Emotion Recognition Model Based on

Facial Expressions in Farm Animals

*Index Terms* — emotional recognition; facial recognition; behavior recognition; farm animals;

1. **Introduction**

There is a dire need for a facial recognition platform that can be used to identify individual animals. A facial recognition system can help detect the behavior of dairy cows and pigs. In dairy cows, the identification of individual animals would benefit the recognition of social interactions between the cows through their licking behavior and identifying the emotions and stress. This paper concerns development of a facial recognition software for identifying individual animals such as dairy cows and pigs.

It is a well-known fact that ears are indicators of a cow's emotional states. Relaxed ear postures indicate cows are feeling well. Cows like to play. When cows go through fear or anxiety, whiter in the eye is seen. When cows feel better, less eye white is seen. When cow mothers are separated from calves, more eye white is seen. Cows display a full range of personalities such as boldness, shyness, sociability, gregariousness and being temperamental. Eye white and ear posture are strongly correlated, and that can be used as complementary measures to interpret emotions. Facial recognition technologies are making their way into animal farming. The older methods of animal recognition and identification were tagging and RFID chips that are not tamper-proof and not easy to manage. With the increase in the intelligent algorithms of the facial recognition system, the older methods are becoming more obsolete (Hansen, 2018).

The potential applications of a facial recognition system are enormous and this technology can be employed in the biometric identification of various animals. Monitoring of large animals like cows to smaller organisms like pigs the technology has revolutionary advantages (Neethirajan, 2020). Animals farming is a complex and sensitive process where timely actions can prevent disease outbreaks and disasters. Emerging technologies hold the key to managing the complex farming system. However the questions arise what aspects the technology should look at and what is the accuracy and efficacy of the system and is it worth the attentions in the future or not (Matkowski, 2019).

In animal farming, there are many factors that determine the welfare and growth of the organism. One of the biggest factor is the stress related to the food or from pain, usually caused by living conditions and poor farm management strategies. There is no way humans can effectively recognize the subtle cues of the animal facial change. The algorithms developed over the years have solved the problem by identifying the expression and the behavioral changes of the farm animals. These changes help identify the environments that are causing unnecessary stress. One good example is the proactive identification of respiratory diseases by recording the pigs cough and then subjecting the data to computers (Loeb, 2018). This gives the veterinary doctors to closely work with the farmers to identify the diseases in advance. One of the main advantages of the facial recognition system is its noninvasive nature that reduces the stress factor. Mostly the technology used is remote and consists of cameras and microphones placed at specific spots to record the data. The video camera for the recognition can easily be installed on the food stations to capture the 2D and 3D images. The images thus obtained are subjected to data processing (Singh, 2019). In some cases an accuracy rate of 96% is obtained in sheep farms. Similar results have been reported in case of pigs (Corkery, 2007). Scientists believe that the facial recognition system can also be employed with other biometric systems to increase accuracy and efficacy (Burghardt, 2013).

In all mammals pain is an expression of unease and indicative of an uncomfortable situation. Identifying pain is a critical step in increasing the survival chances of farm animals. The subtle changes in facial expression after a surgery or tail docking have been successfully identified in piglets, which have led to higher survival rates as timely interventions help reduce the pain (Hyde, 2016). This model is essential for the welfare of the animals in big farms where thousands of animals are managed at the same time and manual screening of the animals in real-time is not possible (Giminiani, 2016). The latest deep neural network technologies and machine learning algorithms are getting more powerful in recognizing the pain levels and the difference in the subtle changes in facial expression. To achieve such sophistication the scientists directly integrated the facial cues with different pain level data. They formed a catalog of sheep and pigs pain perception index for various species to ensure the data is accurate (Marsot, 2020).

Noninvasive biometric identification of animals is not reduced to facial expressions or retinal scans. Recent advances have produced algorithms able to identify cows using various body positions as well. The success rate of such technologies when used in an inaccurate and ideal condition is 86%. However, the major hurdles are the changes in external conditions such as lightning, that tend to affect this system (Zin, Phyo, Tin, Hama, & Kobayashi, 2018). As the technology is in its infancy its success rate is not even in all types of organisms. The facial recognition system is less accurate in case of cattle farming where the accuracy rate is 70% as compared to other small animals. Multiple factors have proved to hamper its growth as the first line of surveillance and identification method including the complex facial expression system of higher animals. The software or the algorithms in use also makes a big difference as some programs are more accurate than others but are complex to handle (Wang, 2020). Algorithms like the “bag of Visual words Muzzle” or BoVWM are more accurate with a 97% success rate. Other algorithms like the SURF and MSER are more manageable and easy to understand but have less success rate (Hassaballah, 2019). This is, however, not a major hurdle in the way of facial recognition systems in animal farming. The facial recognition system used in farming are also employed in identification and tracking of pet animals in smart cities (Singh, 2018). The mathematical models that have helped find the lost dogs based on facial recognition technologies, the scientist believe the same systems can be used in the future to help identify the wild animals and the migration pattern of organisms (Brust et al., 2017). In the future, the facial recognition system in animal farming will get more importance and central attention as it ensures real-time and more accurate identification, the welfare of the animals, a concern for many farmers.

