNOWLEDGEMENTS

The authors express their gratitude to Australian Country Choice and the Brisbane Valley farm staff for their help with data collection. Also particular thanks are due to Matthew Truman, Amanda Doughty, Ashraf Islam and Mike Reynolds for their involvement and efforts in the whole experiment organization and cattle information collection.

REFERENCES

- Andrew, W., Greatwood, C., and Burghardt, T. (2017). Visual localisation and individual identification of holstein friesian cattle via deep learning. In *Proc. IEEE International Conference on Computer Vision (ICCV), Venice, Italy*, 22–29.
- Andrew, W., Hannuna, S., Campbell, N., and Burghardt, T. (2016). Automatic individual holstein friesian cattle identification via selective local coat pattern matching in rgb-d imagery. In *Image Processing (ICIP), 2016 IEEE International Conference on*, 484–488. IEEE.
- Arslan, A.C., Akar, M., and Alagoz, F. (2014). 3D cow identification in cattle farms. In *Signal Processing and Communications Applications Conference (SIU), 2014 22nd*, 1347–1350. IEEE.
- Awad, A.I. (2016). From classical methods to animal biometrics: A review on cattle identification and tracking. *Computers and Electronics in Agriculture*, 123, 423–435.
- Bahlo, C., Dahlhaus, P., Thompson, H., and Trotter, M. (2019). The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review. *Computers and Electronics in Agriculture*, 156, 459–466.
- Banhazi, T.M., Lehr, H., Black, J., Crabtree, H., Schofield, P., Tschärke, M., and Berckmans, D. (2012). Precision livestock farming: an international review of scientific and commercial aspects. *International Journal of Agricultural and Biological Engineering*, 5(3), 1–9.
- Berckmans, D. (2014). Precision livestock farming technologies for welfare management in intensive livestock systems. *Scientific and Technical Review of the Office International des Epizooties*, 33(1), 189–196.
- Cai, C. and Li, J. (2013). Cattle face recognition using local binary pattern descriptor. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, 1–4. IEEE.
- Chen, Z., Jacobson, A., Sünderhauf, N., Upcroft, B., Liu, L., Shen, C., Reid, I., and Milford, M. (2017). Deep learning features at scale for visual place recognition. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, 3223–3230. IEEE.
- Chollet, F. et al. (2018). Keras: The python deep learning library. *Astrophysics Source Code Library*.
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. Ieee.
- Frost, A., Schofield, C., Beaulah, S., Mottram, T., Lines, J., and Wathes, C. (1997). A review of livestock monitoring and the need for integrated systems. *Computers and Electronics in Agriculture*, 17(2), 139–159.
- Gaber, T., Tharwat, A., Hassanien, A.E., and Snasel, V. (2016). Biometric cattle identification approach based on webers local descriptor and adaboost classifier. *Computers and Electronics in Agriculture*, 122, 55–66.
- Halachmi, I. and Guarino, M. (2016). Precision livestock farming: a per animal approach using advanced monitoring technologies. *Animal*, 10(9), 1482–1483.
- He, D., Liu, D., and Zhao, K. (2016). Review of perceiving animal information and behavior in precision livestock farming. *Transactions of the Chinese Society for Agricultural Machinery*, 47(5), 231–244.
- Itakura, K., Saito, Y., Suzuki, T., Kondo, N., and Hosoi, F. (2019). Classification of soymilk and tofu with diffuse reflection light using a deep learning technique. *AgriEngineering*, 1(2), 235–245.
- Karim, F., Majumdar, S., Darabi, H., and Harford, S. (2019). Multivariate lstm-fcns for time series classification. *Neural Networks*, 116, 237–245.
- Kumar, S., Pandey, A., Satwik, K.S.R., Kumar, S., Singh, S.K., Singh, A.K., and Mohan, A. (2018). Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement*, 116, 1–17.
- Kumar, S., Singh, S.K., Singh, R., and Singh, A.K. (2017). Recognition of cattle using face images. In *Animal Biometrics*, 79–110. Springer.
- Li, W., Ji, Z., Wang, L., Sun, C., and Yang, X. (2017). Automatic individual identification of holstein dairy cows using tailhead images. *Computers and electronics in agriculture*, 142, 622–631.
- Liao, Y., Kodagoda, S., Wang, Y., Shi, L., and Liu, Y. (2016). Place classification with a graph regularized deep neural network. *IEEE Transactions on Cognitive and Developmental Systems*, 9(4), 304–315.
- McCabe, B.K., Tait, S., Marchuk, S., Lee, S., Antille, D.L., Eberhard, J., and Baillie, C.P. (2019). Identification

- and analysis of organic waste produced from australian intensive livestock industries: opportunities for aggregation and treatment via co-digestion. In *2019 ASABE Annual International Meeting*, 1. American Society of Agricultural and Biological Engineers.
- McPhee, M., Walmsley, B., Skinner, B., Littler, B., Siddell, J., Cafe, L., Wilkins, J., Oddy, V., and Alempijevic, A. (2017). Live animal assessments of rump fat and muscle score in angus cows and steers using 3-dimensional imaging. *Journal of Animal Science*, 95(4), 1847–1857.
- Ordóñez, F. and Roggen, D. (2016). Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1), 115.
- Peng, Y., Kondo, N., Fujiura, T., Suzuki, T., Yoshioka, H., Itoyama, E., et al. (2019). 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*, 157, 247–253.
- Ploegaert, P. (1976). Technical aspects of cow identification in combination with milk yield recording and concentrate feeding in and outside the milking parlour. In *Proceedings of the Symposium on Cow Identification and Their Applications*. Wageningen.
- Qiao, Y., Cappelle, C., Ruichek, Y., and Yang, T. (2019a). Convnet and LSH-based visual localization using localized sequence matching. *Sensors*, 19(11), 2439.
- Qiao, Y., Truman, M., and Sukkarieh, S. (2019b). Cattle segmentation and contour extraction based on mask r-cnn for precision livestock farming. *Computers and Electronics in Agriculture*, 165, 104958.
- Shen, W., Hu, H., Dai, B., Wei, X., Sun, J., Jiang, L., and Sun, Y. (2019). Individual identification of dairy cows based on convolutional neural networks. *Multimedia Tools and Applications*, 1–14.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2818–2826.
- Tompson, J.J., Jain, A., LeCun, Y., and Bregler, C. (2014). Joint training of a convolutional network and a graphical model for human pose estimation. In *Advances in Neural Information Processing Systems*, 1799–1807.
- Van Hertem, T., Viazzi, S., Steensels, M., Maltz, E., Antler, A., Alchanatis, V., Schlageter-Tello, A.A., Lokhorst, K., Romanini, E.C., Bahr, C., et al. (2014). Automatic lameness detection based on consecutive 3d-video recordings. *Biosystems Engineering*, 119, 108–116.
- Zhao, K. and 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., and Kobayashi, I. (2018). Image technology based cow identification system using deep learning. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1.