
**ST. JOSEPH'S COLLEGE (AUTONOMOUS)**

**#36, LALBAGH ROAD, BENGALURU 560027**

**ST. JOSEPH'S INSTITUTE OF INFORMATION TECHNOLOGY**

**DEPARTMENT OF ADVANCED COMPUTING**

**PROJECT REPORT**

**ON**

# PREDICTING DOG EMOTIONS

# BY

# MACHINE LEARNING

**TEAM MEMBERS**

GEETHU SUNNY-21BDA03

CHRISTINE BIJU JACOB-21BDA06

ANGEL PHILIP-21BDA07

MURREL MIRANDA-21BDA36

TONY TOMY-21BDA44

AKSHAY S-21BDA62

# ABSTRACT

Nowadays, there are plenty of machine learning projects focused on image recognition and caption generation. Great chunk of it is about recognizing human face and naming emotion it express. We decided to build image recognition with emotion capture, not for human faces though, but for dog's muzzles, as there is no good model to do that yet and far more fun. Our application is able to determine a dog's emotional state visually with an accuracy between 56% and 60%, exceeding human capability to recognize dog emotions.

# INTRODUCTION

Dogs play an important social role in western society and can assist with various tasks. Service dogs assist visually impaired people, patients fighting anxiety issues, and mentally or physically disabled people. Law enforcement uses dogs in several fields, for example, officer protection, track search, or drug inspection. The field of medicine also benefits from dogs since their superior sense of smell can perceive illnesses such as Parkinson's disease.

Along with many other possible uses, all share the need for extensive training adapted to the dog. Even for a pet dog, training and education are usually required to live along with people. In order to train dogs optimally, it is relevant to possess the ability to read and interpret their body language correctly. Thus, assessing a dog's emotions is essential for resource-efficient, successful training.

Computer-aided emotion recognition ensures dogs' well-being across distances or numerous animals simultaneously without high manual effort. Furthermore, dog emotion detection can assist users who have little experience in dealing with dogs to recognize angry or aggressive dogs early, and thus evade a potential attack. Another application area would be veterinary medicine, which could initiate diagnostics earlier if a dog has an increased perception of dissatisfaction.

# DATA PREPARATION

- DOGS EMOTIONS RECOGNITION DATASET:
  The dataset used for the project was found in Kaggle with the name "doggos_emotion_recognition", which contains around 2000 images belonging to four classes ie, train, test and validation, containg four emotions :-
  • happy dogs - dogs with the smile, mainly with open muzzle and tongue out
  • angry dogs - dogs that looks scary, with their teeth showing
  • sleepy dogs - dogs that are taking a nap, or are about to, with closed or squinted eyes,
  • good dogs - dogs that looks polite, with neutral muzzle
- Seperating into train, validation and test sets. Lastly, as our dataset

was small, we proceed to separate it using 60/20/20 proportion:

- 60% - 1200 (300 per class) images in training set
- 20% - 400 (100 per class) images in validation set
- 20% - 400 (100 per class) images in test set

# TASK DESCRIPTION

Our task consists of two major parts:

- Prepration:
  - create a dataset of dogs images
  - tag images with proper emotions
- AI implementation:
  - find doggo on the image (detect the doggo's muzzle)
  - recognise the emotion (from a set of possible emotions, assess what it express on the image)

