Identification of cattle facial features via deep learning

A. Chlingaryan[1,*], S. Ye[2], M. Patching[3] and C.E.F. Clark[1]
[1]*Dairy Science Group, Livestock Production and Welfare Group, School of Life and Environmental Sciences, University of Sydney, NSW, 2570, Australia*
[2]*School of Computer Science, University of Sydney, NSW, 2006, Australia*
[3]*Alta Food and Agriculture Pte Ltd, 538 Upper Cross St, #04-98, Singapore 051531*
[*]Corresponding author: Anna Chlingaryan, anna.chlingaryan@sydney.edu.au

Abstract

Cattle identification plays an irreplaceable role in understanding disease trajectories, animal traceability, and animal ownership assignment. Although traditional non-biometric methods of cattle identification have been used extensively, issues with retention, cost and intentional removal remain. Here we developed a robust Deep Learning based cattle facial biometric feature identification system using cattle head images extracted from video as the input. The developed methodology differentiated the cattle face from the background via a combination of instance and semantic segmentation. The images as video frames were then normalized to ensure consistency in the sense of image size and cattle head location. The output of the system was a set of identified biometric features for each image.

A case study was conducted using videos of 25 cattle from 2-year-old beef cattle recorded at the University of Sydney John Bruce Pye Farm. The videos were pre-processed, and a database of 5,000 high quality frontal images was created. The performance of different neural networks previously used for human facial feature identification was then compared. The MobileNetV2 neural network demonstrated the best performance for automated identification of cattle facial key regions achieving 99% precision and recall for muzzle detection, and 98% precision and recall for eye detection. Thus, the proposed approach enables an automated computation of a range of biometric features including, and not limited to, the relative distances between eyes, between eye and muzzle and the angle between different facial features. The availability of automatically calculated biometric features enables work towards the identification of individual cattle.

Keywords: biometric identification, object detection, cattle traceability, image segmentation, deep learning

Introduction

Artificial intelligence (AI) with the aid of computer vision is boosting various sectors for quality production with high efficiency. In the field of livestock production, AI extends its help to farmers and manufacturers to improve decision making. Cattle identification plays an irreplaceable role in understanding disease trajectories, vaccination and production control, animal traceability, and animal ownership assignment. As traditional methods are susceptible to loss, fraud, and duplication, more robust and non-invasive biometric-based cattle identification via AI and computer vision techniques have been developed (Awad, 2016). Cattle biometric recognition approaches include coat pattern analysis, facial recognition, muzzle print pattern analysis, retinal scanning (Baranov et al., 1993; Kumar and Singh, 2017; Kumar et al., 2016; Shojaeipour et al., 2021). Biometric identification of cattle from muzzle patterns has previously indicated promising results using a two-stage YOLOv3-ResNet50 algorithm for muzzle detection followed by biometric identification (Shojaeipour et al., (2021). However, those studies relied on high-resolution and high-quality muzzle print or eye pattern images which are hard to capture in real-world scenarios. Here we utilised video of cattle heads captured on-farm and created a database of manually extracted frontal images for automatic computation of cattle facial biometrics with potential applications to cattle traceability throughout livestock supply chain.

Materials and methods

Our two-stage object detection approach is shown in Figure 1. The first stage of the approach consists of getting eyes and muzzle regions proposals per image generated by the EdgeBoxes (Zitnick and Dollár, 2014). The second stage is the classification of the proposed regions followed by multiple filtering layers to extract the regions of interest (ROI).

