Visual Re-identification within Large Herds of Holstein Friesian Cattle

Mark Lennox
CattleEye Ltd
Northern Ireland Science Park, Belfast
BT3 9DT
Email: mark.lennox@cattleeye.com

Niall McLaughlin
EEECS, QUB
University Rd, Belfast
BT7 1NN
Email: n.mclaughlin@qub.ac.uk

Jesus Martinez del Rincon
EEECS, QUB
University Rd, Belfast
BT7 1NN
Email: j.martinez-del-rincon@qub.ac.uk

***Abstract*—Most modern farms use radio frequency identification (RFID) devices to scan tags/collars to identify unique individual cows. However, keeping such devices operational can be expensive and time-consuming as each animal must have a tag or collar attached. Furthermore, when these systems are functional, they will often miss a percentage of the herd during milking because the cows are out of scanner range. Therefore, farmers require a new solution to identifying their animals that will improve upon the performance of a standard RFID machine while also requiring less maintenance for the farmer. This paper seeks to investigate a deep learning alternative to this problem. By installing stationary top-down cameras in high traffic areas within the farm, we aim to test the performance of our identification solution when compared to the current standard RFID machines. The presented deep learning solution was trained on a dataset of over 200,000+ instances based on 3054 unique individuals across two different cameras. In addition, we test our approach on three additional farms, with 307, 1198 and 2643 unique cows, respectfully. Our key contributions include testing the limits of standard reID solutions within this space and producing an approach capable of a rank-1 accuracy of >86%. In closing, we identify the shortcomings of these techniques and outline future work toward using deep learning to replace the current RFID systems.**

I. Introduction

Intelligent systems are increasingly being deployed on farms and in the wild to monitor and improve animal welfare [1, 2, 3]. In recent years, these systems have capitalised on the advancement in deep learning to automatically detect and identify objects (i.e. typically humans or animals). However, while previously existing radio frequency identification (RFID) [4] plays a vital role in modern herd management concerning metrics based on milking, locomotion, feeding and resting, they are also time-consuming, infrequent and expensive due to the cost of deploying and maintaining individual sensors. On the contrary, video-based solutions are easy to scale and require little human intervention.

This paper uses Deep Convolutional Neural Networks (CNNs) to develop and deploy a practical and cost-effective herd monitoring system to recognise and distinguish the different cow individuals. Rather than follow a traditional classification/identification setting, which requires updating the model for any new individual and that losses performance when scaling up to large herds, we based our solution on a reidentification approach, which can generalise to unique

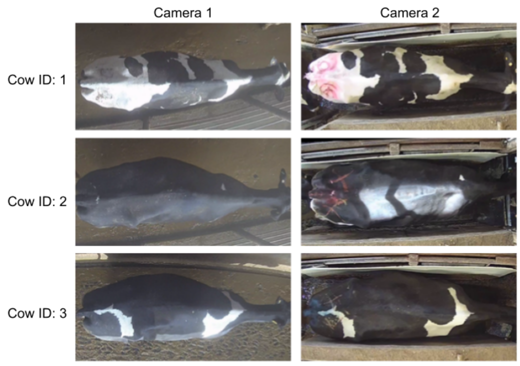


Fig. 1: **Cattle Re-Identification Training Examples:** Cows are captured at two different points on the same farm. The discrepancy between images to the right and the left for the same identity reflects the difficulty of the problem, with very different illumination conditions, level of soiled cows and presence of cows without any distinctive marks.

individuals and therefore be directly deployed in other farms without retraining or tuning to the new herd population and with minimal effort by the farmer.

The main contributions of this paper are: to the best of our knowledge, (1) we present the first reidentification system for large herds in real live conditions. (2) We investigate the most up-to-date reID loss function in human reidentification and their validity for cow monitoring. (3) We validate our solution in a massive cow population, testing the limits of a reID system and the generalisation properties in a cross-farm setting. Finally, (4) We achieve a rank one accuracy >86%, comparable to RFID solutions with a fraction of their cost.

