Label a Herd in Minutes:  
Individual Holstein-Friesian Cattle Identification*  

Jing Gao, Tilo Burghardt, and Neill W. Campbell  

Department of Computer Science, University of Bristol, Bristol, United Kingdom  
jing.gao@bristol.ac.uk, tilo@cs.bris.ac.uk, neill.campbell@bristol.ac.uk  

Abstract. We describe a practically evaluated approach for training visual cattle ID systems for a whole farm requiring only ten minutes of labelling effort. In particular, for the task of automatic identification of individual Holstein-Friesians in real-world farm CCTV, we show that self-supervision, metric learning, cluster analysis, and active learning can complement each other to significantly reduce the annotation requirements usually needed to train cattle identification frameworks. Evaluating the approach on the test portion of the publicly available Cows2021 dataset, for training we use 23,350 frames across 435 single individual tracklets generated by automated oriented cattle detection and tracking in operational farm footage. Self-supervised metric learning is first employed to initialise a candidate identity space where each tracklet is considered a distinct entity. Grouping entities into equivalence classes representing cattle identities is then performed by automated merging via cluster analysis and active learning. Critically, we identify the inflection point at which automated choices cannot replicate improvements based on human intervention to reduce annotation to a minimum. Experimental results show that cluster analysis and a few minutes of labelling after automated self-supervision can improve the test identification accuracy of 153 identities to 92.44% (ARI=0.93) from the 74.9% (ARI=0.754) obtained by self-supervision only. These promising results indicate that a tailored combination of human and machine reasoning in visual cattle ID pipelines can be highly effective whilst requiring only minimal labelling effort. We provide all key source code and network weights with this paper for easy result reproduction.

Keywords: Precision Farming · Self-Supervision · Active and Metric Learning.

1 Introduction  

Background. Individual animal identification is mandatory [21] in dairy farming and critical for managing aspects such as disease outbreaks and animal welfare. To date, invasive identification methods [5] such as ear tags, tattoos, radio-frequency tags or branding are most often deployed. However, ethical considerations aside, these techniques cannot provide continuous location and ID information which still requires specialist tracking systems. For Holstein-Friesians, which constitute the most numerous and also highest milk-yielding [31] cattle breed, contactless visual biometric (re)identification methods [16,28] using their characteristic black-and-white skin markings [12,17,4,3,1,2] have become viable due to advances in deep learning. These approaches produce continuous coverage as long as cameras cover the whole farming area of interest. Similarly, face [30], muzzle [5], retina [5], or rear [23] biometrics may also be used in specific settings. However, modern biometric deep learning approaches that underpin systems for larger herds require significant amounts of identity-annotated visual data, demanding weeks of human annotation efforts [2].  

* Jing Gao was supported by the China Scholarship Council. This work also benefitted from ground work carried out under the EPSRC grant EP/N510129/1 and the John Oldacre Foundation. Many thanks to Andrew Dowsey and Will Andrew for their contributions.
**Conceptual Approach.** To address this problem and reduce labelling requirements, the literature has recently fielded self-supervision methods [9], which learn by exploiting the internal structure of data. However, although superior to traditional unsupervised approaches [37], the accuracy achieved with such systems still lags significantly behind benchmarks using supervised deep learning [2]. In response, here we advocate combining self-supervision, the analysis of the constructed identity space, and minimal active learning [29] to improve performance whilst limiting annotation requirements. Noting that research into visual cattle ID systems with reduced labelling is in its infancy [2,9,32], we propose a hybrid training approach [35,26,34,33] for learning cattle IDs from RGB videos. To isolate the impact of different sources of training information, we follow a three-phase strategy:

- **PHASE #1:** Exploit data-internal structure (via self-supervised metric learning).
- **PHASE #2:** Exploit latent identity space structure (via cluster analysis).
- **PHASE #3:** Utilise limited and targeted user input (via active learning).