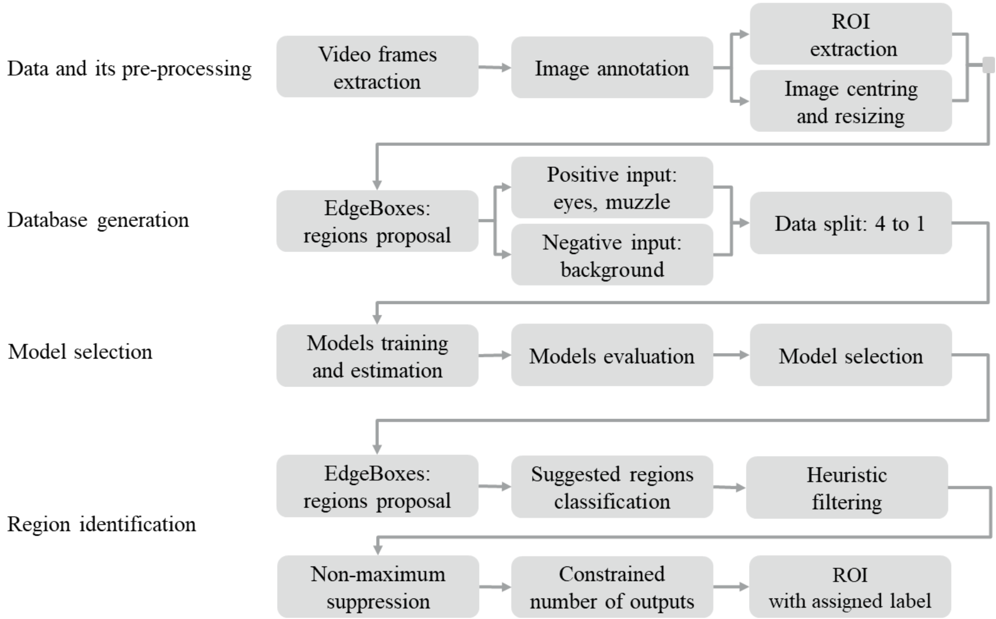

Figure 1: Workflow of eyes and muzzle automated identification in cattle

The choice of the two-stage approach was driven by the need to have a light-weight model as the one-stage object detection algorithms require more computational resources, memory, and time. The bibliometrics for one-stage object detection and two-stage object detection are detailed by Lohia et al. (2021).

Data and pre-processing

Animal use in this experiment was approved by the University of Sydney Animal Ethics Committee (#2021/1916). Videos of 25 white two-year-old Charolais cattle were recorded at the University of Sydney John Bruce Pye Farm. A GoPro HERO 10 camera was used, set at 120fps and 4K resolution. While the cattle were constrained by a crush, 12 seconds of video was recorded via camera that was manually held at the cattle eye level and approximately 50cm horizontally away from its head.

Python code was developed for the pre-processing of the recorded videos and creating the database required for this experiment. The first step was an automated extraction of the frames from the video of each individual cattle followed by manual selection of frontal images. As a result, a database of 5,000 frontal images (i.e., images with visible two eyes and muzzle) with resolution of 3,840×2,160 pixels was created (left image in Figure 2).

The second step was an automated detection of the animal in the image via a combination of deep learning-based instance and semantic segmentation methodologies (Bolya et al., 2019; Garcia-Garcia et al., 2017). The background of each image was then removed using the image’s masks obtained through the segmentation. (Figure 2, image in the middle).

As the distance between the cattle head and camera was not fixed, the images were normalized by centering and cropping them to the same size. To calculate the center point of the cattle head in image, the regions of eyes and muzzle were manually annotated via VGG Image Annotator (Dutta et al., 2016). The following three labels were introduced: muzzle, right eye, left eye. The labels were assigned to the corresponding regions around the features visible in the image (Figure 3). The center point of the cattle head in image was defined as the center of the triangle with the vertices in the centers of the regions for left eye, right eye, and muzzle. Finally, all the images were cropped using the calculated center point and the fixed size of 720×720 pixels (Figure 2, right image).

Figure 2: Frontal image through the pre-processing flow

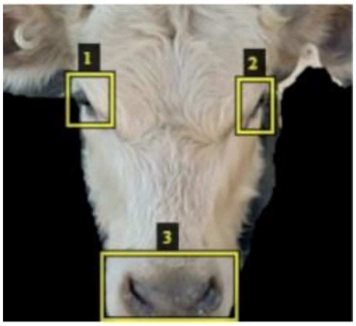

Figure 3: Annotation of the regions of interest in image: muzzle, right eye, left eye.