II. Related Work

This section explores the recent scarce approaches that apply deep learning within this space. Andrew et al. proposed a complete deep learning Holstein Friesian cattle detection and identification system [5], which was one of the first examples of automated visual bovine identification. They tested their approach on footage taken from both fixed cameras within the farm and Unmanned Aerial Vehicles (UAVs) in the fields. Their in-barn footage contained 940 frames for

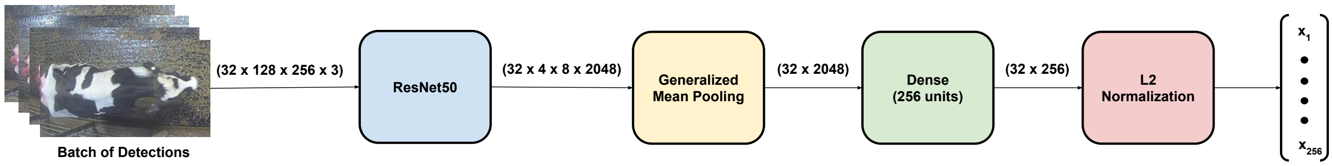


Fig. 2: **ResNet-50 cattle re-identification architecture:** A batch of RGB detections is collected from a pre-trained detector and resized to (128x256x3) during training. The ResNet backbone then encodes the detections before being summarised by the generalised mean pooling layer. A dense layer is then used to reduce the dimensionality to 256 features, which are finally normalised by an L2-normalisation layer and used to compare against other samples using a cosine distance.

only 89 individual cows and 46,430 frames originating from 34 clips of UAV footage taken from 23 individuals and obtained promising results with 86.1% in-barn accuracy and 98.1% UAV accuracy. Furthermore, Andrew et al. suggested that their marker-less Friesian cattle identification system was good enough to assist existing tagging methods [5]. However, they only tested their approach on a minimal dataset. This investigation aims to take it a step further by testing our approaches on separate herds from the training dataset with as many as 2643 unique IDs.

Qiao et al . also explored using deep learning in precision livestock farming to identify individual cattle [6]. They proposed a deep learning method combining a CNN and an LSTM (Long Short-Term Memory) network. The CNN would first encode each detection of a cow; the features generated would then be passed to an LSTM, which then analysed the temporal information within consecutive detections of the same animal. Qiao et al . benchmarked their approach on a dataset gathered from 516 videos based on 41 cows taken from a rear-view camera, concluding that increasing the number of frames improved the network's overall performance. Qiao et al. later improved on their original approach by using a bi-directional LSTM [7], with marginal improvement.

In addition to the aforementioned two orders of magnitude larger benchmark, we will demonstrate that only ten frames of each cow are needed to form a successful final representation for reID. This means we can include as many cows in our monitoring system as possible without bias towards cows with more detections. This is relevant, as in a real-world scenario, we can't ensure that an animal will take the same time to pass under the camera each time.

III. Dataset

The reidentification dataset used to train and validate the performance of our approach is outlined in Table I. This training dataset contains ground truth instances from one farm, which spans two cameras and includes 3054 unique identities across six days of footage. Each image was selected and cropped around the cow using a pre-trained detector and resized (128x256x3). The cameras provide a top-down view of the entire animal, as seen in Figure 1. The training and validation datasets used in this investigation far surpass all previous approaches in applying deep learning to identify cows. We selected three additional farms with previously installed RFID machines to test our new system. Testing farms only require one camera, with reidentification applied across two different days. Images captured on the first day will be considered the enrolment reference (gallery) that needs to be matched with the detection on the second operative day (probes). Our farm selection also contains a variety of herd sizes, beginning with a smaller herd (307 cows) and then eventually scaling up to a larger herd (2643 cows). Testing various farms from the training one (i.e. cross-farm setting) will ensure that our approach has not over-fitted to the original training farm while also testing its performance on a larger scale.

