Emotion recognition in horses with Convolutional Neural Networks

Luis A. Corujo, Peter Gloor, Emily Kieson

MIT Center for Collective Intelligence

January 2020

Abstract

Creating intelligent systems capable of recognizing emotions is a difficult task, specially if those emotions belong to animals. This paper describes the process of designing such a system in order to recognize emotions in horses. This system is formed by two elements, a detector and a model. The first one, is a faster region-based convolutional neural network that detects horses in an image. The second one, is a convolutional neural network that predicts the emotion of those horses. These two models were trained with multiple images of horses until they achieved high accuracy in their tasks, creating therefore the desired system.

This paper proves that it is possible to predict emotions in animals using autonomous intelligent systems, and it is a good first approach that can be enhanced in many ways in the future to obtain incredible results. This system has multiple applications and can help us better understand how animals feel, if they are sick or if they need any kind of help.

1 Introduction

If you were asked to define what an emotion is, you would probably think that you know what it is and that you would be capable of providing a good definition, however, there is no scientific consensus to date on what exactly an emotion is. According to Wikipedia, an emotion is a mental state associated with the nervous system [1].

So, if an emotion is something so diffuse and complex that we can't agree on what it is, how can we identify and recognize them? Emotion recognition is the process of identifying emotions [2]. This is a complicated process that all humans do many times per day every time we encounter someone. Every time we see a person, we can tell based on facial cues, voice tone, posture and other hints how that person is feeling. However, our prediction might be wrong, a person can be smiling appearing to be happy and in reality be sad. This creates two ground truths. The first one being the collective truth that observers can perceive from the subject, and the second one being how the subject really feels, which some times might be even difficult to recognize by the subject himself.

With the improvements in technology during the last few decades, researchers have studied multiple ways of how to recognize emotions using different techniques such as Markov models, artificial neural networks and Bayesian networks among others.

The automation of emotion recognition has three main approaches. Knowledge-based techniques that predict emotions based on semantic and syntactic knowledge, statistical methods which commonly are machine learning algorithms that predict emotions given a big enough data set, and hybrid approaches, which are a combination of the previous two. Many companies have been successful in predicting emotions on humans using these approaches, a good example is Affectiva [3], a company born at MIT that uses facial and voice cues to predict emotions.

So far you might have been thinking about emotions in humans, however, animals experience emotions in a similar way. Since Charles Darwin, who was one of the first to write about this topic, many other scientists have done research on the topic. Jaak Panksepp, who is considered one of the pioneers in affective neuroscience, identified seven different emotions that animals can feel [4]. Even though a lot of research has been conducted, due to the complex idea of what emotions are and the difficulty of measuring them on animals in a rigorous manner, this is still a controversial topic. It is hard to say what emotions animals have, if they all have the same emotions and if they are all capable of having emotions. Most of the research has been done in mammals because of their similarity to humans and because many of them produce facial expressions that bear clear resemblance to the expressions seen in humans [5].

So the questions arises, can we create an intelligent system with the ability to predict emotions on animals based on facial expressions?

2 Related Work

As We have mentioned earlier, most of the research conducted on animal emotions has been done in mammals. A vast majority, has been conducted from a psychological and neuroscience perspective [5], [6].

In addition, most of the research focuses on the ability of animals to interpret our emotions and not on how we can interpret theirs. Within the group of those that focus on interpreting the expressions of the animals, most of them do it without using an intelligent or autonomous system, a good example is AnimalFacs, a tool for identifying facial movements in non human species. They have multiple papers that explain how we, as humans, can analyze the facial expressions of different primates, dogs, cats and horses [5] ,[7], [8].

There is little research done on how to interpret animal emotions using an autonomous or intelligent system, one example is the work done by Laura Niklas and Kim Ferres in predicting dog emotions from images [9]. Another example is the work done at the University of Augsburg on recognizing dog emotions from bark sequences using an autonomous model [10]. Thus, the idea of predicting animal emotions this way is something that has not been explored in depth.

3 Hypothesis

This paper explores the possibility of creating an intelligent system capable of predicting the emotion of a horse based on its face and neck traits. In order to do this, two different elements must be created.

First, we need to develop a detector capable of recognizing a horse in an image, more specifically, it needs to detect a region of interest (ROI from now on), which in this case, will be the head and neck of a horse. The task of the detector is extremely important, if the detector doesn't recognize the appropriate ROI, the second part of the system won't be able to predict the emotion correctly regardless of how well it can accomplish that task.

Secondly, we need a model that, once it receives the detected ROI, is capable of predicting the emotion of the horse. This means this model must be able to detect facial cues and different neck positions and relate them to the appropriate emotion.

Altogether, this should be an intelligent system capable of detecting a horse, analyzing its face and neck features, and predicting its emotion with reasonable accuracy.