Database generation

To generate a comprehensive database of eyes, muzzle and background for the model selection step described in the workflow (Figure 1), the EdgeBoxes approach (Zitnick and Dollár, 2014) was applied to the 4000 pre-processed images of randomly selected 20 cattle. The intersection over union (IOU) was calculated between the regions suggested by the EdgeBoxes and the manually annotated ROI for each image. All the annotated regions were considered as the positive inputs of eyes and muzzle. The suggested regions with the IOU of 0% were considered as the negative inputs and were labeled as background. All the inputs were resized to 224×224 pixels, which is a commonly used input size for neural network applications. The generated database resulted in 8,000 images of eyes, 4,000 images of muzzle, and 16,000 images of background as shown Figure 4.

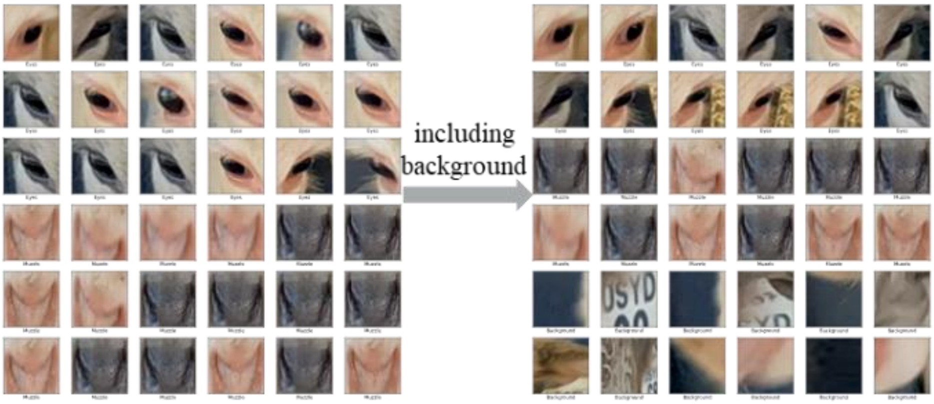

Figure 4: An example of the generated database: extracted and re-sized images of eyes and muzzle (left) and with the addition of background images (right).

Model selection

The architecture of MobileNet, VGG, ResNet50, DenseNet and Xception neural networks (and their versions) commonly used in image classification was slightly modified to fit the context of usage. The backbone of the networks was kept the same and the output layers were modified - as shown in Figure 5 for the MobileNetV2 (Sandler et al., 2018). Specifically, the final dense layers of the networks were replaced by an average pool layer with pool size 7×7 followed by the flatten layer and the dense layer with 128 neurons. To prevent overfitting, the dropout rate was set to 0.5. The final decision dense layer was replaced by 3 outputs representing the likelihood of eye, muzzle, and background. This modification aimed to reduce the inferring time and to enhance the robustness.

To evaluate the performance of the modified five networks, the generated database of eyes, muzzle, and background was randomly split into five parts. One part was used for testing while the other four parts were further split into training and validation sets using 5-fold split (4 for training and 1 for validation). This data split process was stratified resulting in almost equal proportions of eyes, muzzle, and background in each fold. To compute the cross-entropy loss between the labels and predictions, categorical cross-entropy loss function was used. The batch size and learning rate were set to 256 and 0.001, respectively. As an adaptive optimizer can reach higher accuracy at the early stage of training (Zhou et al., 2020), Adam optimizer was utilised.

The performance of the modified five networks was compared to select the most suitable network. The comparison was based on the three metrics: running time, accuracy and number of parameters, aiming to select the network that achieves a high accuracy with as few parameters as possible in an acceptable time.

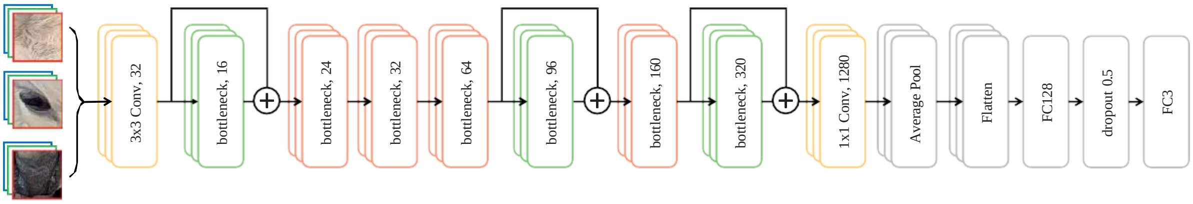

Figure 5: Architecture of the modified MobileNetV2 network

Region identification

The output of the region identification step was used for automated evaluation of the performance of the proposed approach for identification of eyes and muzzle in cattle Figure 1. As the database of images of the 20 cattle were already used for model selection, 100 pre-processed images of the remaining 5 cattle were utilised for the performance evaluation.

The EdgeBoxes approach was applied to each of the 100 images resulting in 3000 suggested regions per image. The regions were filtered by selecting only the ones that had their dimensions within the variability range of eye and muzzle sizes observed in the dataset used for the model selection. All the selected regions were classified via the chosen neural network and the regions with the estimated probability higher than 68% were selected. This threshold was chosen based on visual inspection of classification outcomes on the model selection database. Further filtering of the regions identified as eye or muzzle was conducted using the dimension variation range of the corresponding feature. To replace the overlapping regions classified as the same feature by a single representative region, the Non-Maximum Suppression (NMS) algorithm (Hosang et al., 2017) was applied. The input to the NMS algorithm was the regions with the corresponding estimated probabilities. The final step was the selection of the region with the highest estimated probability per the general area in the image where the corresponding feature (i.e., left eye, right eye and muzzle) was expected to be.

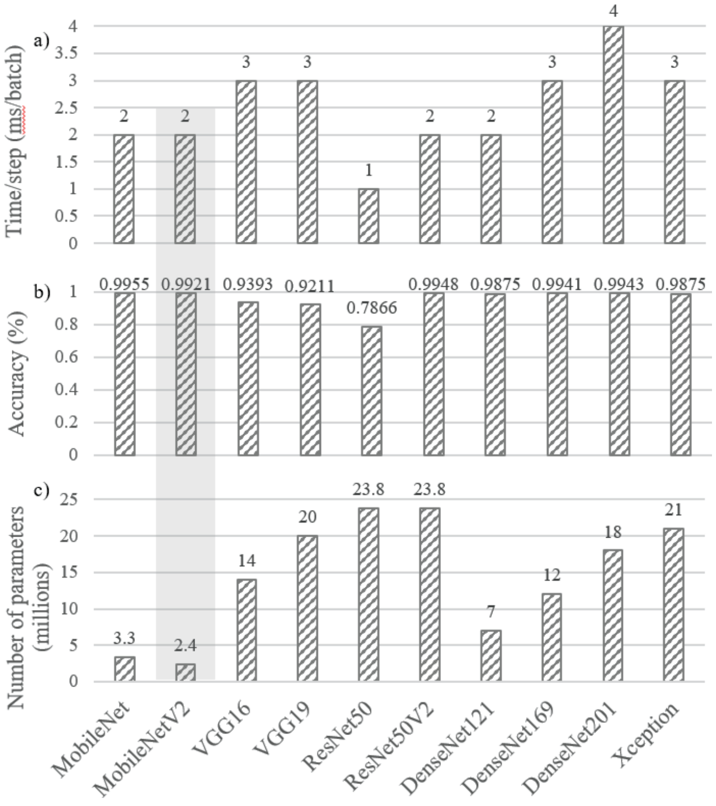

Figure 6: Performance of the five modified networks and their versions: (a) performance of the running time; (b) performance of the accuracy; (c) performance of the number of parameters. The selected network is highlighted in light gray.