Facial biometric identification in animals relies on small cues mostly overlooked by humans. The learning outcomes from the animal farming system has a potential application in future human facial recognition technologies (Mohan, 2018). With the changing learning environment in the world where most of the education has shifted on the internet and distance learning is getting more attention, the facial recognition system could help understand the learner’s emotions in real-time. The state of the mind of the students and teachers would be accurately assessed by looking at the data produced through the algorithms and deep machine learning systems. However, the systems are still learning, and the data set is limited with the systems producing an accuracy of 65% or less in the case of humans (Mukhopadhyay et al., 2020). This does mean that animal facial recognition systems would help us solve many problems related to farming. Humans, however, have a more complex emotion system so an accuracy of 65% in humans means a higher accuracy in animals with the same set of algorithms (Yang, Alsadoon, Prasad, Singh, & Elchouemi, 2017)(Yang et al., 2018). However, more research is still needed because such technologies are still learning (Moreira, 2017).

The overall objective of this paper is to develop an algorithm to automatically determine the emotions of farm animals such as dairy cows and pigs.

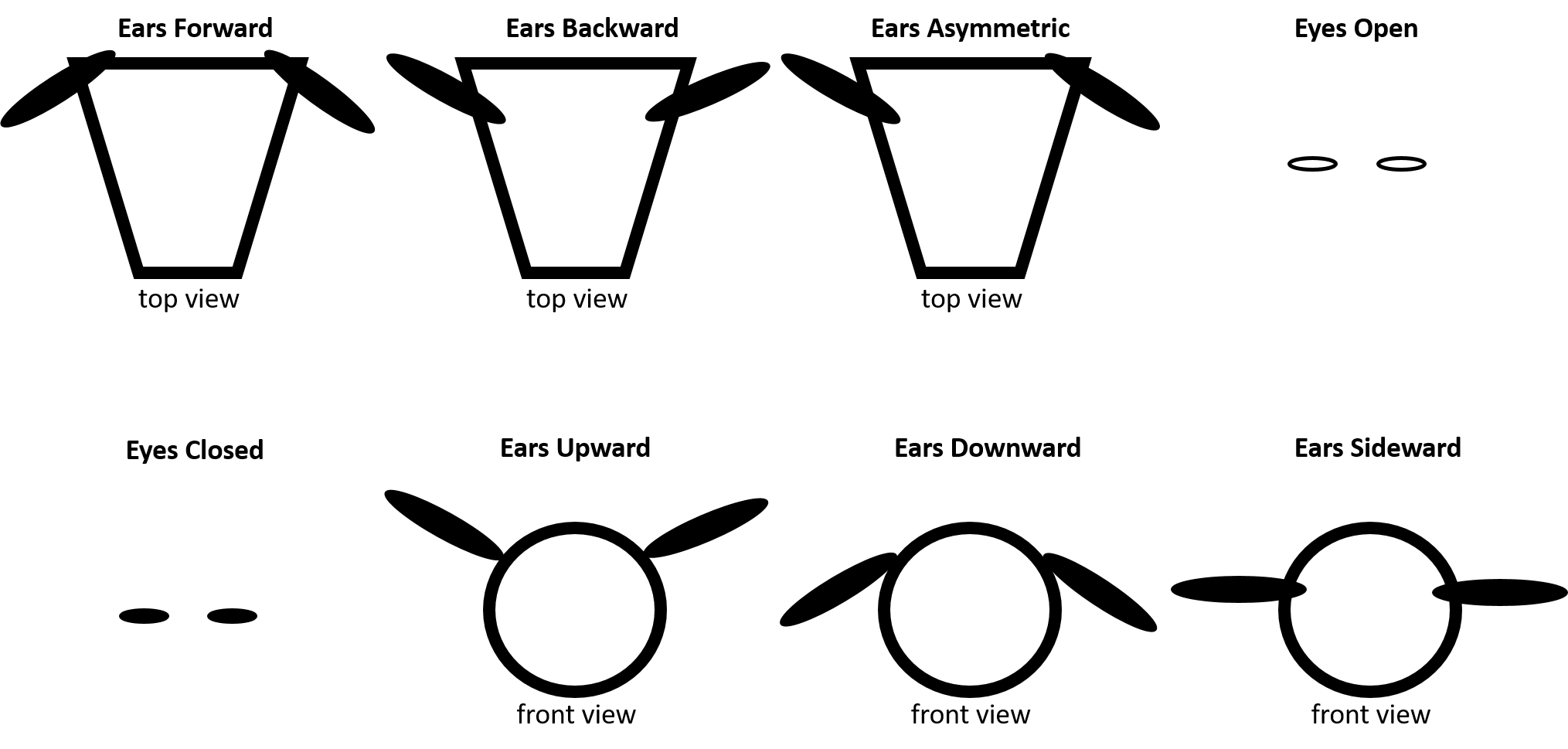
1. **How do Animals Express Emotions**

Ears are indicators of cow's emotional states. Relaxed ear postures indicate cows are feeling well. Cows like to play. When cows go through fear or anxiety, whiter in the eye is seen. When cows feel better, less eye white is seen. When cow mothers are separated from calves, more eye white is seen. Cows display a full range of personalities such as boldness, shyness, sociability, gregariousness and being temperamental. Eye white and ear posture are strongly correlated, and that can be used as complementary measures to interpret emotions. An excited state in a cow is indicated by the ears directed forward and an eye white that is clearly visible. A relaxed state is indicated by half-closed eyes and ears hung down or backwards. Distress in a goat is shown by ears in a forward position, which is indicative of a negative emotion.

1. **Materials and Method**

A machine learning (ML) model was used for identifying ear and eye poses on pigs. Two separate classifiers were trained to identify facial states in cows and pigs. The first classifier identifies the facial state of cows. In this case the classifier outputs two sublabels corresponding to the state of the ears (i.e., up, down) and the state of the eyes (i.e., opened, closed). The second classifier identifies the facial state of pigs. In this case the classifier outputs two sublabels corresponding to the depth of the ear position (i.e., backward, forward, or asymmetric) and the position across the facial plane (i.e., upward, downward, sideward). The different classes are shown in Fig. 1.