4 Method

4.1 First Steps

The first thing needed was to define the different emotions. Four emotions were defined with respect to neck and facial cues:

Alarmed

- Eyes: open eyes with little or no sclera
- Ears: stiffly forward
- Nose: open nostrils, usually slightly tense mouth or muzzle
- Neck: above parallel, head higher than back

Annoyed

- Eyes: open with perhaps some sclera
- Ears: stiffly back or pinned back, close to the horse's head
- Nose: nostrils slightly closed, tense mouth or muzzle
- Neck: usually parallel or above parallel

Curious

- Eyes: open with little or no sclera
- Ears: pointing forward/sides but relaxed
- Nose: open nostrils, relaxed mouth and muzzle
- Neck: usually parallel to ground but may be slightly below or above

Relaxed

- Eyes: partially to mostly shut
- Ears: relaxed, opening pointing to the sides
- Nose: relaxed mouth and muzzle
- Neck: approximately parallel or below

Once the emotions were defined, the second thing to do was to collect the data, in this case, a total of 480 images of horses were collected, 120 images per emotion. This images were split in training and validation sets in order to train the detector and the model.

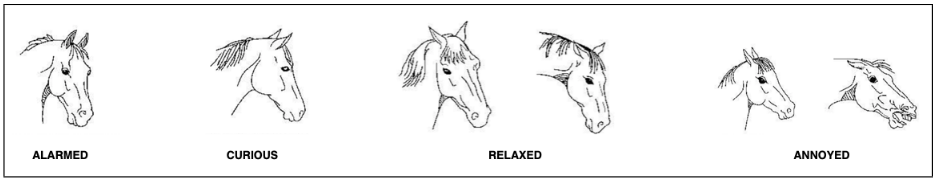

Figure 1: Horse emotions

4.2 Detector

To train the detector, 400 images were used as training data while the 80 left were used as validation data. Every picture was rescaled to 200 pixels in height keeping the original ratio, this was done to facilitate the work of the detector, since having less pixeles requires less computations, and there is not a significant loss of information in the process of rescaling to this size. All images were labeled. In this case, labeling the images meant to manually highlight the ROI that the detector was supposed to find.

Once the images were ready for the detector, its architecture was chosen. The architecture used was the faster region-based convolutional neural network (faster R-CNN) [11], one of the most famous architectures for object detection. This architecture is composed of three different parts. First, the convolutional layers, which filter the images in order to extract useful features, secondly, the RPN (Region Proposal Network), whose duty is to identify the possible regions where objects (in this case horses) can be located, and finally a dense neural network that predicts what kind of object is in each proposed region (in our case whether there is a horse in each proposed region or not).

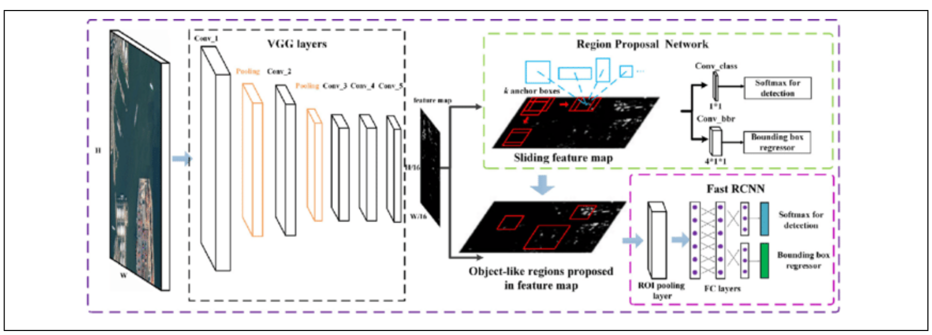

Figure 2: Faster R-CNN architecture

During the first epochs of training the detector had difficulties detecting any ROIs, but as training progressed it began making more accurate detections. After 4000 epochs, it was capable of finding the region of interest with high precision.

4.3 Model

Next, a model to predict the emotions was created. This model would receive images of 150x150 pixels (the rescaled ROI found by the detector), and output predictions (Alarmed, Annoyed, Curious or Relaxed). Once again, the rescaling of the images to a smaller size is a question of making the work of the model easier.

The architecture of this model was formed by a convolutional base, a flatten layer, two fully connected layers (256 and 128 nodes respectively) and a softmax layer (4 nodes, 1 for each emotion). Three different convolutional bases were tested, the base from ResNet50v2, Xception and VGG16, all with weights from imagenet. These three showed accuracies higher than 50%, but the VGG16 base was the best and therefore the one that was chosen. To train the model, only the last convolutional block of layers and the layers on top of this one were trained. They were trained for 25 epochs using 400 pictures (100 for each category) as the data set, and 80 pictures (20 for each category) as the validation set. Training the first layers of the base didn't make sense in this case, the reason was that these layers learn common patterns that are present in all images, such as corners or straight lines, and since this base was trained using the 14 million images of imagenet, achieving a better performance with only 400 images was very unlikely.

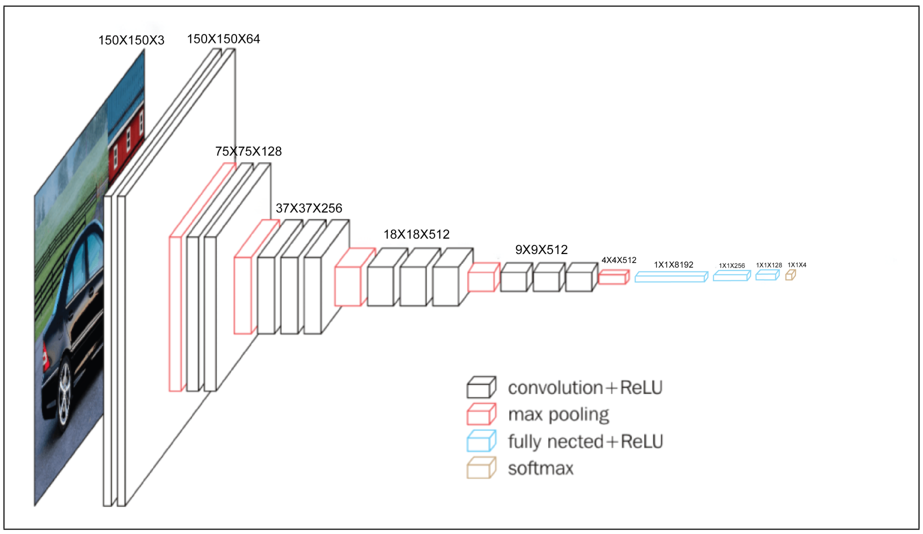

Figure 3: Model architechture

4.4 Final steps

Lastly, in order to facilitate the use of this system, a desktop graphical user interface was created. The GUI allows any user to upload an image, processes this image, and displays the ROI with a rectangle (which should be the head and neck of a horse) and the predicted emotion.

In a nutshell, the system created consists of two well distinguished parts, a detector and a model. The detector receives an image previously rescaled to 200 pixeles in height, and outputs a ROI, a region of the image with a horse in it. This ROI is rescaled to 150x150 pixels and is passed to the model, which predicts and outputs the final emotion. A diagram of the entire process can be seen in the figure below.

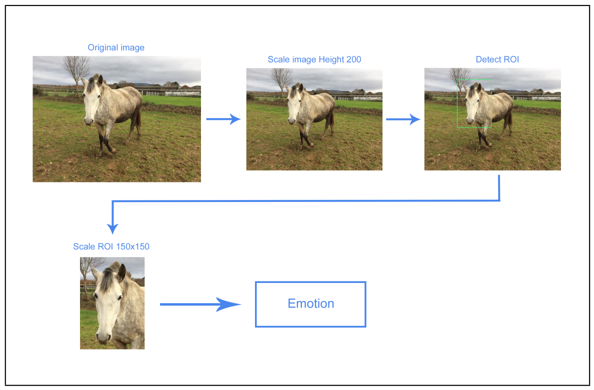

Figure 4: System process diagram

5 Results

The work resulted in a detector capable of finding a horse's face in an image, a model capable of predicting the emotion of a horse given a picture of its head, and a GUI that is user friendly.

The detector's precision is very high, it labeled all the 20 test images without a single error, some of these detections are shown in figure 5.

Figure 5: Detector predictions

The model's 10-fold cross-validation average accuracy (both for the training and validation set) can be seen in figure 6 along with the results of the models trained with the ResNet and Xception convolutional bases. The model achieved an accuracy between 63% and 68%.

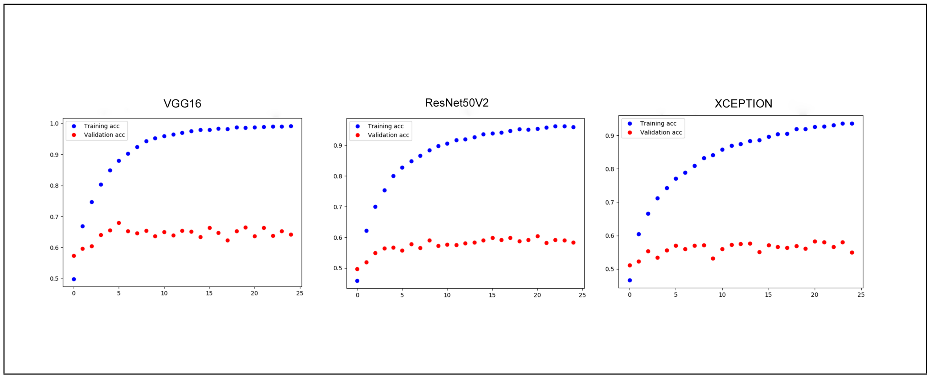

Figure 6: 10-fold cross-validation average accuracy over 25 epochs

Finally, the entire system is brought together by the desktop GUI, which includes the two previous results and makes them easy for anyone to use. This makes it easy for a user to upload an image and see the ROI and the emotion detected. The graphical user interface is shown in figure 7.

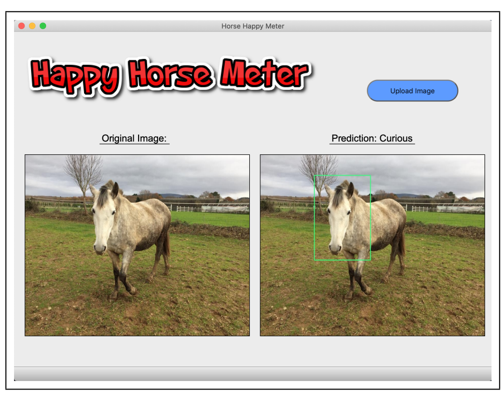

Figure 7: Graphical user interface

6 Discussion

In the introduction section, the question of whether it was possible to create an intelligent system capable of detecting emotions on animals was asked. This paper shows how such a system was created. This system is not only capable of predicting emotions on horses, but it does so in an autonomous way and with good results. This proves that we can recognize emotions on all animals that have the ability to produce facial expressions, and that it might be possible to detect these emotions by other methods such as measuring the animal's heart rate, its temperature, or recording the sounds that they produce and feeding all this data into a system similar to the one created here. While the system works with reasonable accuracy, it is worth pointing out that it could be improved in many ways, this is just a first approach.

First of all, we have to keep in mind that predicting emotions is a complex task, it is hard in humans and it is even harder in animals.

Secondly, there were only 480 images in total, which is not a large number for systems like this one. Having more images, thousands or millions of them split evenly between every emotion would have made the model have a higher accuracy.

Thirdly, only the head and neck of the horse were used to predict its emotion, analyzing cues from the entire body would probably yield better results, however, since we would have more features to analyze, more data would be needed. In addition, if the entire body is used, the emotions have to be reviewed, since they were defined only by cues found in the head and neck of the horse.

Finally, another way to improve the results obtained in this paper would be to use information from other sources in combination with the images. Sensors to measure heart rate and temperature could be put on horses, their sound could be recorded, etc.

In conclusion, this system is able of predicting emotions in horses and it is a great foundation on which to build new and better systems to predict animal emotions.

References

[1] Wikipedia, "Emotion." https://en.wikipedia.org/wiki/Emotion.

[2] Wikipedia, "Emotion recognition." https://en.wikipedia.org/wiki/Emotionrecognition.

[3] Wikipedia, "Affectiva." https://en.wikipedia.org/wiki/Affectiva.

[4] J. Panksepp, "Affective consciousness: Core emotional feelings in animals and humans," *Consciousness and cognition*, vol. 14, no. 1, pp. 30–80, 2005.

[5] B. M. Waller, K. Peirce, C. C. Caeiro, L. Scheider, A. M. Burrows, S. McCune, and J. Kaminski, "Paedomorphic facial expressions give dogs a selective advantage," *PLoS one*, vol. 8, no. 12, 2013.

[6] P. Martens, M.-J. Enders-Slegers, and J. K. Walker, "The emotional lives of companion animals: Attachment and subjective claims by owners of cats and dogs," *Anthrozoös*, vol. 29, no. 1, pp. 73–88, 2016.

[7] J. Wathan, A. M. Burrows, B. M. Waller, and K. McComb, "Equifacs: the equine facial action coding system," *PLoS one*, vol. 10, no. 8, 2015.

[8] S.-J. Vick, B. M. Waller, L. A. Parr, M. C. S. Pasqualini, and K. A. Bard, "A cross-species comparison of facial morphology and movement in humans and chimpanzees using the facial action coding system (facs)," *Journal of nonverbal behavior*, vol. 31, no. 1, pp. 1–20, 2007.

[9] "Dog-emotions." https://github.com/cangaroo6/Dog-Emotions, 2019.

[10] S. Hantke, N. Cummins, and B. Schuller, "What is my dog trying to tell me? the automatic recognition of the context and perceived emotion of dog barks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5134–5138, IEEE, 2018.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems 28* (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds.), pp. 91–99, Curran Associates, Inc., 2015.