Results and discussion

Model selection

Running time significantly impacts the usability of the network: the shorter running time means faster processing of a new image, which is crucial when the number of processed images is large or when a real-time response is needed (e.g., feature identification in live-streaming video). The larger number of network's parameters leads to higher computational cost increasing the running time. Finally, higher accuracy of the network ensures more accurate estimation, in our case more accurate assignment of the labels to the regions. These three metrics were utilised to compare the performance of the five modified networks and their versions as shown in Figure 6.

Comparison of the networks' performance based on the running time (Figure 6a). demonstrated that the VGG16, VGG19, DenseNet169, DenseNet201 and Xception required longer time per batch than other networks, i.e., approximately three seconds or more per batch. While the ResNet50 achieved the accuracy of about 80%, the accuracy of the remaining networks (MobileNet, MobileNetV2, ResNet50V2 and DenseNet121) was about 99% (Figure 6b). Finally, the MobileNetV2 network was selected as it used a smaller number of parameters (about 2.3 million) than the MobileNet, ResNet50V2 and DenseNet121 (Figure 6c).

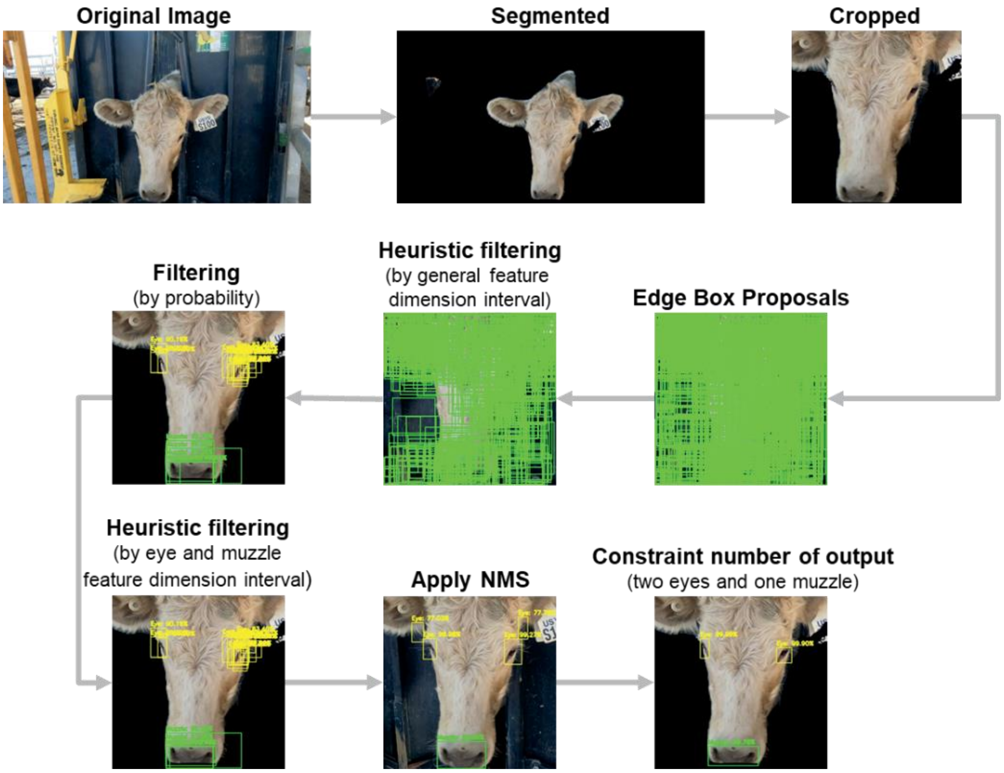

Figure 7: Demonstration of the region identification process

Region identification

Figure 7 presents the steps of the region identification process described in the workflow of eyes and muzzle automated identification (Figure 1) as applied to a new given image of cattle head.