![Fig. 1: Conceptual Overview](image)

Using the public Cows2021 dataset (purple) we put forward a 3-phase identity learning process (blue-green-orange) using Holstein-Friesian detection sequences (yellow) in RGB videos from an operating farm. Our approach combines self-supervised metric learning (blue), cluster analysis (green), and active learning (orange) to iteratively improve performance (middle plot) under a minimal labelling regime. First, single animal tracklets are formed via automated oriented object recognition and tracking (yellow). All frames from the tracklets are then mapped into a 128-dimensional identity space via metric learning and self-supervision (blue). On this space we perform cluster analysis and merge tracklets to form growing clusters of individual cattle (green) used for fine-tuning. We identify the inflection point at which machine merging cannot compete with user input and thus determine when active labelling (orange) is needed. We track the evolving identity space via t-SNE and show significant improvements beyond self-supervision, reducing annotation requirements to around ten minutes.

We will show that each of these phases can contribute to improving learning iteratively and together lead to benchmarks closer to fully supervised learning performance – yet at a fraction of the labelling effort. Fig. 1 illustrates the proposed approach visually.
Paper Contribution. Overall, this paper makes three key contributions to the field of visual learning in precision farming:

1. A principled and practical three-phase hybrid deep learning framework with reduced labelling requirements for training coat-based biometric cattle ID systems.
2. Integration of deep learning with a fast labelling approach (one-click same or different ID selection at 30 queries per minute) avoiding specific identity annotation, cold start issues [14], or seed labelling requirements.
3. A detailed system evaluation and comparative analysis on real-world farming videos across the public Cows2021 dataset including the identification of inflection points in the learning where human intervention becomes vital.

We proceed by providing details on the dataset, implementation, experiments, and results. Finally, we discuss how these methods and insights can be utilised for more rapid ID system roll out in precision farming.

2 Dataset
The Cows2021 dataset [6] contains 720p HD digital RGB video data from a working dairy farm in the United Kingdom. It was captured by a single-view top-down camera placed 4m above the ground between milking parlour and holding pens and operating at 25Hz (see Fig. 2 top). Excluding three particular animal identities (IDs: 155, 169, 182) with too little data for effective verification options or poor quality images, we utilise the remaining herd covering 179 individual cattle (see Fig. 2 bottom) in our study. The dataset is first automatically processed into 435 tracklets from 301 videos produced by a deep object detection (see Fig. 2 top) and tracking framework for cattle detailed in [9]. This ID-agnostic extraction performs rotational normalisation resulting in tracklets (see Fig. 2 middle) that contain exactly one individual with an average number of 1.45 tracklets per video. Further statistics and data splits are given in Fig. 2. Note that four totally black cows (IDs: 54, 69, 73, 173) were treated as one individual during training and were excluded in the validation and test sets to avoid systematic errors arising by mixing this particular physical anomaly with other aspects of performance. Finally, for open-set testing we withheld 24 individual cattle from training and regular testing altogether. Since each tracklet contains data of only a single individual, self-supervision can be used to associate images to identity classes.

3 Implementation
3.1 PHASE #1: Self-Supervision
Metric Deep Learning and Identity Space Construction. As our base architecture we use a ResNet50 [10] backbone pre-trained on ImageNet [7] modified to have a fully-connected final layer mapping to a latent 128-dimensional vector from triplet image inputs detailed in [2]. To initialise this space for identity information without any supervision we treat each tracklet as a unique class representing the same, unknown individual forming a set of ‘positive’ image samples. One video may contain one or more tracklets. We pair these sets against ‘negative’ samples from cattle shown in other tracklets from different videos, i.e. we use a video-aware sampling strategy. Statistically, the fact that the same ‘positive’ individual may – with a small chance – appear in some ‘negative’ sample in a different video is accepted as training noise during this initialisation stage.
Fig. 2: Dataset. Our training data contains 23,350 frames (4670 frames and their four augmentations) from 301 RGB videos resolved at 1280 × 720 pixels. Automated oriented object detection and tracking [9] of these frames (top) yields 435 normalised tracklets (middle). (bottom) The tracklets cover 155 individuals, all of which feature in the training set. 153 individuals are used in the test set (5344 extra images), and 149 in the validation set (2480 extra images). The four all-black cattle are represented here by a single image. Red boundaries denote individuals not included in the test set, both red and blue boundaries are not included in the validation set, and the last 24 individuals (654 images) after the blank area are unseen cattle during training used for open-set evaluation only.

For self-supervised learning we use reciprocal triplet loss \( L_{RTL} \) [20] leading to an initial version of the identity space that optimises:

\[
L_{RTL} = d(x_a, x_p) + d(x_a, x_n)^{-1}
\]

where \( d \) denotes the Euclidean distance, \( x_a \) and \( x_p \) are sampled from the ‘positive’ set and \( x_n \) is a ‘negative’ sample. We utilise online batch hard mining [11] expanded to a search that exploits both anchor and negative samples (see code base for implementation details). Training of this stage took approx. 7 hours on an RTX2080 node using SGD [27] over 50 epochs with batch size 16, learning rate \( 1 \times 10^{-3} \), margin \( \alpha = 2 \), and weight decay \( 1 \times 10^{-4} \). The pocket algorithm [30] against the validation set was used to address network overfitting according to the Adjusted Rand Index (ARI).

Using t-distributed Stochastic Neighbour Embedding (t-SNE) [19] for visualisation, Fig. 3(a) depicts the initial identity space after ImageNet pre-training and before self-supervision. As described above, the self-supervised metric learning then initialises this identity space as given in Fig. 3(b) leading to distinct local groupings of data points that relate to tracklets (i.e. image groups) of single cattle identities.

3.2 PHASE #2: Cluster Analysis

Clustering the Identity Space. Our aim is that the identity space should be partitioned such that all tracklets are grouped to relate to individual cattle identities. However, we start by discovering clusters as given by the data in the space so far. We achieve such a partitioning by fitting a Gaussian Mixture Model (GMM) [25,22] with \( g = 450 \) components (this parameter is non-critical, provided that it is set to a number a little over the total number of tracklets) over 150 iterations to the tracklet-initialised identity space (see Fig. 4(left)). Our approach is to utilise the discovered clusters to merge tracklets automatically into groups that represent the same individuals.
Label a Herd in Minutes

(a) START: Identity Space (ImageNet-Initialised)
(b) PHASE #1: Self-Supervised Metric Learning
(c) PHASE #2: Automatic Cluster Merging
(d) PHASE #3: 2 Minutes Merging (Human Annotation)
(e) PHASE #3: 10 Minutes Merging (Human Annotation)
(f) Ideal Partitioning

Fig. 3: Identity Space Evolution. 2D t-SNE visualisations of the training data projected into the generated 128D latent identity space through the different phases of learning where colours relate to ground truth identities. Note that after (a) ImageNet initialisation, (b) self-supervision yields a tracklet-structured space where around three times more clusters are present than individual identities. (c)-(e) Merging of these groupings and subsequent fine-tuning based on this leads to the step-wise discovery of actual identities and fewer and fewer overall clusters, first conducted via automatic cluster analysis in Phase #2 and finally via active learning in Phase #3.

**Associating GMM Clusters with Tracklets.** We hypothesise that different GMM clusters which cover tracklets of the same individual should be near to each other in identity space and that the similarity between tracklets can be evaluated based on Euclidean tracklet-to-tracklet distances. Therefore, we avoid calculating the distance of all the combinations of tracklet pairs and instead work with the top-$k$ matches to each tracklet utilising the GMM for calculation. The pseudocode is presented in Algorithm 1.

First, all data points $t \in T$ in the 128-dimensional embedding space are partitioned via the fitted GMM yielding $g = 450$ data point clusters (i.e. GMM components) $T^m \subset T$ where $m \in [1, g]$ denotes the cluster label. Next, the dominant tracklet ID marked as $\tau_m$ is assigned to each cluster label $m$. Let $T_\theta \subset T$ be the set of datapoints associated with the input tracklet ID $\theta$. The intersection set $T_\theta \cap T^m$ denoted as $T_\theta^m$ is then used to find the most frequent tracklet ID $\theta$ therein denoted $\tau_m$ and derived via a function ‘Mode’ (line 8 of the Algorithm).

**Selecting Candidates for Merging.** Next, we find candidate tracklet pairs for merging based on the associated GMM information. We use the fact that the GMM can return
Algorithm 1: Find candidates of tracklet pairs to merge based on GMM fit

**Input:** Data Points $T$ in Identity Space;
Data Points $T_\theta \subset T$ with tracklet ID $\theta$;

**Output:** Candidate List $R$ of tracklet ID pairs;

1. $T_m \leftarrow \text{GMM}_{g=450}(T)$, $T$ is partitioned into clusters $T_m$ with labels $m$;
2. $T^m_\theta = T_\theta \cap T^m$, Data Points with tracklet ID $\theta$ in cluster $T^m$;
3. for $k \leftarrow 1$ to 4 do
   4. if $k = 1$ then
      5. $\tau_m \leftarrow \text{Mode}(T^m_\theta)$;
   6. else
      7. $l^m_k \leftarrow \text{Select}(k, \text{GMM}, T^m, T)$;
      8. $\tau_{l^m_k} \leftarrow \text{Mode}(T^m_{l^m_k})$;
   9. $R \leftarrow \text{list.append}(\tau_m, \tau_{l^m_k})$;
4. end
5. end

