Multi-views Embedding for Cattle Re-identification

Luca Bergamini*
luca.bergamini24@unimore.it

Angelo Porrello*
angelo.porrello@unimore.it

Andrea Capobianco Dondona†‡, Ercole Del Negro†‡, Mauro Mattioli‡, Nicola D'Alterio‡, Simone Calderara*

†Farm4Trade Srl, Chieti, Italy
*AImageLab, University of Modena and Reggio Emilia, Modena, Italy
‡Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise 'G.Caporale', Teramo, Italy

***Abstract*—People re-identification task has seen enormous improvements in the latest years, mainly due to the development of better image features extraction from deep Convolutional Neural Networks (CNN) and the availability of large datasets. However, little research has been conducted on animal identification and re-identification, even if this knowledge may be useful in a rich variety of different scenarios. Here, we tackle cattle re-identification exploiting deep CNN and show how this task is poorly related with the human one, presenting unique challenges that makes it far from being solved. We present various baselines, both based on deep architectures or on standard machine learning algorithms, and compared them with our solution. Finally, a rich ablation study has been conducted to further investigate the unique peculiarities of this task.**

***Index Terms*—Cattle, Identification, Convolutional Deep Network, Multi-view Embedding, Animal biometrics**

I. Introduction

A. Animal Re-identification Motivations

Animal Re-identification shares some of the aims of the human task, while also including new challenges. The identification process represents a pillar for national and international trade, especially for animals representing crucial economic assets. Furthermore, it constitutes a method for validating the quality and the "authenticity" of the animal being traded. Similarly, for animals supplying products intended for human consumption, the animal identity and the traceability along the entire value chain are prerequisites for the certification of the quality and the safeness of the product for final consumers. In fact, as some of these animals may host and transmit pathogens, a monitoring system is essential to avoid the spread of such diseases to humans and animals and it is necessary to easily identify and track the origin of infected products. Finally, stock theft represents an issue that often outbreaks into a social challenge in developing countries. As an example, India reported almost ten thousand cattle thefts in 2015 while the number of horse theft has grown past forty thousands world wide [1], [2].

On the other hand, re-identification systems for pets has seen some interest in the computer vision community [3], mainly aiming to retrieving lost "family members".

Finally, re-identification systems may represent an opportunity for safeguarding endangered species, acting as a crucial aid for studying wildlife and for conservation actions. For such animals, traditional identification systems make use of electronic chips placed in collars and require the animal to be captured and immobilised at least once. Such practice may be unfeasible for aggressive or elusive species, and typically requires GSM or satellite transmitters, the latters being very expensive and often impractical. Again, only few noticeable novel and unobtrusive methods [4], [5], [6] have been proposed, mainly due to the lack of large datasets publicly available.

B. Cattle Re-identification Motivations

The number of cattle in Europe in 2016 stood at almost 122 million, of which 23.3 million were dairy cows [7], and the number is growing past 1400 million in the world based on the latest surveys [8]. Although efficient identification and re-identification systems already exists, it is mandatory to develop new tools that can support the existing ones not only to ensure milk and meat safeness, but also to avoid kidnappings and counterfeits while improving animal health and animal welfare. Nowadays, the following methods are employed:

- RFID subcutaneous chips or rumen bolus, which can be read using a dedicated electronic reader;
- Ear tags, holding the animal identification number according to the country legislation format;
- Brand code, marked on the animal skin and used as a traditional identification system mainly in developing countries.

Authors from [9] reported further details on RFID and electronic identification system for cattle in the US market, while the reader may refer to [10] for an extensive review of cattle identification systems worldwide.