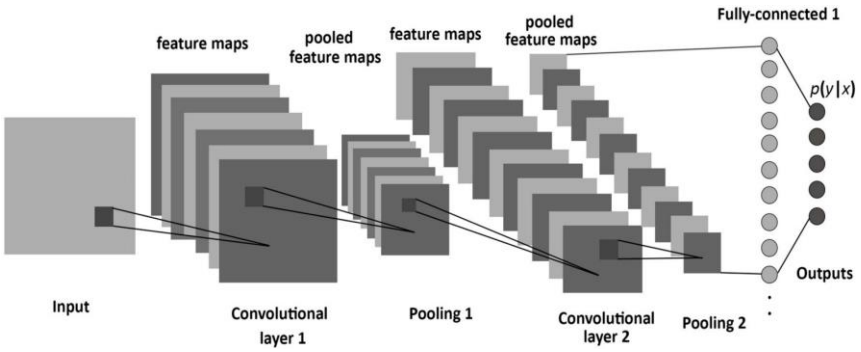
# COMMON APPROACHES

Computer vision is rapidly growing interdisciplinary field, that deals with computers gaining understanding from photos. In our research, we focused on two topics, mainly: how human emotion recognition from pictures is done and how can it be applied to captioning doggo's emotion (is it possible).

In this type of tasks (working with pictures, recognizing what's on them) neural networks are the most useful. They are ones of the most popular machine learning algorithms at present and are proven to outperform other algorithms in accuracy and speed. In recent scientific papers, two types of neural networks seem to be the most popular:

- CNN (Convolutional Neutral Network) - it derives its name from the type

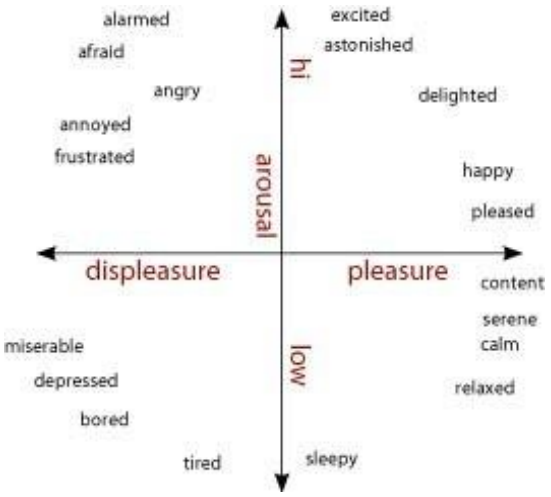
of hidden layers it consists of - convolutional layers, pooling layers, fully connected layers, and normalization layers. An input is processed through them and an output is produced, with an assumption that two successive inputs are independent of each other. Convolution and pooling functions are used as activation functions (functions that are applied to an input of the neuron) (example in usage: Deep Face)




- RNN (Recursive Neural Network) - this variation of the network allows two successive inputs to be dependent of each other - as it happened very often in real life. It performs the same task for every element of a sequence, with the output being depended on the previous computations. Another way to think about RNNs is that they have a "memory" which captures information about what has been calculated so far.

# OUR SOLUTION

We are yet to decide whether we want to just named the emotions or use arousal - valence scale suited for doggo's emotions (so each photo will be tagged with two values x, y representing score on the according axis).