Fig. 1. Facial subclasses to be identified



A convolutional neural network (CNN) was used for the classification. A CNN is a deep learning algorithm. CNNs are made of alternated layers of convolution and pooling. Convolution layers convolve an incoming image with a learned kernel while pooling layers reduce the spatial size of the image in order to reduce computations.

The software for the facial classification was developed in Python. It relies on the open source OpenCV (Bradski, 2000), Numpy (Harris, Millman, & van der Walt, 2020), and YOLOv3 (J. Redmon, 2016) libraries. The YOLO library is used to perform the classification. It relies on OpenCV for image processing. YOLO performs object detection because it not only classifies the object but it also outputs its position. YOLO CNNs can be initialized to a set of pretrained weights, which can lead to faster convergence when training the network on a different dataset.

The algorithm can work in real time and is extremely fast because it runs on a single network evaluation. The LabelImg software was used to annotate the images (Tzutalin, 2015).

* 1. Training the classifiers

A total of approximately 500 images were used to train each of the two classifiers (i.e., pig classifier and cow classifier). Only images in which the eyes and ears are clearly visible were used to train the classifier. For the cow facial state classifier, the images were spread evenly across the following five image types: ear posture longer and upright (102 images), cow ears hanging down (100), cow forward facing ear posture (100), half-closed eyes (100), and visible eye white (100). For the pig facial state classifier, the images were spread evenly among the following types: ears asymmetrical (101 images), ears backward (100), ears forward (103), ears sideward (101), and ears upward (100).

LabelImg was used to annotate the images by drawing a box around the regions of interest. When using LabelImg the user annotates the training images the program will output files with class names. Such files are compatible with YOLO so they can be used to train the YOLO algorithm. The images were reduced to a size of 416 x 416 pixels. The images were normalized by subtracting mean and dividing by the maximum intensity (255).

In the implementation, two separate classifiers are developed for each animal. In the case of the cow facial state classifier, for example, the state of the ears is classified separately from the state of the eyes. Each classifier is stored in two files, one with weights, and another one that describe other parameters of the network configuration.