The robust evaluation of the performance of the proposed region identification process requires the availability of reliable ground truth. The regions of eyes and muzzle manually annotated during the pre-processing of the 100 images of the five cattle were employed as the ground truth. For each facial feature in the image, the IOU between the ground truth and the corresponding estimated region was computed. The region estimate for a feature was considered to be correct if it had the correct label (i.e., eye or muzzle) and its IOU with the ground truth region was greater than 50%. The performance evaluation on the 100 images resulted in precision and recall being both 99% for muzzle detection, and 98% and 97.5%, respectively, for eyes detection.

Conclusions

State-of-the-art AI techniques and object detection algorithms present new opportunities to develop more robust and non-invasive capabilities for automated identification of individual cattle. Facial biometrics identification can be a key step towards this. We demonstrated that a two-stage deep learning-based object detection approach can be adopted for an automated identification of eyes and muzzle in an image. To enable this, we generated a database of 5,000 frontal images of cattle heads via manual extraction of frames from the videos captured on-farm. The performance evaluation of the approach resulted in a high precision and recall (98%-99%) in eyes and muzzle detection. While the suggested approach was tested on white Charolais cattle, it can be extended to other breeds, such as black Angus. This can open new possibilities for automatic computation of cattle facial biometrics with potential applications to cattle traceability throughout the livestock supply chain.

Acknowledgments

We acknowledge the financial support of MLA (Meat and Livestock Australia) and the assistance of the group from University of Sydney John Bruce Pye Farm that enabled this work.

References

Awad, A.I. (2016) From classical methods to animal biometrics: A review on cattle identification and tracking. *Computers and Electronics in Agriculture* 123, 423-435.

Baranov, A., Graml, R., Pirchner, F., and Schmid, D. (1993) Breed differences and intra-breed genetic variability of dermatoglyphic pattern of cattle. *Journal of Animal Breeding and Genetics* 110(1-6), 385-392.

Bolya, D., Zhou, C., Xiao, F., and Lee, Y.J. (2019) Yolact: Real-time instance segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision* 9157-9166.

Dutta, A., Gupta, A., and Zissermann, A. (2016) Vgg image annotator (via). *http://www.robots.ox.ac.uk/~vgg/software/via*.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., and Garcia-Rodriguez, J. (2017) A review on deep learning techniques applied to semantic segmentation. *ArXiv preprint arXiv:1704.06857*.

Hosang, J., Benenson, R., and Schiele, B. (2017) Learning non-maximum suppression. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4507-4515.

Kumar, S., and Singh, S.K. (2017) Automatic identification of cattle using muzzle point pattern: A hybrid feature extraction and classification paradigm. *Multimedia Tools and Applications* 76, 26551-26580.

Kumar, S., Tiwari, S., and Singh, S.K. (2016) Face recognition of cattle: Can it be done? *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences* 86, 137-148.

Lohia, A., Kadam, K.D., Joshi, R.R., and Bongale, A.M. (2021) Bibliometric analysis of one-stage and two-stage object detection. *Library Philosophy and Practice* 4910, 34.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018) Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4510- 4520.

Shojaeipour, A., Falzon, G., Kwan, P., Hadavi, N., Cowley, F.C., and Paul, D. (2021) Automated muzzle detection and biometric identification via few-shot deep transfer learning of mixed breed cattle. *Agronomy* 11(11), 2365.

Zhou, P., Feng, J., Ma, C., Xiong, C., and Hoi, S.C.H. (2020) Towards theoretically understanding why sgd generalizes better than adam in deep learning. *Advances in Neural Information Processing Systems* 33, 21285-21296.

Zitnick, C.L., and Dollár, P. (2014) Edge boxes: Locating object proposals from edges. In: *Computer Vision–ECCV 2014: 13th European Conference* Zurich, Switzerland, 391-405.