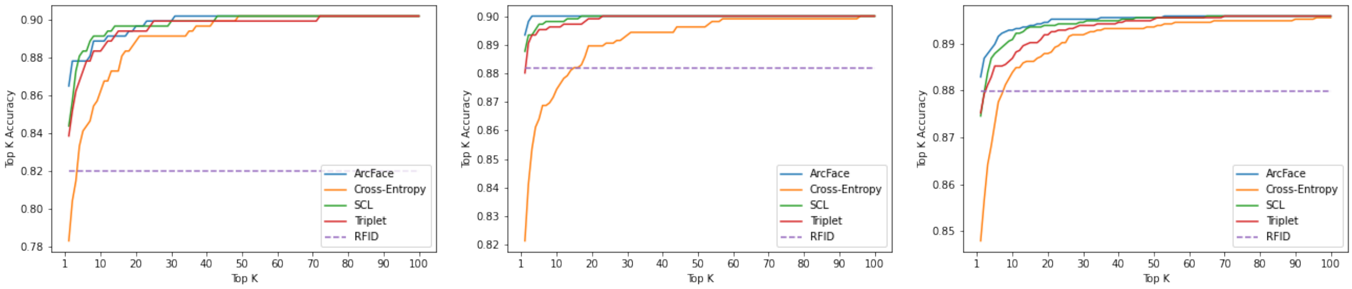
TABLE I: Cattle Re-Identification Dataset.

Farm	Dataset	Camera	No. of Unique IDs	Instances
A	Train	1	3054	97606
A	Train	2	3054	104685
A	Validation	1	3054	6108
A	Validation	2	3054	6108
B	Test	1	307	6140
C	Test	1	1198	23960
D	Test	1	2643	52860

IV. Methods

Typically RFID scanners are placed in high-traffic areas within the farm to ensure that the machine can identify as many animals as possible during the milking. We have installed stationary cameras in similar locations on each farm. The footage captured spans several milkings at 1080p at ten fps. A pre-trained YOLOv4 [8] model was used first to detect and crop any cows that passed under our camera before a SMOT-based tracking algorithm [9] was used to determine the complete tracklets. Finally, the tracklets were summarised to select the ten best detections based on the confidence scores of the object detection algorithm. Each set of detections is then split randomly, with eight images being used in the training dataset and two images being placed in the validation dataset. We will use all ten detections for each cow in the test dataset across two days.

The previous short tacklets containing the detections are passed to our reidentification system. Figure 2 outlines the main backbone of our reidentifier. This backbone generates a short 256-element feature vector for each detection that can be



(a) Farm B (b) Farm C (c) Farm D

Fig. 3: **CMC curves:** A comparison of the four training methods across three test farms.

used as a unique signature. By comparing the detected cow feature vector (probe) against all existing cows in the herd (gallery), the animal can be recognised by applying a cosine distance between the averaged feature vectors of the tracklet, with the chosen identity being the one with the smallest distance. Finally, a threshold with the minimum acceptable similarity can be introduced to deal with open-world situations where new cows appear on the farm but cannot be reidentified since they have not yet enrolled or are simply distracters. A deep metric learning method is used to train our encoder for this approach to work. Three state-of-art reID losses were considered triplet-loss [10], supervised contrastive learning [11] and ArcFace [12].

TABLE II: Cattle Re-Identification Results.

Farm	System	Top One Accuracy	Top One F1 Score
B	ArcFace [12]	**0.865**	**0.927**
B	SCL [11]	0.843	0.915
B	Triplet [10]	0.838	0.912
B	Cross-Entropy	0.783	0.878
B	RFID	0.82	0.901
C	ArcFace [12]	**0.893**	**0.942**
C	SCL [11]	0.887	0.939
C	Triplet [10]	0.88	0.939
C	Cross-Entropy	0.821	0.897
C	RFID	0.881	0.939
D	ArcFace [12]	**0.882**	**0.937**
D	SCL [11]	0.875	0.933
D	Triplet [10]	0.875	0.933
D	Cross-Entropy	0.847	0.917
D	RFID	0.879	0.936

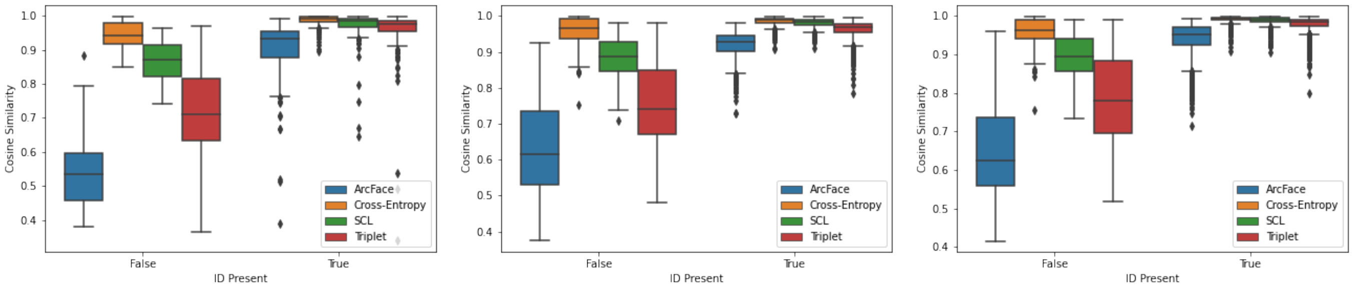
V. Results

In this investigation, we explore the performance of our re-identification system. First, we train a simple classifier using categorical cross-entropy loss as an initial baseline to show the advantage of a reID training versus a conventional classification training. This baseline uses the same backbone as Figure 2 with the addition of one final softmax layer for training purposes using categorical cross-entropy. Next, to ensure each reID approach does not overfit, we validated the performance using a separate set of detections that included all cows used during training. In addition, we wanted to stress-test this set of deep learning solutions on unseen farms to avoid any bias that would be seen if we were to test on the same farm. Please note that no fine-tuning or retraining was done for the model trained in farm A to be then tested in farms B, C and D. Table II outlines the results for all three test farms, with ArcFace outperforming all other techniques on each farm. The CMC curves presented in Figure 3 further showcase that the ArcFace encoder produced the best rank accuracy across all three test farms, with the three reID approaches outperforming the baseline classification setting as well as getting comparable or better results than the RFID system.

As an extension of our evaluation of each training technique, in Figure 4, we present a set of boxplots that show the cosine similarity values between feature vectors for cows that appear on both days and for those that do not. Cows appearing on both days (present IDs in Table II) represent those that should be reidentified since they have been previously enrolled. In contrast, cows only appear once missing IDs in Table II represent an open-world setting of cows, i.e. that should not be matched to anyone. As such, present cows should have a higher similarity score by definition, as demonstrated in Table II. These plots let us determine if a natural threshold exists between cows reappearing from two consecutive days. Like most of the approaches tested, ArcFace had a favourable rank accuracy, and Figure 4 reveals that the encoder trained using ArcFace is better at identifying when a cow reappears each day and therefore makes it easier to determine a threshold for accepting the identification as explained in section III.

VI. Qualitative Discussion

As an extension of our evaluation of each training technique, in Figure 4, we present a set of boxplots that show the cosine similarity values between feature vectors for cows that appear on both days and for those that do not. Cows appearing on both days (present IDs in Table II) represent those that should be reidentified since they have been previously enrolled. In contrast, cows only appear once missing IDs in Table II represent an open-world setting of cows, i.e. that should not be matched to anyone. As such, present cows should have a higher similarity score by definition, as demonstrated in Table II. Furthermore, these plots let us determine if there is a natural threshold between cows that reappear from two consecutive



(a) Farm B (b) Farm C (c) Farm D

Fig. 4: **Evaluating Cosine Similarity Strength:** Box plots presenting the cosine similarity for predicted matches for cows that appear on both days (True) across the three test farms.

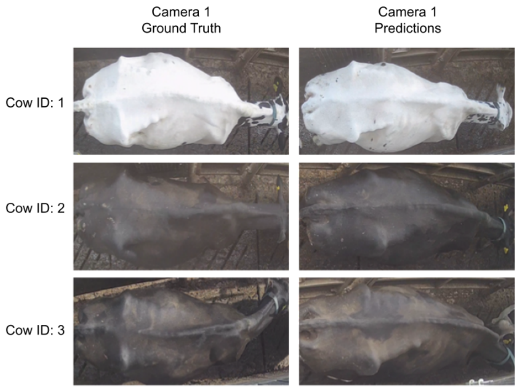


Fig. 5: **Misclassified Examples.** Three misclassified cows were taken from one of the test farms. In the left column, we have the ground truth thumbnails for each animal, and on the right column, we have the thumbnail for the predicted identity. The lack of a clear pattern on these cows is a theme in all the misclassified examples on all three farms.

days. Like most of the approaches tested, ArcFace had a favourable rank accuracy, and Figure 4 reveals that the encoder trained using ArcFace is better at identifying when a cow reappears each day and therefore makes it easier to determine a threshold for accepting the identification as explained in section III.

VII. Conclusion

This paper investigated the potential of using deep learning to replace an onsite RFID system to reidentify cows within large herds of Friesian cattle. We considered various deep metric learning methods and covered a range of herd sizes to test the robustness of each solution. Our results indicate a real potential for deep learning within this space, achieving strong results by averaging the final set of encodings for a collection of detections for a cow. In future work, there is still an open challenge in identifying cows with no distinct patterns (e.g. Figure 5). However, these misclassifications show a clear limitation of using a deep learning model that focuses predominantly on the visual appearance of the animal's full body. Therefore, a deep learning model must consider more than just the animal's colour and pattern to outperform an RFID system. In future work, we will explore additional techniques that will build on this work by identifying each animal through other means, such as using a keypoint model to identify important parts of the animal and then using their positional information to correct and scale each detection. We hope that the proof of concept provided in this paper can aid the research of applying deep learning to agriculture and propel the field forward to help ensure the well-being of the animals on the farm.

References

[1] Pang Wei Koh et al. "Wilds: A benchmark of in-the-wild distribution shifts". In: *International Conference on Machine Learning*. PMLR. 2021, pp. 5637–5664.

[2] Clemens-Alexander Brust et al. "Towards automated visual monitoring of individual gorillas in the wild". In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017, pp. 2820–2830.

[3] Matthias Körschens, Björn Barz, and Joachim Denzler. "Towards automatic identification of elephants in the wild". In: *arXiv preprint arXiv:1812.04418* (2018).

[4] Dan Grooms. "Radio frequency identification (RFID) technology for cattle". In: (2007).

[5] William Andrew, Colin Greatwood, and Tilo Burghardt. "Visual localisation and individual identification of holstein friesian cattle via deep learning". In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017, pp. 2850–2859.

[6] Yongliang Qiao et al. "Individual cattle identification using a deep learning based framework". In: *IFAC-PapersOnLine* 52.30 (2019), pp. 318–323.

[7] Yongliang Qiao et al. "BiLSTM-based individual cattle identification for automated precision livestock farming". In: *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*. IEEE. 2020, pp. 967–972.

[8] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection". In: *arXiv preprint arXiv:2004.10934* (2020).

[9] Wei Li et al. "Smot: Single-shot multi object tracking". In: *arXiv preprint arXiv:2010.16031* (2020).

[10] Florian Schroff, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 815–823.

[11] Prannay Khosla et al. "Supervised contrastive learning". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 18661–18673.

[12] Jiankang Deng et al. "Arcface: Additive angular margin loss for deep face recognition". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 4690–4699.