C. Machine Learning Motivations and Insights

The methods reported above suffer from major drawbacks. In particular, the use of RFID devices entails a significant cost for farmers of developing countries, because of the installation fees and the need of electronic readers during re-identification procedures. Moreover, they may, sometimes, have a sensible impact on animal welfare. On the other hand, ear tags are cheaper to buy but can be easily counterfeit or even removed through ear excision, beside being often lost by the animal

itself. Consequently, there is room and a strong need to develop new methods with the following requirements:

- Cheap for both installation and maintenance, including the supporting hardware;
- Able to be easily and rapidly used in real scenarios, as instance in the field or in a stable;
- Hard to be counterfeit or removed.

Re-identification based on images holds all these properties, and can be exploited using the latest techniques and advances from Deep Learning and Computer Vision. Differently from traditional machine learning, deep learning techniques do not require any human hand-crafted features as they learn those representations directly from data, identifying features that may be more robust to pose or backgrounds variations as well as illumination changes. This is especially needed for cattle, since it is not easy to obtain images with a predefined pose of the animal, as it tends to move constantly while roaming or eating.

While for humans it is widely recognised that the face holds a great importance for visual identification purposes, in the animal kingdom a similar certainty still lacks, as almost no studies for this specific task have been conducted yet. Cattle present a high inter-breed variance in both body proportions and skin textures. On one hand, this makes fairly easy to distinguish cows of different breeds even for novices, on the other hand, due to the genetic selection made by humans in the past centuries and even more in the last decades, cows present a lower inter-breed variance compared to humans. However, despite this quite high degree of inbreeding, many cow breeds hold a unique texture pattern that is different from animal to animal, while also behaviour and social interaction contribute with marks, scratch and other defects that remain evident on the animal skin.

In this work we used pictures of the head of cows taken from different angles of rotation and inclination, essentially for the following reasons:

- The head of a cow shows a sufficient characteristic set of textures, shapes and patches. Even for textures-less breeds (such as the Bruna Alpina), the presence of horns and their length or the fur colour contributes to this variety. Furthermore, [11] showed how cattle face muscles are sufficiently developed to exhibit different facial expression, that may be used to distinguish one from another.
- Most of our images have been collected in farms with cows restrained during veterinary procedures, and only the head is easily accessible;
- Pictures of the full cow would introduce variance in both the animal pose and the background and could possibly require even more images of the same animal to perform the re-identification task.

Furthermore, an approach based on "facial" images could be compared with the current literature available on human re-identification.

D. Main contributions and Novelties

The contributions of this research are two-fold. Firstly, we provide a deep learning based framework for cattle re-identification. Secondly, we demonstrate how the use of multiple views of the same cow leads to superior performance w.r.t. standard approaches, the latters typically using only the front view image of the subject.

E. Paper Structure

The rest of this paper is structured as follows. Section II presents re-identification methods on both human and animal, focusing in particular on cattle. Section III introduces our proposed method, detailing the base block of our CNN architecture. Our cattle dataset is introduced and described in IV. Section V shows extensive experiments and comparison with the baselines, while also discussing their performances. Section VI further investigates our proposed architecture and its parts function. Finally, section VII summarises the contributions and results of the article.

II. Related Works

A. Human Re-identification

Human Re-identification has a long history of both research and practical uses. Among early methods, EigenFaces [12] has proved to perform well on cropped and aligned faces, such as the Olivetti dataset [13], where it achieves a re-identification accuracy of 95%. FisherFaces [14] employed a classifier based on the Linear Discriminant Analysis (LDA), exploiting features coming from a preprocessing phase involving a Principal Component Analysis (PCA) stage. In this way, it merges information from multiple views of the same subject in the final classifier. However, both these methods are unable to deal with unaligned faces and suffer from illumination changes.

Since they are widely known for extracting more invariant features, Local Binary Pattern Histograms (LBPH) [15], Histogram of Oriented Gradients (HOG) and Scale Invariant Features Transform (SIFT) [16] descriptors have been widely used. As some of the more extend variations usually occur in illumination changes and scale, these descriptors have been designed to be robust for these applications. However, the human face has more than 15 muscles producing some of the most complex expression in nature, which can alter dramatically the final appearance of a person face. In the latest years, DCNN trained on huge datasets, such as [17], [18] and [19] have provided features learned directly from examples with a growing interest from the computer vision and machine learning communities. Among them, [20] and [21] show state-of-the-art performances, as they can handle different facial expressions as well as face aging. Even with some differences, both architectures produce a low-dimensional feature vector, namely an embedding of the input face, efficient to be compared with others using Nearest Neighbours classifiers.

