Towards Self-Supervision for Video Identification of Individual Holstein-Friesian Cattle: The *Cows2021* Dataset

Jing Gao[1] Tilo Burghardt[1] William Andrew[1,2] Andrew W. Dowsey[2] Neill W. Campbell[1]

[1]Department of Computer Science
University of Bristol
Bristol, United Kingdom

[2]Bristol Veterinary School
University of Bristol
Bristol, United Kingdom

Abstract

In this paper we publish the largest identity-annotated Holstein-Friesian cattle dataset (Cows2021) and a first self-supervision framework for video identification of individual animals. The dataset contains $10,402$ *RGB images with labels for localisation and identity as well as* 301 *videos from the same herd. The data shows top-down in-barn imagery, which captures the breed's individually distinctive black and white coat pattern. Motivated by the labelling burden involved in constructing visual cattle identification systems, we propose exploiting the temporal coat pattern appearance across videos as a self-supervision signal for animal identity learning. Using an individual-agnostic cattle detector that yields oriented bounding-boxes, rotation-normalised tracklets of individuals are formed via tracking-by-detection and enriched via augmentations. This produces a 'positive' sample set per tracklet, which is paired against a 'negative' set sampled from random cattle of other videos. Frame-triplet contrastive learning is then employed to construct a metric latent space. The fitting of a Gaussian Mixture Model to this space yields a cattle identity classifier. Results show an accuracy of Top-1:* 57.0% *and Top-4:* 76.9% *and an Adjusted Rand Index:* 0.53 *compared to the ground truth. Whilst supervised training surpasses this benchmark by a large margin, we conclude that self-supervision can nevertheless play a highly effective role in speeding up labelling efforts when initially constructing supervision information. We provide all data and full source code alongside an analysis and evaluation of the system.*

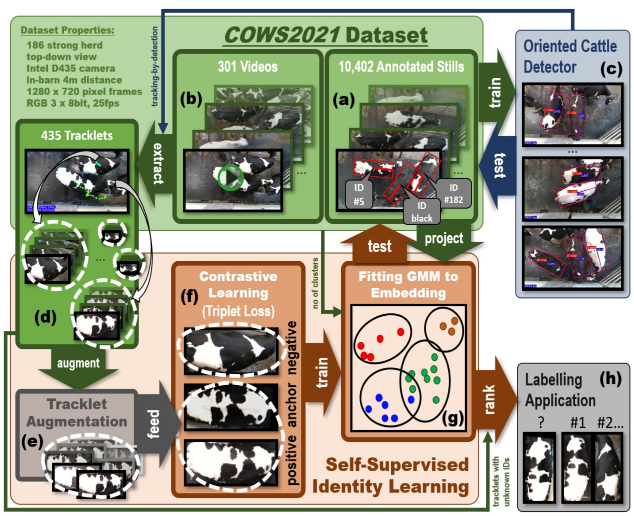


Figure 1: **Conceptual Overview**. Our dataset $Cows2021$ provides both (a) test images with oriented bounding-box and ID annotations, and (b) unlabelled training videos of the same herd. (c) ID-agnostic cattle tracking-by-detection across such videos yields (d) scale and orientation-normalised tracklets, which are (e) enhanced by augmentation. (f) Frame-triplets with in-tracklet anchor and positive ROI vs. out-of-video negative ROI are used for contrastive learning of a latent embedding, wherein (g) a GMM is fitted yielding an identity classifier by interpreting clusters as IDs. (h) ID labelling applications for building productions systems from video datasets can significantly benefit from having a confidence-ranked list of possible identities provided to the user.

1. Introduction and Background

Holstein-Friesians are, with a global population of 70 million [16] animals, the most numerous and also highest milk-yielding [41] cattle breed in the world. Cattle identification (ID) via tags [19, 12, 39] is mandatory [32, 31], yet transponders [23], branding [1, 9] and biometric ID [25] via face [11], muzzle [34, 26, 22, 42, 10, 14], retina [2], rear [36], or coat patterns [29, 20, 27, 8] are also viable. The last can conveniently operate from a distance above and has recently been implemented via supervised deep learning [6, 3]. However, research into reducing manual labelling efforts for creating and maintaining such ID systems is in its infancy [5, 44]. Particularly, unsupervised learning for coat pattern identification of Holstein-Friesians has not been tried and public datasets [4] are small to date.