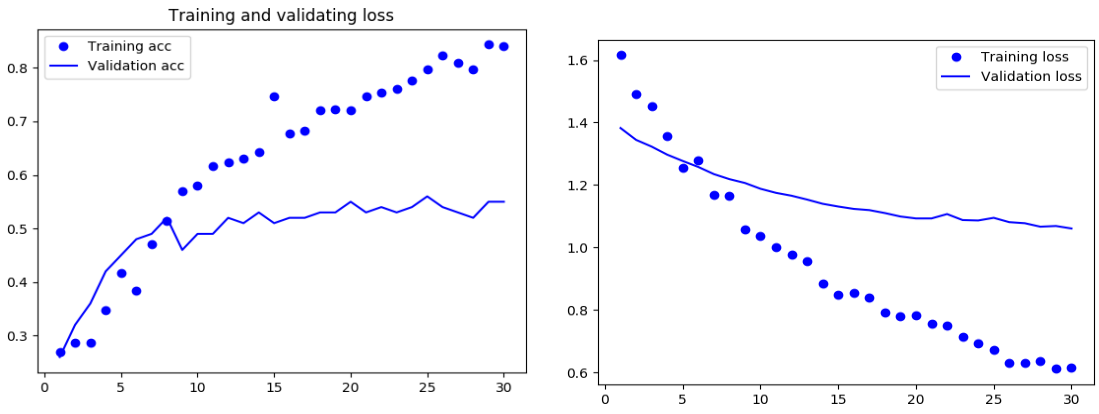



Our choice of algorithm is now settled on Convolution Neural Network, for a couple of reasons: it is better suited for our's type of task (CNN can be stacked into a very deep model, which has been proven to be very effective in capturing features. On the contrary, due to the gradient vanishing and exploding problems, RNN models are usually very shallow, 2 or 3 layers), we had experience with building CNN from another project, we can head start our project by taking pre-trained model, such as deep fake and cutting last layers of CNN (as they are the most detailed) to check whether learing emotion recognition in humans can be used to recognize doggo's emotions.

We will procede to represent out images as matrixes of pixels with 3 values for color (RGB) - enables our later algorithm to read them and learn from them. To train our model, we will use the algorithm that consists of convolutional layers that will detect patters. It will be done by matrix multiplication of the part of an image matrix with filter, then applying activating function - ReLu, polling layers that will generalize and minimize the size of the image (matrix) by taking maximum value of the pixel or the average of the chunk. In the end, fully connected layer will classify processed matrixes with the output being the name of the emotion or values for valence-arousal scale (which can be easily interpreted with names of emotions). After classification, we can compute loss function and bearing in mind those values adjust the filters and improve our algorithm.

# IMPLEMENTATION

- pure CNN We started with the simplest neural network we could think of, one consisting of four layers - every one used relu as its activation function - and one with softmax at the end. We went for 30 epochs and achieved 30% accuracy on validation data. Despite tremendous overitting we were off to a good start.
- CNN with data augmentation Overfitting was due to small size of our dataset, but it turned out that keras has perfect solution for it - easy to implement data augumentation - process that from one sample creates more by subjecting ir to various transformations. This time we decided on 100 epochs, so after some time (even using GPU) we got 56% accuracy with overfitting at smaller scale.
- CNN with convolutional base taken from 'imagenet' To counter overfitting resulting from using small dataset, we decided to use transfer learning and introduce convolutional base, pre-trained on famous imagenet dataset. Imagenet includes pictures of all sorts of animals, including various breeds of dogs, so we cocluded it would be suitable for our needs. It performed much better than pure version, with 60% accuracy and similar to the version with data augmentation.




(a) Accuracy function (b) Loss function

Figure 1: Accuracy and loss functions for CNN with 'imagenet'

# CONCLUSION

The project shows promising results for machine learning dog emotion recognition, especially regarding the small size of the training data set. Multiple models were developed to determine a dog's emotional state based on input images with an accuracy between 56% and 60%.

Image-based emotion recognition promises wide everyday application for automatic emotion analysis of animals. The neural network seems to be better suited for practical applications since it can handle missing landmark points and therefore provides results even for low-quality input images. This work focuses mainly on posture and less on facial expression.

# REFERENCE


1. Hasegawa, M.; Ohtani, N.; Ohta, M. Dogs' Body Language Relevant to Learning Achievement. *Animals* **2014**, *4*, 45–58.
2. Amici, F.; Waterman, J.; Kellermann, C.; Karim, K.; Bräuer, J. The Ability to Recognize Dog Emotions Depends on the Cultural Milieu in Which We Grow Up. *Sci. Rep.* **2019**, *9*, 1–9.
3. Waller, B.; Peirce, K.; Correia-Caeiro, C.; Oña, L.; Burrows, A.; Mccune, S.; Kaminski, J. Paedomorphic Facial Expressions Give Dogs a Selective Advantage. *PLoS ONE* **2013**, *8*, e82686.
4. Franzoni, V.; Milani, A.; Biondi, G.; Micheli, F. A Preliminary Work on Dog Emotion Recognition. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence—Companion Volume, New York, NY, USA, 14–17 October 2019; pp. 91–96.