Probabilities that an input $t$ belongs to any particular cluster $m$. Across all data points in any GMM cluster $T^m$ we collect the $k^{th}$ ($k \in [2, 4]$) most likely cluster labels, that is for each cluster $T^m$ and $k$ we find the most frequent cluster label $l^m_k$ via a function ‘Select’, which effectively identifies the $k^{th}$ nearest cluster to $T^m$. Each cluster identified by $l^m_k$ has itself a dominant tracklet ID $\tau_{l^m_k}$. With this information in hand, we finally compile the list $R$ of candidate tracklet ID pairs by pairing the dominant tracklet ID $\tau_m$ with $\tau_{l^m_k}$.

**Candidate Ranking and Cluster Merging.** We now rank the generated candidate tracklet pairs $(T_\theta, T_\eta)$ associated to the list of tracklet ID pairs in $R$ based on the tracklet-to-tracklet distance between each two tracklets as suggested in previous work [34]. The distance between two tracklets is calculated as the mean Euclidean distance between all point pairs that exist between the tracklets. Candidates ranked towards the top are considered closest in identity space and have indeed a high chance of representing the same individual. The dotted brown curve in Fig. 5 (left) confirms this hypothesis against the ground truth by tracking the rate of correct merging. The graph shows a clear downwards trend of this rate when moving lower down the ranking. The rate first declines slowly, but then drops rapidly.

**Identity Space Update via Fine-Tuning.** After merging, we fine-tune the network for a further 5h of training as described in Section 3.1 in order to incorporate the new information derived from cluster merging. This yields an updated network which defines a new embedding function for the identity space. By measuring the validation performance of this network against the number of tracklet mergers (or queries) we identify the point at which automatic merging cannot improve network performance (measured via ARI) anymore. This occurs after 173 queries as shown as the peak of the green curve in Fig. 5 (left). Automatic merging after this point is thus too noisy to further improve performance. Consequently, user interaction via active learning should start at this point if further performance improvements are required.
3.3 PHASE #3: Active Learning

User-guided Fine-Tuning. Active learning [24,29] aims to interactively annotate the most informative samples from the training model, followed by a model update. A domain-relevant sample selection is absolutely critical [18,8,33]. For our case, having constructed a domain-informed ranked list $R$ of potential tracklet ID mergers already, we will further utilise this information to guide active learning. Essentially, we will continue the tracklet merging strategy as before, but exploit user input to verify the merging instead of blindly accepting it. Thus, the final phase refines the identity space with a query sequence of one-click user inputs and utilises the new candidate ranking after another round of self-supervision retrained with the automatically merged tracklets. For manual annotation, a human is shown the next highest ranking potential merger tracklet pair from $R$ presented as a single image from each of two tracklets. By confirming the identity to be the same or different only (one click), the human annotator throughput with this approach is extremely high at approximately 30 answers per minute. The approach avoids specific identity annotations (i.e. selection from a catalogue) and it has no cold start issues [14]. Practically, whenever two tracklets sharing the same identity are identified by an annotator, they are merged into a single tracklet and the number of tracklets is reduced as before during automatic merging. Note that it is possible that across various mergers – more than two tracklets may merge into a single tracklet due to the transitive nature of the process. As before: after merging, fine-tuning for 5 hours, is used to incorporate new information derived from cluster merging into the deep network. The resulting networks are benchmarked as orange curves (marked ‘Active’) in Fig. 5.

4 Experimental Results

Closed Set Testing. The proposed pipeline is first evaluated on the 5,344 unseen test images across 153 individuals. This is closed-set testing since all these individuals have been seen (in different videos) during training. A projection of the closed testset into the learned identity space is visualised via t-SNE in Fig. 7(a)). In order to evaluate
Fig. 5: **Individual Identification Performance.** System performance (closed-set) across the three learning phases shaded as blue, green, and orange. Benchmarks provide *(left)* ARI measurements and *(right)* Top-1 ID accuracy along evolving system training. Evaluation is conducted on 5k+ unseen test images. Phase #1 in blue depicts results of self-supervised metric learning over the last 50 epochs leading up to the 0 mark of the abscissa indicating the start of tracklet merging. Phases #2 and #3 show results of fine-tuning after automated and manual merging, respectively. The abscissa tracks the number of mergers (or queries). Note, that at 173 queries, automatic merging (green curve) peaks and further mergers are too noisy to improve performance without manual correctness checking via active learning (orange curve). Note further that performance after 10 min of labelling at accuracy 92.44% is close to the performance of perfectly merging all clusters exhaustively (95.17%) based on ground truth. The dotted brown curve on the left quantifies the proportion of correct merger suggestions.

the clustering performance in identity space, we use two measures: the Top-N ID prediction accuracy and also the Adjusted Rand Index (ARI) [13] for assessing structural clustering similarity.

**Structural Clustering Benchmarks.** We evaluate the structural similarity of the testset clustering produced by the trained network against the ground truth. We quantify the similarity via the ARI measure which operates on two data partitions and does not require any knowledge of ID labels. In particular, the testset clustering is derived by fitting a GMM (with the component cardinality of the testset) to the test data in the fine-tuned identity space. Fig. 5 *(left)* shows the evolution of ARI across the three learning phases. Self-supervision of Phase #1 as described in Section 3.1 leads to an initial ARI value of 0.75 (see blue curve). In Phase #2, automatic tracklet merging and subsequent fine-tuning as described in Section 3.2 steadily improves the ARI further to a peak value of 0.86 at query 173 (see green curve). At this point, we find that any further automatic merging is not beneficial as depicted by the dotted green curve. In Phase #3, user input is now utilised. The ARI can be increased further to 0.88 after 3.5 minutes of human labelling, 0.91 after 8 min, and 0.93 after 10 min. Perfectly merging all clusters exhaustively only leads to a small increase of ARI at 0.95.

**Accuracy Stipulation.** Next, we evaluate identification accuracy against ground truth ID labels of the testset. In order to obtain IDs for our network output, each fitted GMM cluster used for structural benchmarking is assigned to the one individual ID having the highest overlap ratio with the ground truth, defined as:

\[
\hat{C} = \frac{C}{L}
\]

1 Note that this performance improves on the self-supervision state-of-the-art [9] in the domain by using the same network yet with our extended hard mining regime. Testing our Phase #1 method on their testset improves ARI from their published ARI of 0.53 to 0.65.
where $C$ is the number of images in a GMM cluster that belongs to an individual, and $L$ is the total number of images of that individual. This produces the (GMM Cluster)-(ID Label) pairs required for accuracy evaluation against the ground truth.

**Top-1 Accuracy Benchmarks.** The resulting Top-1 test accuracy for individual identification over the three training phases is given in Fig. 5 (right). Generally, the plot structure is consistent with our ARI analysis in Fig. 5 (left) showing the effectiveness of all three phases of our approach across the two different performance measures. Phase #1 leads to an accuracy of 76.12% already on the test set, emphasising the strength of a tracklet-driven self-supervision signal during metric learning. In Phase #2, automatic tracklet merging can further improve accuracy to a peak value of 85.95% (see green curve). Note that automatic merging is imperfect and accuracy is not optimised directly by fine-tuning, thus a temporary decrease of accuracy is possible. In Phase #3, human labelling allows a further increase in accuracy to 86.28% after 3.5 minutes of human labelling, 90.62% after 8 minutes, and finally 92.44% after 10 minutes.

**Top-N Accuracy Benchmarks.** For each GMM cluster assigned to an ID as per Eq. 2, one can alternatively rank all identities according to $O_l$ where identities that have a $O_l = 0$ form the tail of the ranking sequence with randomly assigned, remaining IDs. This process allows the creation of a Top-N [15] accuracy benchmark as shown in Fig. 6. In this statistic, a prediction is considered correct if and only if the ground truth ID is found amongst the top N ranked IDs. As expected, for $N = 1$ this equates to the traditional definition of accuracy. Fig. 6 emphasises that across Phases #2 and #3 improve accuracy across all N. This setting is particularly interesting if semi-automatic identification is used to present a system user with a set of N candidate identities for a query. For $N = 16$ this setting leads to a near-perfect accuracy benchmark of 96.59% using only 10 minutes of labelling.

**Open-Set Accuracy Benchmarks.** The proposed pipeline is secondly evaluated on the 654 unseen test images of 24 never seen individuals. A projection of this open testset into the learned identity space is visualised via t-SNE in Fig. 7 (c). This ‘Open-Set’ testing is critical for stipulating how far performance can translate across herds or farms in a zero-shot identification paradigm. Fig. 6 depicts Top-1 open-set test performance.
(a) CLOSED TESTSET, 153 animals. (10min Annotation)  (b) CLOSED TESTSET, 153 animals. (Fully Supervised)  (c) OPEN TESTSET, 24 animals. (10min Annotation)

Fig. 7: t-SNE Testset Embeddings. 2D t-SNE visualisation of the closed and open test sets when projected into the identity space. (a) shows the result for the proposed method at accuracy 92.44% after all three stages of training and 10min of manual annotation. (b) For comparison, we show the same testset projected into an identity space built from fully manual annotation at the Top-1 accuracy 96.05% using weeks of frame-by-frame identity labelling. (c) When projecting unseen images of unseen animals into the space used in (a) identities still cluster well (accuracy 90.81%) confirming generalisation to a domain wider than the herd trained on.

across the final two phases of training. It can be seen that open-set performance generally lags behind closed-set performance. However, the margin (gap between blue and yellow curves) is only 1.53% after 10 minutes of manual labelling. This is promising, showing that the identity space created generalises well across the domain of Holstein-Friesians and not just the training herd.

5 Conclusion

We have put forward a practical, three-phase deep learning approach for training an ID system for Holstein-Friesians on a farm, requiring only ten minutes of labelling. We showed that automatic identification of individual cattle in real-world farm CCTV can be achieved effectively by combining self-supervision, metric learning, cluster analysis, and active learning. We provided detailed explanations and key source code\footnote{https://github.com/Wormgit/LabelaHerdinMinutes} for full result reproduction and evaluated the approach using the publicly available Cows2021 dataset. Self-supervised metric learning was first leveraged to initialise an identity space where tracklets are considered a distinct entity. Grouping entities is then performed by automated merging via cluster analysis and active learning feeding into fine-tuning. Experimental results showed that cluster analysis and a few minutes of labelling after automated self-supervision can indeed improve the identification accuracy of closed and open test sets compared to self-supervision only. Despite superior performance of fully supervised systems which required weeks of frame-by-frame labelling in the past, our 10 minute labelling approach shows promising results indicating that human and machine reasoning in tandem can be integrated into visual cattle ID pipelines in a highly effective fashion requiring only minimal labelling effort.
References

1. Andrew, W.: Visual biometric processes for collective identification of individual Friesian cattle. Ph.D. thesis, University of Bristol (2019)
2. Andrew, W., Gao, J., Mullan, S., Campbell, N., Dowsey, A.W., Burghardt, T.: Visual identification of individual holstein-friesian cattle via deep metric learning. Computers and Electronics in Agriculture 185, 106133 (2021)
3. Andrew, W., Greatwood, C., Burghardt, T.: Visual localisation and individual identification of holstein friesian cattle via deep learning. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2850–2859 (2017)
4. Andrew, W., Hannuna, S., Campbell, N., Burghardt, T.: 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), pp. 484–488. IEEE (2016)
5. Awad, A.I.: From classical methods to animal biometrics: A review on cattle identification and tracking. Computers and Electronics in Agriculture 123, 423–435 (2016)
6. Campbell, N., Burghardt, T., Gao, J., Andrew, W.: The cows2021 dataset (2021), https://data.bris.ac.uk/data/dataset/4vnrc7qw1642qlwxjdp87lh7
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
8. Freytag, A., Rodner, E., Denzler, J.: Selecting influential examples: Active learning with expected model output changes. In: European conference on computer vision. pp. 562–577. Springer (2014)
9. Gao, J., Burghardt, T., Andrew, W., Dowsey, A.W., Campbell, N.W.: Towards self-supervision for video identification of individual holstein-friesian cattle: The cows2021 dataset. arXiv preprint arXiv:2105.01938 (2021)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
11. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 (2017)
12. Hu, H., Dai, B., Shen, W., Wei, X., Sun, J., Li, R., Zhang, Y.: Cow identification based on fusion of deep parts features. Biosystems Engineering 192, 245–256 (2020)
13. Hubert, L., Arabie, P.: Comparing partitions. Journal of classification 2(1), 193–218 (1985)
14. Konyushkova, K., Szmitman, R., Fua, P.: Learning active learning from data. In: NIPS. pp. 4228–4238 (2017), http://papers.nips.cc/paper/7010-learning-active-learning-from-data
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 25, pp. 1097–1105. Curran Associates, Inc. (2012), http://papers.nips.cc/paper/14824-imagenet-classification-with-deep-convolutional-neural-networks.pdf
16. Kühl, H.S., Burghardt, T.: Animal biometrics: quantifying and detecting phenotypic appearance. Trends in ecology & evolution 28(7), 432–441 (2013)
17. Li, W., Ji, Z., Wang, L., Sun, C., Yang, X.: Automatic individual identification of holstein dairy cows using tailhead images. Computers and electronics in agriculture 142, 622–631 (2017)
18. Lindenbaum, M., Markovitch, S., Rusakov, D.: Selective sampling for nearest neighbor classifiers. Machine learning 54(2), 125–152 (2004)
19. van der Maaten, L.J., Hinton, G.E.: Visualizing high-dimensional data using t-sne. Journal of machine learning research 9(nov), 2579–2605 (2008)
20. Masullo, A., Burghardt, T., Damen, D., Perrett, T., Mirmehd, M.: 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. pp. 0–0 (2019)
21. Parliament, E., 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]
22. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830 (2011)
23. Qiao, Y., Su, D., Kong, H., Sukkarieh, S., Lomax, S., Clark, C.: Individual cattle identification using a deep learning based framework. IFAC-PapersOnLine 52(30), 318–323 (2019)
24. Ren, P., Xiao, Y., Chang, X., Huang, P.Y., Li, Z., Gupta, B.B., Chen, X., Wang, X.: A survey of deep active learning. ACM Computing Surveys (CSUR) 54(9), 1–40 (2021)
25. Reynolds, D.A.: Gaussian mixture models. Encyclopedia of biometrics 741, 659–663 (2009)
26. Rizve, M.N., Duarte, K., Rawat, Y.S., Shah, M.: In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. In: International Conference on Learning Representations (2020)
27. Robbins, H., Monro, S.: A stochastic approximation method. The annals of mathematical statistics pp. 400–407 (1951)
28. Schneider, S., Taylor, G.W., Kremer, S.C.: Similarity learning networks for animal individual re-identification-beyond the capabilities of a human observer. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops. pp. 44–52 (2020)
29. Settles, B.: Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin-Madison (2009)
30. Stephen, I.: Perceptron-based learning algorithms. IEEE Transactions on neural networks 50(2), 179 (1990)
31. Tadesse, M., Dessie, T.: Milk production performance of zebu, holstein friesian and their crosses in ethiopia. Livestock Research for Rural Development 15(3), 1–9 (2003)
32. Vidal, M., Wolf, N., Rosenberg, B., Harris, B.P., Mathis, A.: Perspectives on individual animal identification from biology and computer vision. Integrative and comparative biology 61(3), 900–916 (2021)
33. Wang, K., Zhang, D., Li, Y., Zhang, R., Lin, L.: Cost-effective active learning for deep image classification. IEEE Transactions on Circuits and Systems for Video Technology 27(12), 2591–2600 (2016)
34. Wang, M., Lai, B., Jin, Z., Gong, X., Huang, J., Hua, X.: Deep active learning for video-based person re-identification. arXiv preprint arXiv:1812.05785 (2018)
35. Wu, Y., Lin, Y., Dong, X., Yan, Y., Ouyang, W., Yang, Y.: Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5177–5186 (2018)
36. Yao, L., Hu, Z., Liu, C., Liu, H., Kuang, Y., Gao, Y.: Cow face detection and recognition based on automatic feature extraction algorithm. In: Proceedings of the ACM Turing Celebration Conference-China. pp. 1–5 (2019)
37. Yu, H.X., Wu, A., Zheng, W.S.: Cross-view asymmetric metric learning for unsupervised person re-identification. In: Proceedings of the IEEE international conference on computer vision. pp. 994–1002 (2017)