B. Animal Re-identification

Nowadays, little efforts have been made for the animal Re-identification task, with the noticeable exception of apes, since

they share common traits with human beings. [4] achieved 98.7% on facial images coming from 100 red-bellied lemurs (Eulemur rubriventer) of the Ranomafana National Park, Madagascar. The authors employed multi-local binary patterns histograms (MLBPH) and Linear Discrimant Analysis (LDA) on cropped and aligned faces. The authors show how human faces recognition algorithms can be adapted to primates, as their faces share the same underlying features. Following their work, [5] expanded the re-identification task to multiple apes species using DCNN. On one hand, the authors tried some traditional baseline such as EigenFaces and LBPH, on the other hand, they exploited two state-of-the art human faces identification networks, namely FaceNet and SphereFace. Using the last, the authors showed how a CNN trained for solving a human re-identification task may also be adapted to the primates one, leveraging a fine-tuning strategy. However, they achieved slightly superior performance by means of a smaller CNN (in terms of number of layers and parameters), which is trained from scratch on apes' faces from three different species. In both above mentioned approaches, the underlying assumption lies on the presence of some similarity between the human and the ape faces. This evidence has been corroborated by two surprising results: firstly, the network trained on human faces perform enough well also on apes (showing comparable results with a network trained from scratch on apes [5]) and secondly, it is able to extract and work with facial landmarks.

Other endangered species have also attracted interest in the latest years, from zebras [22] to tigers [23]. [6] proposed a deep learning technique aiming to automatically identify different wildlife species, as well as counting the occurrences of each species in the image. Differently from us, such methods work on images depicting the entire animal's body, exploiting the characteristic stripe patterns of such animals.

Finally, pets represent an opportunity to build larger dataset, as they outnumber the above mentioned animals of a large margin, but public datasets still do not exist, and collecting this data requires huge resources and efforts. As an example, using pictures of two dogs breeds gathered from Flickr, [3] achieved remarkable performances. As dog faces differ from humans ones, the authors developed two Deep CNN trained from scratch on dogs images only, after a pre-processing phase consisting of a tight crop to suppress most of the background.

C. Cattle Re-identification

Due to their economic value, a literature regarding cattle-identification methods is slowly stemming in the latest years. However, to the best of our knowledge, this is the first attempt in doing it using the animal face with Deep Learning techniques.

[24] employed images from Unmanned Aerial Vehicle (UAV) to identify cattle of a single breed using individual stripes and patches. Firstly, the authors gathered a dataset of 89 different cows depicted in 980 RGB images, the latters being captured by a camera placed over the walkway between holding pens and milking stations of a farm. Secondly, the authors presented a CNN trained from images, as well as a complete pipeline involving a Long-Short Term Memory (LSTM) layer to exploit temporal information. They achieved 86.07% identification accuracy on a random train-test split and 99.3% detection and localisation accuracy.

Similarly, [25] developed a system based on histograms and movements to record images of the backs of 45 cows from a camera placed on the Rotary Milking Parlour, for a period of 22 days. The authors trained a DCNN to perform individual identification, achieving an outstanding 98.97% of accuracy. The collection system was able to correctly detect the back of the cow and crop it from the image. Using this approach, the system was able to record a huge amount of data with a great variation of light condition.

Since in the above described methods cows' pictures are taken from above, they are not suited and comparable with our scenario. Indeed, both approaches would require the constant presence of an expensive UAV for image acquisition or several difficulties for an operator. Therefore, their use could be extremely limited especially in developing countries. Furthermore, for textures-less animals, a picture of the back holds little details compared with one of the face, where traits such as the eyes and the muzzle vary greatly.

III. Proposed Method

In order to leverage the textures and details of both cattle profiles (i.e. the frontal and one of the two sides), we built an embedding DCNN starting from two images of the same cow.

At a high level, the network takes in input the two images and subsequently outputs a 128 dimensional embedding with unitary L2 norm. More in detail, each of the two images is independently processed by a separate convolutional branch, and their outputs are concatenated to form a single feature vector. It is worth noting that the two branches do not share any parameter, since different features may be required for the two animal profiles.

Our multi-view network has been designed by means of two building blocks:

- **ConvBlock**: A single Convolutional layer followed by InstanceNorm [26] and LReLU activation, reducing the feature maps' spatial resolution by a half;
- **ResBlock**: A residual unit [27] with LReLU activation, preserving the spatial resolution.

A scheme of these blocks is shown in Figure 1.

Instead of using the more popular batch normalisation, we address the internal covariate shift problem by means of instance normalisation. Indeed, with the second, we observed improvements in terms of stability during the training phase. The network ends with a single 2D Convolution, with kernel size equals the feature map size and number of filters equals to the desired embedding size (i.e. 128). In this way, each input map is reduced to a single scalar value, leading to a fully convolutional architecture. The overall architecture is presented in Figure 2.

We employed the Histogram Loss from [28] as the only loss function of our architecture. After a batch of anchors, positives and negatives is embedded into a high-dimensional

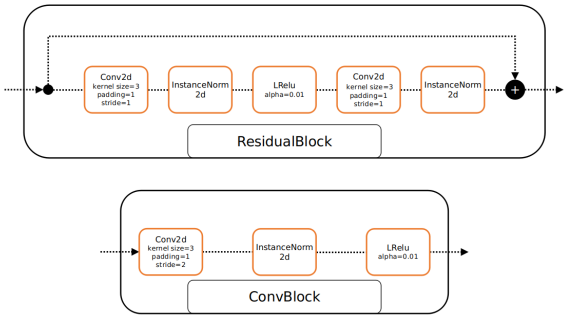

Fig. 1: Base blocks used in our architectures.

space by a deep network, the loss computes the histograms of similarities of positive and negative pairs. The integral of the product between the negative distribution and the cumulative density function for the positive distribution is evaluated, corresponding to a probability that a randomly sampled positive pair has smaller similarity than a randomly sampled negative pair. We performed extensive comparison using the triplet loss function described in [20], but found the latter more unstable during the train phase.

IV. Dataset

A potential drawback in the use of Deep Learning methods is that a huge amount of pictures depicting cows heads is required to achieve reasonable performances on unseen examples. Furthermore, the training set should include a great variety of poses, illumination changes and background for each subject, with the acquisition process spanning potentially in multiple days. However, to the best of our knowledge, such dataset still does not exist for cows' faces. Thus, we collected pictures and video of cattle from four Italian farms distributed in three regions. We collected videos and extracted images from those for the training process, while employing only pictures acquisitions for the test phases. We leveraged the Vatic tool [29] to annotate the cows' faces with a bounding box for each frame. Finally, we discarded some of the extracted frames aiming to ensure a high inter-frames variance.

Such activity should be considered as mandatory, since the animal usually moves slowly during video acquisition, introducing a lot of redundancy if all the frames are kept. Moreover, a traditional setting for the re-identification task consists of few different pictures per single identity.

Eventually, we obtained the following splits:

- Train Set; consisting of 12952 pictures from 387 different subjects;
- Database Set; consisting of 4289 pictures from 52 different subjects, recorded during two different days;
- Test Set; consisting of 561 pictures from 52 different subjects. These cows are the same included in the Database Set;

Some random samples from the last two set are shown in Table I. It is worth noting that, given an image, one cannot make any

TABLE I: Some randomly drawn samples from our dataset.

assumptions regarding the cow's face location and orientation. Moreover, because of the oblong shape of cow faces, any alignment would lead to part of the cow face being cropped. Finally, only few landmarks detector work on animal faces [30], but they need to be fine-tuned on the animal domain, thus requiring expensive landmarks annotations.

V. Experiments

A. Metrics

Following [5] and [21], we test our solution in two different settings:

- **Open-Set**: Identities of the test images are included in the train set.
- **Closed-Set**: Identities of the test images are separated from the train set ones.

Regarding the last, we consider it as more challenging and general, being also able to provide a good estimation of the generalisation capacity of the proposed model. For both settings, we conducted the same experiment, namely the **Identification**. Given images and correspondent ground-truth identities from a test set, the matching strategy returns the k nearest neighbours from the database set. It is worth noting that the above mentioned "matching strategy" can be implemented by every classifier.

B. Baselines

We include both deep and non-deep baselines to further present and motivate the main challenges of this novel task. For non-deep baselines, we include methods traditionally employed by the computer vision community for the human re-identification task. The reason behind this choice is not only to emphasise the differences between the human related task and the cattle one, but also because the great majority of them are provided as open-source verified software with the open-cv [31] package.

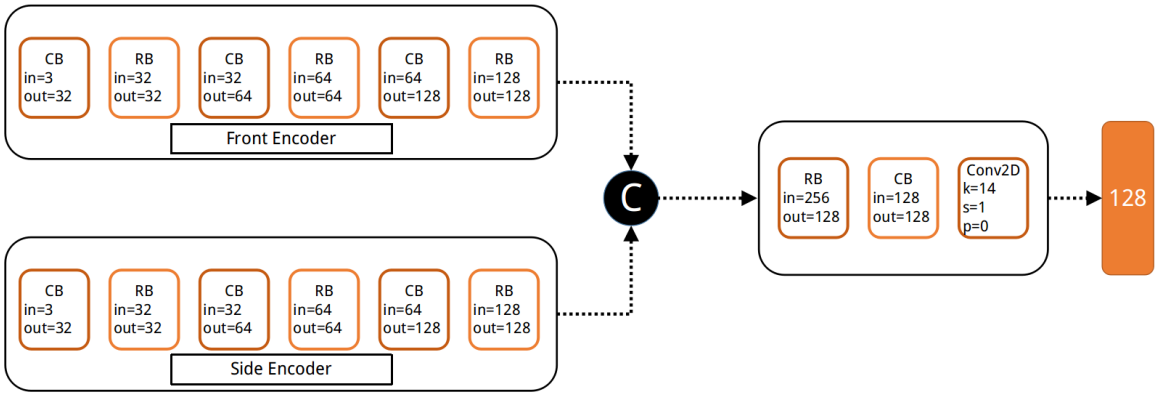

Fig. 2: Our multi-views architecture. Note that CB stands for ConvBlock, while RB for ResidualBlock, as described in Figure 1

EigenFaces [12] method consists of a Principal Component Analysis (PCA) applied to images of human faces. The first k eigenvectors, sorted by their respective eigenvalues magnitudes, can be seen as prototypes used to build the data. Test images can then be projected to extract k coefficients, each one describing how much a prototype contributes to the image.

FisherFaces [14] extends Eigenfaces by means of the Fisher Discriminant Analysis (FDA), aiming to force multiple images of the same identities to lie in a nearby region of the subspace. While PCA preserves maximum variance during the projection, FDA attempts to preserves the discrimination capability at the end of such transformation. Indeed, FDA may be consider a supervised embedding algorithm, since it finds a projection that maximises the scatter between classes and minimises scatter within classes. Thanks to this, the Fisherface method shows superior performances w.r.t. Eigenfaces under the presence of variation in lighting and expression.

LBPH [15] has been widely used as a texture description. It builds a circular neighbour with a certain radius for each pixel, and extracts features based on the relationships between each pixel and its neighbour. The latters are carried out by a histogram, which is then used to describe the image, and can act as a feature descriptor for a further PCA.

HOG, similarly to LBPH, builds a histogram using neighbours but, instead of pixels values, the spatial derivatives are employed. The histogram takes into account both the magnitude and the orientation of the gradients, with the latter being quantized to contribute to achieving invariance to orientation.

The authors of SphereFace [21] present a DCNN trained on a large human faces dataset, achieving state-of-the-art performances in an open-set setting. The images are firstly aligned and cropped, and a 512-dimensional feature vector is extracted from the second-last layer of the network. A classification loss, named Angular Softmax, is proposed: on one hand, it requires examples from the same identity to lie nearby on the output landscape. On the other hand, it forces examples from different identities to be spaced by a considerable margin, the latter being in the form of an angle on a hypersphere. As versions of the network pretrained on human faces are available, results with and without a training phase on cattle are reported in subsection V-D.

It is worth noting that all methods listed above share the same output representation; in particular a feature vector (embedding) is produced from a given input image and, as such, the "matching strategy" can be the same for all of them.

C. Implementation Details

As far as it regards the train methodology, it is worth noting that:

- For the **closed set** scenario we train on both the Train and the Database Set, while for the **open set** one we train only on the first;
- We pre-processed the images by scaling them to a fixed size (i.e. 224).
- We performed data augmentation by randomly rotating, cropping and projecting images, while also changing the hue and saturation of the images. We didn't perform horizontal flip, as it causes a noticeable drop in performances;
- We employed the Histogram Loss using a batch size of 64 triplets and 200 bins for the histograms;
- We mined both hard positives and negatives during the training phase. The firsts are positives with an embedding extremely far away from the average embedding of the identity, while the lasts are the closer negatives to the identity in the embedding space;

D. Results

As shown in Table II, we compare our results with the baselines discussed in V-B. For each method, the hyper-parameters tuning activity has been conducted using a grid-search strategy. For the identification task, results are reported in terms of Top1 and Top3 match, for both open and close settings. For every methods, KNN (k-nearest neighbours) has been employed as final classifier, as it requires no hyper-parameters grid-search nor supervised training, while also scaling linearly with the number of identities.

	EigenFaces		FisherFaces		LBPH		HOG		SphereFace		SphereFace(cattle)		Ours	
	Open	Close	Open	Close	Open	Close	Open	Close	Open	Close	Open	Close	Open	Close
Top1	0.227	0.229	0.242	0.237	0.263	0.265	0.39	0.4	0.139	-	0.367	0.556	0.558	**0.817**
Top3	0.352	0.354	0.345	0.345	0.406	0.415	0.522	0.532	0.226	-	0.46	0.636	0.742	**0.891**

TABLE II: Results for the Identification task.

Fig. 3: Illustration of the K-NN retrievals, computed on two individuals using a single view or a double view architecture. Input images have been labelled with blue contour, while green and red have been used respectively for correct and misclassified examples. Best view in colours.

The input of each method is a single image, whereas our proposed method can be provided with two different profiles images. In order to enable a fair comparison and to motivate our design choices, results with only one profile image are reported in subsection VI-A.

Looking at results showed in Table II, the following conclusion may be drawn:

- Methods based on local and stationary property (i.e. LBPH, HOG and CNNs) achieve better performance than Eigenfaces and Fisherfaces, which do not implicitly exploit nearby pixel correlations or local pattern's presence.
- Deep models trained on cattle performs better than shallow ones, due to their robustness to different poses and other major source of variations (i.e background or illumination changes).
- SpereFace, when trained on aligned human faces, does not generalise to cattle. Such result show how the cattle re-identification task has nothing to do with the same task in the human domain, differently from what happen with apes.
- Our solution outperforms by a consistent margin all the other competitors, including a state-of-the-art human re-identification network as SphereFace (even if trained from scratch on cattle). Such improvement is achieved by

	Single-View		Multi-View	
	Open	Close	Open	Close
Top1	0.443	0.688	0.558	**0.817**
Top3	0.575	0.748	0.742	**0.891**

TABLE III: Comparison between single and multi views methods.

	Database Set		Extended Set	
	Open	Close	Open	Close
Random3	0.057	0.057	0.006	0.006
Top1	0.558	**0.817**	0.396	0.732
Top3	0.742	**0.891**	0.583	0.820

TABLE IV: Comparison with different set during the test phase. Random3 stands for a random predictor scoring 1 if the correct identities lies among the first 3 prediction.

leveraging two different cow's profile, which leads to a higher discriminative capability for similar subjects.

VI. Ablation Study

A. Multi-view vs Single-view

Results for the Identification task reported in Table III highlights the superiority of the multi-view approach over the single-view one. 3 shows some cherry-picked example where the single-view model fails, while the multi-view approach effectively fuses features from both views to produce a more representative and robust embedding vector. This justifies the request of both views for a single prediction. Indeed, if two cows may have the same visual appearance under some poses or particular light conditions, on the other hand this possibility has a lower probability under the presence of both profiles.

Moreover, cattle usually do not share a symmetry between left and right profile, as they often present very different spots and patterns. Also for such reason, the use of two profiles instead of one should be considered useful to find a meaningful discrepancy between two cows.

B. Extended Database

In Table IV we report results, in term of accuracy, showing how the use of the only database set during the test phase leads to better performances with respect to the union of the train and database sets. However, for the close set scenario, the drop of performance between the two settings is much lower w.r.t the open one, highlighting how the network improves its behaviour if the animal's images are available during the train phase. In this way, such knowledge may be used during the test phase to reject other subjects with similar patterns or characteristics. To further highlight the differences in terms of difficulty between the two settings a random predictor has been included.

VII. Conclusions

In this work we propose a Deep Learning base method for cattle re-identification in unconstrained environment from single and multi-views. We present extend baseline comparisons both with non-deep and deep methods. We show that human and cattle re-identification are slightly similar tasks, but present important and significant differences. Finally, we highlight how a multi-views method (i.e a method combining information from multiple profiles) clearly outperforms both baselines and single-view methods.

Acknowledgment

The authors thank Allevamento Martin, Allevamento Costantini, Cooperativa Venditti and Allevamento Hombre for providing access to their farms during the dataset acquisition. Moreover, the authors would like to acknowledge Francesco di Tondo, Lisa Leonzi Giuseppe Carolla and Carmen Sabia for their precious help with data acquisition.

References

[1] "Cattle theft statistics," https://en.wikipedia.org/wiki/Cattle_theft_in_India, accessed: 2018-09-316.

[2] "Cattle theft statistics," https://www.nytimes.com/2013/05/27/world/asia/cow-thefts-on-the-rise-in-india.html, accessed: 2018-09-316.

[3] T. P. Moreira, M. L. Perez, R. de Oliveira Werneck, and E. Valle, "Where is my puppy? retrieving lost dogs by facial features," *Multimedia Tools and Applications*, vol. 76, no. 14, pp. 15 325–15 340, 2017.

[4] D. Crouse, R. L. Jacobs, Z. Richardson, S. Klum, A. Jain, A. L. Baden, and S. R. Tecot, "Lemurfaceid: a face recognition system to facilitate individual identification of lemurs," *BMC Zoology*, vol. 2, no. 1, p. 2, 2017.

[5] D. Deb, S. Wiper, A. Russo, S. Gong, Y. Shi, C. Tymoszek, and A. Jain, "Face recognition: Primates in the wild," *arXiv preprint arXiv:1804.08790*, 2018.

[6] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," *Proceedings of the National Academy of Sciences*, p. 201719367, 2018.

[7] "Dairy cattle in europe statistics," https://dairy.ahdb.org.uk/market-information/farming-data/cow-numbers/eu-cow-numbers/#.W6aLCRyxVhF, accessed: 2018-09-316.