This paper addresses these shortcomings and introduces the largest ID-annotated dataset of Holstein-Friesians: *Cows2021* so far, alongside a basic self-supervision system for video identification of individual animals (see Fig. 1).

Figure 2: **The *Cows2021* Herd**. Top-down, right facing view of the 186 individuals in the dataset normalised from RGB oriented bounding-box detections. Individually-characteristic black and white coat pattern patches are resolved at around 500×200 pixels. Note that 4 animals (2.2% of herd) carry no white markings and are excluded as 'un-enrollable' from the identification study.

2. Dataset *Cows2021*

We introduce the RGB image dataset *Cows2021*[1], which features a herd of 186 Holstein-Friesian cattle (see Fig. 2) and was acquired via an Intel D435 at University of Bristol's Wyndhurst Farm in Langford Village, UK. The camera pointed downwards from $4m$ above the ground over a walkway (see Fig. 3) between milking parlour and holding pens. Motion-triggered recordings took place after milking across 1 month of filming.

The dataset is resolved at 1280×720 pixels per frame with 8bit per RGB channel. It contains $10,402$ still images, in addition to 301 videos (each of length $5.5s$) at 30fps. The distribution of stills across individuals and time reflects the natural workings of the farm (see Fig. 4). Various expert ground truth (GT) annotations are provided alongside the acquired dataset.

Oriented Bounding-Box Cattle Annotations. Adhering to the VOC 2012 guidelines [15] for object annotation, we manually labelled[2] all visible cattle torso instances across the still image set. Annotations excluded the head, neck, legs and tail. Significantly clipped torso instances were (following [15]) not used further and given a 'clipped' tag. Example images from the resulting set of $13,165$ non-clipped cattle torso annotations are given in red in Fig. 3 (bottom). Each oriented bounding-box label is parameterised by a tuple: (c_x, c_y, w, h, θ) corresponding to the box centrepoint, width, height, and head direction.

[1] Available online at https://data.bris.ac.uk

[2] Tool used: https://github.com/cgvict/roLabelImg

Figure 3: **Dataset and Cattle Localisation**. Representative frames characterising the datatset. (top) Frames with varying animal orientation, crowding, clipping, and differing walking directions; $(middle)$ Frames with some motion blur, a mainly white cow with and without motion blur, and a near-boundary animal; $(bottom)$ Test images with oriented bounding-box annotations in red, output of ID-agnostic cattle detector in blue.

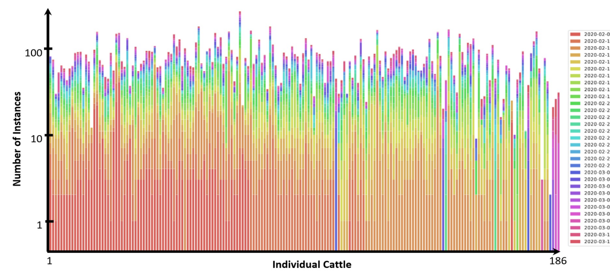
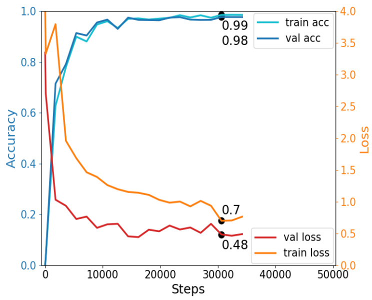


Figure 4: **Acquisition Across Individuals**. Number of still images captured of the 186 individuals, with time of acquisition across the month of recording shown as colour values.

Animal Identity Annotations. Overall $13,784$ detected (see Sec. 3) cattle instances were manually ID-assigned to one of 182 individuals (see Fig. 2). The 4 all-black cows were excluded from the ID study subject to future research. The number of occurrences of individuals varies from 2 to 273 with mean $\mu = 77.5$ and standard deviation $\sigma = 39.9$ (see Fig. 4). $8,670$ of these annotations were filmed on different days to the video data. These were used to form the identity test data.

Video Data and Tracklet Annotations. In addition to still images, the dataset contains videos with tracklet information designed for utilisation as a rich source of self-supervision in identity learning. Using a highly reliable ID-agnostic cattle detector (see Sec. 3) and sampling at 5Hz, tracking-by-detection was employed to connect nearest centrepoints of detections in neighbouring frames and thereby extract entire tracklets of the *same* individual (see Fig. 1). Manual checking ensured no tracking errors occurred. The average number of tracklets per video is 1.45.



Test AP	0.97
@IoU	0.7
@Conf-Thr	0.3
@NMS-Thr	0.28
#Train Images	7,248
#Val Images	1,023
#Test Images	2,131

Figure 5: **Cattle Detector Performance**. Training and validation curves, working point (approx. @$29k$ steps), test Average Precision (AP), and setup parameters for ID-agnostic cattle detector performing single frame oriented bounding-box detection.

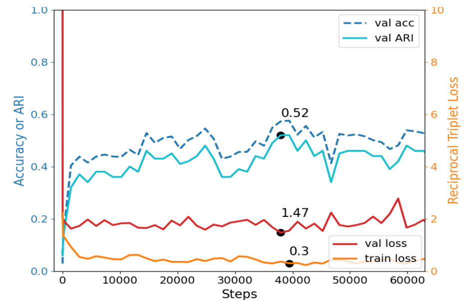


Figure 6: **Self-Supervised Identity Learning**. Training and validation curves, working point (approx. @38k Steps), accuracy and Adjusted Rand Index benchmarks for learning cattle identities via triplet loss.

3. ID-agnostic Cattle Detector

Existing multi-object single-frame cattle detectors [11, 7, 5] produce image-aligned bounding-boxes that cannot avoid capturing several individuals in crowded scenes (see Fig. 3), which is problematic for subsequent identity assignment. In response, we constructed a first orientation-aware cattle detector (see Fig. 3 blue) by modifying RetinaNet [28] with an ImageNet-pretrained [13] ResNet50 backbone [17]. We added additional target parameters for orientation encoding and rotated anchors implemented in 5 layers (P3 - P7). To train the network, we partitioned the still image set approximately $7:1:2$ for training, validation and testing, respectively. We used timestamps to split data so any temporal bias is reduced. We then trained the network against Focal Loss [28] with settings $\gamma = 2$, $\alpha = 0.25$, $\lambda = 1$ via SGD [38] with a learning rate of 1×10^{-5}, momentum of 0.9 [35], and weight decay of 1×10^{-4}. Fig. 5 illustrates training and depicts full performance benchmarks for the detector. For the test set, it operates at an Average Precision of 97.3% using an Intersection over Union (IoU) threshold of 0.7, reliably translating in-barn videos to tracklets.

4. Self-Supervised Animal Identity Learning

Given an ID-agnostic cattle detector (see Sec. 3), reliable tracklets can be generated (see Sec. 2) from readily available in-barn videos of a Holstein-Friesian herd. We investigated how far this data can be used to self-supervise the learning of filmed individual animals to aid the time-consuming task of manual labelling.

4.1. Contrastive Training

Identification Network and Triplet Loss. We use a ResNet50 [17] pretrained on ImageNet [13], modified to have a fully-connected final layer to learn a latent 128-dimensional ID-space. Across the training data of all videos, we normalise each tracklet for rotation (as seen in Fig. 2) and organise it into a 'positive' ID sample set representing the same, unknown individual. We pair this set against 'negative' samples from random cattle of other videos, which have a high chance of containing a different individual. All sets are enhanced via rotational augmentation (max. angle $\pm 7°$). The separate image data was used as a validation and testing base, split $1:3$. Reciprocal triplet loss (RTL) [30] is then employed for learning an ID-encoding latent space via an online batch hard mining strategy [18]:

$$\mathbb{L}_{RTL} = d(x_a, x_p) + \frac{1}{d(x_a, x_n)} \tag{1}$$

where x_a and x_p are sampled from the 'positive' set and x_n is a 'negative' sample. We trained the network for 7 hours via SGD [38] over 50 epochs with batch size 16, learning rate 1×10^{-3}, margin $\alpha = 2$, and weight decay 1×10^{-4}. The pocket algorithm [40] against the validation set was used to tackle overfitting (see Fig. 6).

4.2. Animal Identity Discovery via Clustering

Clustering. We then fitted [33] a Gaussian Mixture Model (GMM) [37] to the generated 128-dimensional space by setting the cluster cardinality to the known $k = 182$ patterned individual animals with 200 iterations. Resulting clusters are then interpreted as representing separate animal identities. A t-distributed Stochastic Neighbour Embedding (t-SNE) [43] of the training set projected into the clustered space is visualised in Fig. 7. In order to evaluate the clustering performance, we used two measures: the Adjusted Rand Index (ARI) [21] and ID prediction accuracy. For the latter, each GMM cluster is assigned to the one individual ID with the highest overlap which is defined as:

$$\mathbb{O}_l = C/L \tag{2}$$

where C is the number of images in a GMM cluster that belong to an individual, and L is the total number of images of the individual. This produces (GMM Cluster)-(ID Label) pairs for accuracy evaluation.

Top-N Accuracy. In order to quantitatively evaluate the capacity to aid human annotation, we consider a scenario

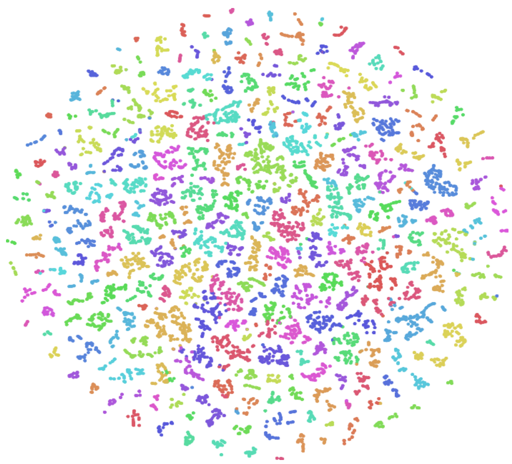


Figure 7: **Training Embeddings**. t-SNE plot of training data projected into the latent space and partitioned by the GMM into 182 identity clusters shown using random colours.

where a user annotates IDs as a one-out-of-N pick (expanding N if the correct ID is not present). Thus, the Top-N system accuracy [24] is a key measure to investigate. For each cluster one can rank all identities according to $\mathbb{O}_l$. Identities that have a $\mathbb{O}_l = 0$ form the randomly assigned tail of the sequence. For every data point this provides a general Top-N assigned ID. Finding the GT identity amongst the Top-N assigned IDs is then counted as correct identification.

5. Experimental Results and Discussion

Structural Clustering Similarity. In order to characterise the ID performance as if this were a new, unknown herd, we calculated the ARI to be 0.53 for the test set when measured between the partitioning derived from the clustering provided by the GMM versus the identity GT. This measure captures the (purely structural) similarities between the two clusterings.

Clustering Accuracy. In order to characterise the ID performance with class labels, we calculated Top-N accuracy for the test set as depicted in Table 1. Figure 8 visualises the identification performance and misclassification severity using a t-SNE plot.

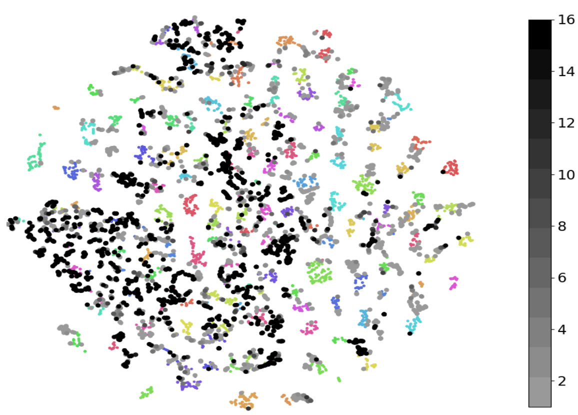


Figure 8: **Clustered Embedding**. t-SNE plot across test images. Colour indicates correct ID assignments of test data points, gray-to-black indicates Top-N severity of mismatch.

Context and Result Discussion. Considering that 182 classes were used and absolutely no training labelling was provided, results of 57.0% Top-1 accuracy and 76.9% Top-4 accuracy are an encouraging and practically relevant first step towards self-supervision in this domain. We know that individual Holstein-Friesian identification via *supervised* deep learning is a widely solved task with systems achieving near-perfect benchmarks when using multi-frame LRCNs [7] and good results even in partial annotation settings [5]. However, labelling efforts are laborious for supervised systems of larger herds; they require days if not weeks of manual annotation effort using visual dictionaries of animal ground truth. Humans can efficiently compare small sets of images. Thus, using the described pipeline we could present the user with a set of (e.g. 4) images that contain the correct individual with a chance better than 3-in-4. As part of a toolchain, the approach presented can potentially dramatically reduce labelling times and help bootstrap production systems via combinations of self-supervised learning followed by open set fine-tuning [5].

Top-N	N=1	N=2	N=4	N=8	N=16
Accuracy (%)	57.0	71.8	76.9	79.7	81.8

Table 1: **Top − N Performance**. Shown is ID accuracy for a variety of N across all $8,670$ test instances of the 182 identities.

6. Conclusion

In this paper we presented the largest identity-annotated Holstein-Friesian cattle dataset, *Cows2021*, made available to date. We also showed a first self-supervision framework for identifying individual animals. Driven by the enormous labelling effort involved in constructing visual cattle identification systems, we proposed exploiting coat pattern appearance across videos as a self-supervision signal. A generic cattle detector yielded oriented bounding-boxes which were normalised and augmented. Triplet loss contrastive learning was then used to construct a latent space wherein we fitted a GMM. This yielded a cattle identity classifier which we evaluated. Our results showed that the achieved accuracy levels are strong enough to help speed up ID labelling efforts for supervised systems in the future. Despite the need for even larger datatsets, we hope that the published dataset, code, and benchmark will stimulate research in the area of self-supervision learning for biometric animal (re)identification.

Acknowledgements. This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1 and the John Oldacre Foundation through the John Oldacre Centre for Sustainability and Welfare in Dairy Production, Bristol Veterinary School. Jing Gao was supported by the China Scholarship Council. We thank Kate Robinson and the Wyndhurst Farm staff for their assistance with data collection.

References

[1] Sarah JJ Adcock, Cassandra B Tucker, Gayani Weerasinghe, and Eranda Rajapaksha. Branding practices on four dairies in kantale, sri lanka. *Animals*, 8(8):137, 2018. 1

[2] A Allen, B Golden, M Taylor, D Patterson, D Henriksen, and R Skuce. Evaluation of retinal imaging technology for the biometric identification of bovine animals in northern ireland. *Livestock science*, 116(1-3):42–52, 2008. 1

[3] William Andrew. *Visual biometric processes for collective identification of individual Friesian cattle*. PhD thesis, University of Bristol, 2019. 1

[4] Will Andrew, Tilo Burghardt, Neill Campbell, and Jing Gao. The opencows2020 dataset, 2020. https://data.bris.ac.uk/data/dataset/10m32xl88x2b61zlkkgz3fml17. 1

[5] William Andrew, Jing Gao, Neill Campbell, Andrew W Dowsey, and Tilo Burghardt. Visual identification of individual holstein friesian cattle via deep metric learning. *arXiv preprint arXiv:2006.09205*, 2020. 1, 3, 4

[6] 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*, pages 2850–2859, 2017. 1

[7] William Andrew, Colin Greatwood, and Tilo Burghardt. Aerial animal biometrics: Individual friesian cattle recovery and visual identification via an autonomous uav with onboard deep inference. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 237–243. IEEE, 2019. 3, 4

[8] William Andrew, Sion Hannuna, Neill Campbell, and Tilo Burghardt. Automatic individual holstein friesian cattle identification via selective local coat pattern matching in rgb-d imagery. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 484–488. IEEE, 2016. 1

[9] Ali Ismail Awad. From classical methods to animal biometrics: A review on cattle identification and tracking. *Computers and Electronics in Agriculture*, 123:423–435, 2016. 1

[10] Ali Ismail Awad and M Hassaballah. Bag-of-visual-words for cattle identification from muzzle print images. *Applied Sciences*, 9(22):4914, 2019. 1

[11] Jayme Garcia Arnal Barbedo, Luciano Vieira Koenigkan, Thiago Teixeira Santos, and Patrícia Menezes Santos. A study on the detection of cattle in uav images using deep learning. *Sensors*, 19(24):5436, 2019. 1, 3

[12] W Buick. Animal passports and identification. *Defra Veterinary Journal*, 15:20–26, 2004. 1

[13] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 3

[14] Hagar M El Hadad, Hamdi A Mahmoud, and Farid Ali Mousa. Bovines muzzle classification based on machine learning techniques. *Procedia Computer Science*, 65:864–871, 2015. 1

[15] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html. 2

[16] Food and Agriculture Organization of the United Nations. Gateway to dairy production and products. `http://www.fao.org/dairy-production-products/production/dairy-animals/cattle/en/`. [Online; accessed 4-August-2020]. 1

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[18] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 3

[19] R Houston. A computerised database system for bovine traceability. *Revue Scientifique et Technique-Office International des Epizooties*, 20(2):652, 2001. 1

[20] Hengqi Hu, Baisheng Dai, Weizheng Shen, Xiaoli Wei, Jian Sun, Runze Li, and Yonggen Zhang. Cow identification based on fusion of deep parts features. *Biosystems Engineering*, 192:245–256, 2020. 1

[21] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985. 3

[22] Akio Kimura, Kazushi Itaya, and Takashi Watanabe. Structural pattern recognition of biological textures with growing deformations: A case of cattle's muzzle patterns. *Electronics and Communications in Japan (Part II: Electronics)*, 87(5):54–66, 2004. 1

[23] M Klindtworth, G Wendl, K Klindtworth, and H Pirkelmann. Electronic identification of cattle with injectable transponders. *Computers and electronics in agriculture*, 24(1-2):65–79, 1999. 1

[24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. 4

[25] Hjalmar S Kühl and Tilo Burghardt. Animal biometrics: quantifying and detecting phenotypic appearance. *Trends in ecology & evolution*, 28(7):432–441, 2013. 1

[26] Santosh Kumar and Sanjay Kumar Singh. Automatic identification of cattle using muzzle point pattern: a hybrid feature extraction and classification paradigm. *Multimedia Tools and Applications*, 76(24):26551–26580, 2017. 1

[27] Wenyong Li, Zengtao Ji, Lin Wang, Chuanheng Sun, and Xinting Yang. Automatic individual identification of holstein dairy cows using tailhead images. *Computers and electronics in agriculture*, 142:622–631, 2017. 1

[28] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 3

[29] Carlos A Martinez-Ortiz, Richard M Everson, and Toby Mottram. Video tracking of dairy cows for assessing mobility scores. 2013. 1

[30] Alessandro Masullo, Tilo Burghardt, Dima Damen, Toby Perrett, and Majid Mirmehdi. Who goes there? exploiting silhouettes and wearable signals for subject identification in multi-person environments. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019. 3

[31] United States Department of Agriculture (USDA) Animal and Plant Health Inspection Service. Cattle identification. `https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/nvap/NVAP-Reference-Guide/Animal-Identification/Cattle-Identification`. [Online; accessed 14-November-2018]. 1

[32] European Parliament and Council. Establishing a system for the identification and registration of bovine animals and regarding the labelling of beef and beef products and repealing council regulation (ec) no 820/97. `http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32000R1760`, 1997. [Online; accessed 29-January-2016]. 1

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 3

[34] WE Petersen. The identification of the bovine by means of nose-prints. *Journal of dairy science*, 5(3):249–258, 1922. 1

[35] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999. 3

[36] Yongliang Qiao, Daobilige Su, He Kong, Salah Sukkarieh, Sabrina Lomax, and Cameron Clark. Individual cattle identification using a deep learning based framework. *IFAC-PapersOnLine*, 52(30):318–323, 2019. 1

[37] Douglas A Reynolds. Gaussian mixture models. *Encyclopedia of biometrics*, 741:659–663, 2009. 3

[38] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951. 3

[39] C Shanahan, B Kernan, G Ayalew, K McDonnell, F Butler, and S Ward. A framework for beef traceability from farm to slaughter using global standards: an irish perspective. *Computers and electronics in agriculture*, 66(1):62–69, 2009. 1

[40] I Stephen. Perceptron-based learning algorithms. *IEEE Transactions on neural networks*, 50(2):179, 1990. 3

[41] Million Tadesse and Tadelle Dessie. Milk production performance of zebu, holstein friesian and their crosses in ethiopia. *Livestock Research for Rural Development*, 15(3):1–9, 2003. 1

[42] Alaa Tharwat, Tarek Gaber, Aboul Ella Hassanien, Hasssan A Hassanien, and Mohamed F Tolba. Cattle identification using muzzle print images based on texture features approach. In *Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014*, pages 217–227. Springer, 2014. 1

[43] Laurens JP van der Maaten and Geoffrey E Hinton. Visualizing high-dimensional data using t-sne. *Journal of machine learning research*, 9(nov):2579–2605, 2008. 3

[44] Maxime Vidal, Nathan Wolf, Beth Rosenberg, Bradley P Harris, and Alexander Mathis. Perspectives on individual animal identification from biology and computer vision. *arXiv preprint arXiv:2103.00560*, 2021. 1