

Machine Learning Engineer Nanodegree

Capstone Proposal

Shvet Chakra December 3rd, 2020

Proposal

Emotion recognition using facial expressions

Domain Background

Facial expressions and related changes in facial patterns give us information about the emotional state of the person and help to regulate conversations with the person. Moreover, these expressions help in understanding the overall mood of the person in a better way. Facial expressions play an important role in human interactions and non-verbal communication. Classification of facial expressions could be used as an effective tool in behavioural studies and in medical rehabilitation. Facial expression analysis deals with visually recognizing and analyzing different facial motions and facial feature changes.[1]

Motivation Nowadays, when technology has penetrated into the personal life of humans with mobile devices and interaction of machines with humans is carried out on the daily basis, it is getting important for the machine to recognize the emotion of the fellow human while interacting with him. Recognition of emotion can help machines powered with artificial intelligence to enhance this interaction. Possible application where this project can be helpful in wellbeing of humans. As recognition & acceptance is the first step of any problem, this solution can help the user about their mood and help them in taking appropriate step. For eg. we have facial recognition phone security feature to lock/unlock the device, if the device captures the face and determine the emotion of the users on several intervals and then give a happy mood percentage for the day. It can help them in recognizing about their mood if it is not appropriate and motivate them to take necessary action.

Problem Statement

Facial expressions play an important role in recognition of emotions and are used in the process of non-verbal communication, as well as to identify people. They are very important in daily emotional communication, just next to the tone of voice [2]. They are also an indicator of feelings, allowing a man to express an emotional state [3]. People, can immediately recognize an emotional state of a person. As a consequence, information on the facial expressions are often used in automatic systems of emotion recognition [4]. The aim of the project is to recognize seven basic emotional states: neutral, joy, surprise, anger, sadness, fear and disgust based on facial expressions.[5] Numerous investigators have used neural networks for facial expression classification. The performance of a neural network depends on several factors including the initial random weights, the training data, the activation function used, and the structure of the network including the number of hidden layer neurons, etc. Here I will try to create a CNN based neural network model to classify the images into their seven basic emotional states: neutral, joy, surprise, anger, sadness, fear and disgust.

Datasets and Inputs

The project has been chosen from a kaggle competition and hence the same dataset will be used here also to train

and test the model for the proposed problem. <https://www.kaggle.com/c/emotion-detection-from-facial-expressions/data>

Solution Statement

Convolution networks or convolution neural networks are a specialized kind of neural networks for processing data that has a known grid-like topology. Examples included time-series data and image data which has a grid like structure. Convolution leverages three important ideas that can help improve a machine learning system; sparse interctions, parameter sharing and equivariant representations. Moreover, convolution provides a means for working with inputs of variable sizes. One major advantage of using CNNs over NNs is that you do not need to flatten the input images to 1D as they are capable of working with image data in 2D. This helps in retaining the "spatial" properties of images.

Benchmark Model

For benchmarking we can use Alexnet or VGG models to perform some benchmarking. <https://www.imperial.ac.uk/intelligent-digital-systems/cnn-benchmark-suite/> Our model can also be compared for categorization accuracy as per kaggle competition leaderboard <https://www.kaggle.com/c/emotion-detection-from-facial-expressions/leaderboard>

Evaluation Metrics

We can use following metrics to evaluate our model

1. Classification Accuracy
2. Confusion Matrix
3. F1 Score

Project Design

Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernels), Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1.

The project can be mainly clasified into following steps:

1) Load and preprocess the image dataset:

The dataset will be analyzed for the image attributes and the list of emotions labeled to them. Conversion of labels to lowercase and merge few labels into one may also be explored if required. Then we will divide the dataset into three sets, train set, validation set , and test set.

For the training the model, we will apply some general image transformations such as random scaling, cropping, and flipping.

The input data is center cropped to or resized to 224x224 pixels as required by the pre-trained networks for benchmarking.

Next no scaling or rotation transformations will be performed on validation set and test set but need to resize then crop the images to the appropriate size to measure the model's performance on data it hasn't seen yet.

2) Building and training the classifier

Now that the data is ready, it's time to build and train the classifier. As usual, we should use one of the pretrained models from torchvision.models to get the image features. Build and train a new feed-forward classifier using those features.

2.1) Load a pre-trained network

2.2) Define a new, untrained feed-forward network as a classifier, using ReLU activations and dropout

2.3) Train the classifier layers using backpropagation using the pre-trained network to get the features

2.4) Track the loss and accuracy on the validation set to determine the best hyperparameters

3) Evaluate the model on the test set and evaluation metrics

We can use following metrics to evaluate our model

1. Classification Accuracy
2. Confusion Matrix
3. F1 Score

References:

- [1] <https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/1475-925X-8-16> [2] Ratliff M. S., Patterson E., Emotion recognition using facial expressions with active appearance models, Proceedings of the Third IASTED International Conference on Human Computer Interaction, ACTA Press, Anaheim, CA, USA, 2008, 138–143. [3] Tian Y. I., Kanade T., Cohn J. F., Recognizing action units for facial expression analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (2001), no. 2, 97–115. [4] Mao Q., Pan X., Zhan Y., Shen X., Using Kinect for real-time emotion recognition via facial expressions, Frontiers Inf Technol Electronic Eng, 16 (2015), no. 4, 272–282. [5] International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland Emotion recognition using facial expressions, Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak