

Facial Expression Recognition (Deep Learning)

A PROJECT REPORT

Submitted by

Dilip Kumar (17114026), Balwant Singh (17114018),

Akhilesh Kumar (17114007)

for the fulfillment

of

CSN-400B: B.Tech. Project

Under the guidance of

Prof. R. Balasubramanian



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

ROORKEE-247667

25 May 2021

A. Declaration and Certificate

B. Acknowledgement

C. Abstract

D. Table of Content

- 1) Chapter-1: Introduction**
 - a. Objective**
 - b. Motivation**
 - c. Problem Statement**
- 2) Chapter-2: Literature Survey**
- 3) Chapter-3: Methodology/Experimental Setup**
- 4) Chapter-4: Results**
- 5) Chapter-5: Conclusion and Future Work**
- 6) References**

CANDIDATE'S DECLARATION

We hereby declare that the work carried out in the B.Tech. Project titled **Facial Expression Recognition (Deep Learning)** is presented for the partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in the Computer Science and Engineering and submitted to the Department of Computer Science and Engineering, Indian Institute of Technology Roorkee under the supervision of Prof. R. Balasubramanian, Department of CSE, IIT Roorkee.

The work presented in this report is an authentic record of our work carried out during the period from August 2019 to June 2020. The content of this report has not been submitted by us for the award of any other degree of this or any other institute.

Date: 25 May 2021

Signature: 

Place: IIT Roorkee

Name & Enrollment No.: Dilip Kumar (17114026)

Date: 25 May 2021

Signature: 

Place: IIT Roorkee

Name & Enrollment No.: Balwant Singh (17114018)

Date: 25 May 2021

Signature: 


Place: IIT Roorkee

Name & Enrollment No.: Akhilesh Kumar (17114007)

CERTIFICATE

This is to certify that the statement made by the candidate is correct to the best of my knowledge and belief.

Date: 04 June 2021

Signature: 

Place: IIT Roorkee

Prof. R. Balasubramanian

Professor, Department of CSE, IIT Roorkee

Acknowledgement

We are very grateful to Professor R. Balasubramanian for their support to the project.

Abstract

Automatically performing facial expressions recognition through computer vision will be really important for machines and humans to interact with each other and has many other applications in medical science and computer surveillance. After studying some recent papers on the topic we used Convolutional Neural Networks to extract valid features from the input dataset. We also used transfer learning and data augmentation. By using these methods from recent papers, we achieved an accuracy of 69.02% on the test subset of the Fer2013 dataset.

Introduction

Facial expressions recognition is very important to humans in terms of nonverbal communication and it has been well studied in the past[1].

For human to human conversation and interaction facial emotions is a really important thing, automatically performing facial expressions recognition through computer vision will be really important for machines and humans to interact with each other. Facial expressions recognition is a well studied field and it is considered solved [1] in controlled conditions like in front facing images with good resolution. It has lots of uses in real life like personalised technology, medical science and many other systems that can make use of facial expression recognition.

Human facial expressions have been mostly divided into seven types: angry, disgust, fear, happy, neutral, sad and surprise. Facial expressions recognition in controlled conditions is a solved problem but doing the same in natural conditions is a very challenging problem, because of different lighting conditions, different head poses and side faces. Reliable Facial expressions recognition in natural conditions is a very useful technique but still it hasn't been solved [1], [2].

Motivation

Facial expressions recognition has lots of uses in real life like personalised technology, medical science and many other systems that can make use of facial expression recognition. Through Facial expression recognition machines can better understand human emotions. but facial expression recognition in natural conditions is considered to be a very challenging problem.

Facial expression Recognition wasn't a very widely researched problem in the past. However, recent progress in computer vision technologies like object detection, pattern detection have motivated a lot of researchers to work in this field.

Problem Statement

Facial expression recognition is considered a well studied problem with lots of publicly available datasets. For our project we used the [FER2013](#) dataset. We chose this dataset because it is considered to be one of the hardest dataset with human level accuracy of only $65\pm 5\%$ and highest achieved accuracy of 75.1% [8] on a single model. It is publicly available on Kaggle as a part of a competition. It was created to promote research facial recognition systems in natural conditions. This dataset has 35887 grayscale images of size 48x48 with one of the seven emotions as label. We train CNN models on this dataset and take seven emotions as labels.



This dataset have three subsets

- 1) **Training:** 28709 images, Used for Model training dataset
- 2) **PublicTest:** 3589 images, Used for Model Validation dataset
- 3) **PrivateTest:** 3589 images, Used for Model Testing

Seven types of Emotions(labels) in the dataset

Angry: label 0 and have 4593 grayscale images

Disgust: label 1 and have 547 grayscale images

Fear: label 2 and have 5121 grayscale images

Happy: label 3 and have 8989 grayscale images

Sad: label 4 and have 6077 grayscale images

Surprise: label 5 and have 4002 grayscale images

Neutral: label 6 and have 6198 grayscale images

Objective

In this project we want to get a deeper understanding of facial recognition systems in natural conditions and try to improve the currently available facial expressions recognition systems. We researched some recent papers and used techniques like transfer learning and image data augmentation. We also took use of class weights to improve accuracy.