1. **Results**
   1. Testing the classifiers

In the implementation of the testing algorithm, the two files for each classifiers are read (i.e., network weights and network configuration). Each of these two files can alternatively be generated by the following frameworks: Caffe, TensorFlow, Torch, Darknet, or DLDT. As mentioned before, in our case Darknet was used. The test implementation uses only Numpy and the Deep Neural Networks module from OpenCV. YOLO was only used during the training stage. The implementation outputs the identified sublabels, accuracy for each sublabel, and an image in which a bounding box is displayed to verify that face and eyes were correctly located. The implementation of each classifier can be divided into 4 stages: (1) preprocessing, (2) classification, (3) non-max suppression, and (4) bounding box drawer.

The preprocessing stage includes subtracting pixel value means and dividing by standard deviation. This is very often done because some statistical methods respond differently to variables depending on the mean and standard deviation. This ensures that in a classification problem, no matter what the mean, standard deviation, or units of measurement are, the classification results will be the same for two inputs if all other characteristics of the inputs are the same. Not all deep learning architectures, however, substract the mean and scale the inputs. In the proposed method, however, the pixel values are scaled by the maximum intensity (i.e., 256). In addition to the normalization, the image is put into a *blob*. A blob is a collection of 4-dimensional images with the same width, height, channel depth, and preprocessing. In this case, the size of the images that are feed into the CNN classifier is 416X416.

In the classification stage, the preprocessed data is classified using a CNN created with the OpenCV library. The weights and configuration of the trained network are first read from files and used create a deep network object. The blob is then set as the input to the network. After that, inference is run through the network, a process in which the output layer is computed. The computation output is a vector in the YOLOv3 format, which includes following numbers: x-coordinate, y-coordinate, width, height, objectness score, and class scores or convidences. The x and y-coordinates pertain the center of the rectangle for the detected object. The width and height numbers pertain the size of that rectangle. The objectness score represents the probability that an object is contained inside a bounding box. The class scores or confidences are the probabilities for each class. In other words, the class scores are normalized. In YOLOv3, predictions will be made on feature maps of different scales. For a 416X416 images, these scales will be 13X13, 26X26, and 52X52, and they correspond to different layers in the network, and they allow the network to detect objects of different sizes. Those scales correspond to grid sizes. Each cell in a grid predicts 3 bounding boxes that may contain an object, for a total of ((52 x 52) + (26 x 26) + 13 x 13)) x 3 = 10,647 bounding boxes. The large number of boxes detected must then be filtered. The CNN predicts multiple bounding boxes for each object to be detected and class probabilities for each bounding box. A non-max suppression (NMS) algorithm was used to select the best bounding box for each of the detected objects. However, in the proposed implementation all boxes that have a confidence < .5 are ignored. The NMS algorithm selects the box with the highest score for each object and removes other boxes that overlap more than a given threshold with that box. It then moves iteratively to boxes with lower confidence eliminating overlapping boxes that have less confidence at each iteration. After the last iteration, there will be

ADD MORE TO DISCUSSION AND FUTURE WORK FROM WHAT THEY DO IN OTHERS

ADD MORE FROM THE DESCRIPTIONS FOR EACH PROGRAM

ADD MORE FROM THE FILES THAT WERE SENT

Classifiers (1) and (2) first read the weights and configuration of the trained network in order to create a new one with the same parameters. Each CNN classifier is a deep neural network created with the OpenCV library.

Each of the two classifiers was tested with 100 images. The 100 images were split among each of the 5 image types that were used to train each algorithm. An accuracy higher than 95% was obtained for both classifiers. An example image output for each classifier is shown in Fig. 2.

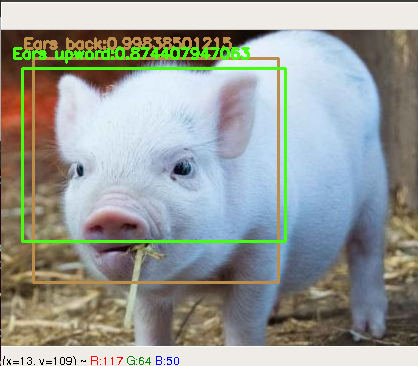
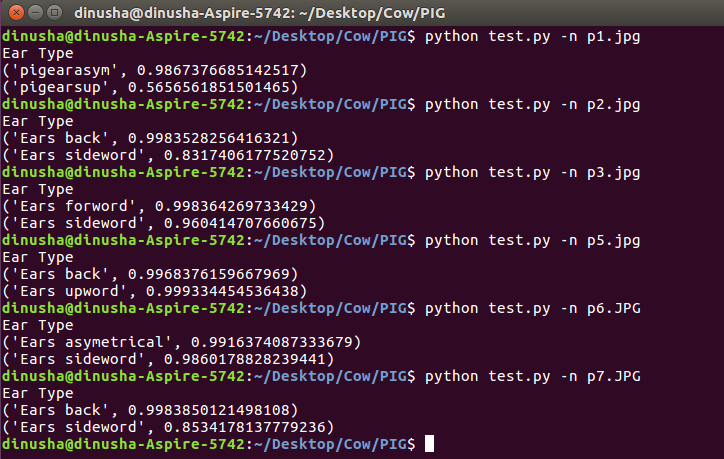
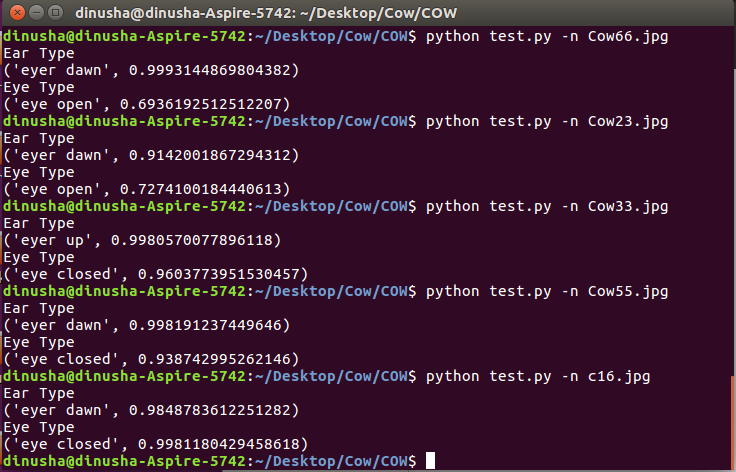


Fig. 2. Close ups for example output images for cow (left) and pig (right) facial state classifiers

The classification software outputs class labels and class probabilities in addition to the labeled images. An example of such output is shown in Fig. 3.

Fig 3. Terminal output for cow (left and back) and pig (front and right) facial state classifiers



1. **Conclusions**

The results from this project show that CNNs are effective in identifying different facial states in cows and pigs. These facial states can be associated with emotional states. Identification of facial states was accomplished by using a divide-and-conquer approach in which each facial state class was associated with the labels for two different subclasses. In the case of the cow, these subclasses correspond to the state of the ears (up vs. down) and the eyes (open vs. closed). In the case of the pig, on the other hand, the subclasses correspond to the depth of the ear position (backward vs. forward vs. asymmetric) and position of the ears on the facial plane (upward, downward, sideward). A possible tracks for future exploration is to determine whether the proposed model translates to images from other animals.

1. **References**

Bradski, G. (2000). The OpenCV Library. Dr. Dobb. *Journal of Software Tools*.

Burghardt, K. &. (2013).

Corkery. (2007).

Giminiani, D. (2016). The assessment of Facial Expressions in Piglets Undergoing Tail Docking and Castration: Toward the Development of the Piglet Grimace Scale. *Frontiers in Veterinary Science*.

Hansen. (2018). Towards on-farm pig face recognition using convolutional neural networks. *Computers in Industry*.

Harris, C., Millman, K., & van der Walt, S. (2020). Array programming with NumPy. *Nature 585*, 357–362.

Hassaballah, A. &. (2019).

Hyde. (2016).

J. Redmon, S. D. (2016). You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 779-788). Las Vegas, NV.

Loeb. (2018).

Marsot. (2020). An adaptive pig face recognition approach using Convolutional Neural Networks. *Computers and Electronics in Agriculture*.

Matkowski. (2019).

Mohan, K. (2018). *Automated Facial Expression Detection Using Machine Learning Algorithm.* School of Computing, National College of Ireland.

Moreira, T. P. (2017). *Where Is My Puppy? Retrieving Lost Dogs by Facial Features.*

Neethirajan. (2020).

Singh, K. &. (2018).

Singh, K. &. (2019).

Tzutalin. (2015). *LabelImg, Git code*. Retrieved from https://github.com/tzutalin/labelImg

Wang. (2020). Cattle Face Recognition Method Based on Parameter Transfer and Deep Learning. *Journal of Physics: Conference Series*.

Yang, D., Alsadoon, A., Prasad, P., Singh, A., & Elchouemi, A. (2017). An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment. *6th International Conference on Smart Computing and Communications.* Kurukshetra, India.

Zin, T. T., Phyo, C. N., Tin, P. T., Hama, H., & Kobayashi, I. (2018). Image Technology based Cow Identification System Using Deep Learning. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2018.* Hong Kong.