[8] "Cattle fao world statistics," http://www.fao.org/faostat/en/, accessed: 2018-09-316.

[9] J. Evans, "Livestock identification," 2005.

[10] M. Bowling, D. Pendell, D. Morris, Y. Yoon, K. Katoh, K. Belk, and G. Smith, "Identification and traceability of cattle in selected countries outside of north america," 2008.

[11] K. B. Gleerup, P. H. Andersen, L. Munksgaard, and B. Forkman, "Pain evaluation in dairy cattle," *Applied Animal Behaviour Science*, vol. 171, pp. 25–32, 2015.

[12] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[13] "Orl dataset," https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[14] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," Yale University New Haven United States, Tech. Rep., 1997.

[15] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1. IEEE, 1994, pp. 582–585.

[16] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2. Ieee, 1999, pp. 1150–1157.

[17] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

[18] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *European Conference on Computer Vision*. Springer, 2016, pp. 87–102.

[19] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[20] F. Schroff, D. Kalenichenko, and J. 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.
[21] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2017, p. 1.
[22] M. Lahiri, C. Tantipathananandh, R. Warungu, D. I. Rubenstein, and T. Y. Berger-Wolf, "Biometric animal databases from field photographs: identification of individual zebra in the wild," in *Proceedings of the 1st ACM international conference on multimedia retrieval*. ACM, 2011, p. 6.
[23] L. Hiby, P. Lovell, N. Patil, N. S. Kumar, A. M. Gopalaswamy, and K. U. Karanth, "A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins," *Biology letters*, pp. rsbl–2009, 2009.
[24] W. Andrew, C. Greatwood, and T. Burghardt, "Visual localisation and individual identification of holstein friesian cattle via deep learning," in *Proc. IEEE International Conference on Computer Vision (ICCV), Venice, Italy*, 2017, pp. 22–29.
[25] T. T. Zin, C. N. Phyo, P. Tin, H. Hama, and I. Kobayashi, "Image technology based cow identification system using deep learning," in *Proc. of the International MultiConference of Engineers and Computer Scientists IMECS 2018, Hong Kong*, vol. 1, 2018.
[26] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *CoRR*, vol. abs/1607.08022, 2016. [Online]. Available: http://arxiv.org/abs/1607.08022
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[28] E. Ustinova and V. Lempitsky, "Learning deep embeddings with histogram loss," in *Advances in Neural Information Processing Systems*, 2016, pp. 4170–4178.
[29] C. Vondrick, D. Patterson, and D. Ramanan, "Efficiently scaling up crowdsourced video annotation," *International Journal of Computer Vision*, pp. 1–21, 10.1007/s11263-012-0564-1. [Online]. Available: http://dx.doi.org/10.1007/s11263-012-0564-1
[30] M. Rashid, X. Gu, and Y. J. Lee, "Interspecies knowledge transfer for facial keypoint detection," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2017.
[31] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

APPENDIX
ALIGNMENT IMPORTANCE FOR HUMAN FACES

As already stated in IV, alignment for animal faces poses major challenges in terms of annotation costs. Furthermore, since cattle posses oblong-shaped heads, an alignment using landmarks may lead to crops with deformed cow's proportions. To investigate the real impact of landmarks alignment, in terms of re-identification accuracy over human faces, we test the SphereFace architecture trained on CASIA-Webfaces on a split from the CELBA dataset, both with and without alignment. It is worth noting that even when the faces are not aligned, they are still cropped tightly around the subject face. Results from Table V show a noticeable drop of performance using non-aligned faces. Even if CNN should be robust to translations, other transformations such as rotations and prospective seem to affect performances.

	SphereFace (Aligned)	SphereFace (Not Aligned)
Top1	0.976	0.390
Top3	0.978	0.565

TABLE V: Results raised comparing SphereFace both on aligned or not faces.