Literature Survey

There has been a lot of research and lots of papers have been written For facial expression recognition in natural conditions and it is still an active field of research. There are a lot of available research papers on this topic. We started with the FER2013 [3] Challenge.

FER2013 is a Challenge/Dataset for human emotions in natural conditions, it was created by Goodfellow et al. to promote research in the field of facial expression recognition. Yinchuan Tang won this challenge with the accuracy of 71.2% by using CNN based model [4]. It is one of the first CNN based models for this kind of problem.

Yu and Zhang [5] used several CNN models and ensembled them to improve accuracy. They also used data augmentation on both training and testing datasets to further improve model accuracy.

Mollahosseini et al. [6] used transfer learning to improve accuracy; they used a model based on Inception Architecture [7].

From our research most recent works make use of data augmentation to increase valid input data except [8]. Some of them also make use of transfer learning to better classify the emotions.

Experimental Setup

- ❖ **Fedora 33 Workstation**
- ❖ **Python (3.7.10)**
- ❖ **TensorFlow (2.4.1)**
- ❖ **Keras (2.4.0)**
- ❖ **Keras-vggface**
- ❖ **Jupyter Notebook**
- ❖ **Google Colaboratory**
 - Google Colab (GPU) was used for training CNN models.
- ❖ **OpenCV(cv2)**
- ❖ **Matplotlib**
- ❖ **Pandas**
- ❖ **NumPy**
- ❖ **Scikit-Learn**

Methodology

We used all these methods in preprocessing of the data.

Data Preparation

The Fer2013 dataset is available [here](#). It is csv file named “example_submission.csv” which has three columns named “emotion”, “usage” and “pixels” where *emotion* is a label between 0 to 6.

usage is data type like Training(train), PublicTest(validation) and PrivateTest(test). And *pixels* is a 48x48 long string separated by a single space, every value in this string represents a single pixel hex color value (gray image).

We read this csv file and simply partitioned this input data into 3 directories for training, testing and validation and each directory contains 7 directories: angry, disgust, fear, happy, neutral, sad and surprise (one for each emotion).

Data Augmentation

We performed data augmentation on input images based on some papers and experimented with commonly known techniques like horizontal mirroring, image rotations, and horizontal shifting, vertical shifting and image zooms.

Class Weighting

We also used class weights to remove label imbalance from the input dataset.

Models

Models used for training the dataset.

Base Model

We created a simple CNN model to better understand the problem from the root. We started creating the model with BatchNormalization and then using four convolutional layers, all the convolutional layers have BatchNormalization and MaxPool layers after them and for output layers we used two fully-connected(Dense) layers and BatchNormalization.

Layer	Output Shape	Parameters
BatchNormalization	48, 48, 1	4
Conv2D	48, 48, 32	320
BatchNormalization	48, 48, 32	128
MaxPooling2D	24, 24, 32	0
Conv2D	24, 24, 64	18496
BatchNormalization	24, 24, 64	256
MaxPooling2D	12, 12, 64	0
Conv2D	12, 12, 64	36928
BatchNormalization	12, 12, 64	256
MaxPooling2D	6, 6, 64	0
Conv2D	6, 6, 64	36928
BatchNormalization	6, 6, 64	256
MaxPooling2D	3, 3, 64	0
Flatten	576	0
FC(Dense)	64	36928
BatchNormalization	64	256
FC(Dense)	7	455

Base Model

We used SGD optimizer with 0.01 learning rate and trained the model for 150 epochs and achieved accuracy of 65.03% on the test dataset.

Fine-Tuned vgg16 model

VGG16 is a very famous CNN based object classification model. It is effective on small and unbalanced datasets. We used Keras Vgg-face library to create a transfer learning model based on vgg16. Input in this new pretrained model is colored images of specific size. We recolored the 48x48 grayscale images, resized them to 128x128 and applied data augmentation on the test dataset during the training process.

Layer	Output Shape	Parameters
Input Layer	128, 128, 3	0
Conv2D	128, 128, 64	1792
Conv2D	128, 128, 64	36928
MaxPooling2D	64, 64, 128	0
Conv2D	64, 64, 128	73856
Conv2D	64, 64, 128	147584
MaxPooling2D	32, 32, 256	0
Conv2D	32, 32, 256	295168
Conv2D	32, 32, 256	590080
Conv2D	32, 32, 256	590080
MaxPooling2D	16, 16, 512	0
Conv2D	16, 16, 512	1180160
Conv2D	16, 16, 512	2359808
Conv2D	16, 16, 512	2359808
MaxPooling2D	8, 8, 512	0
Conv2D	8, 8, 512	2359808
Conv2D	8, 8, 512	2359808
Conv2D	8, 8, 512	2359808
MaxPooling2D	4, 4, 512	0
Flatten	8192	0
Dropout	8192	0
FC(Dense)	4096	33558528
Dropout	4096	0
FC(Dense)	1024	4195328
FC(Dense)	7	7175

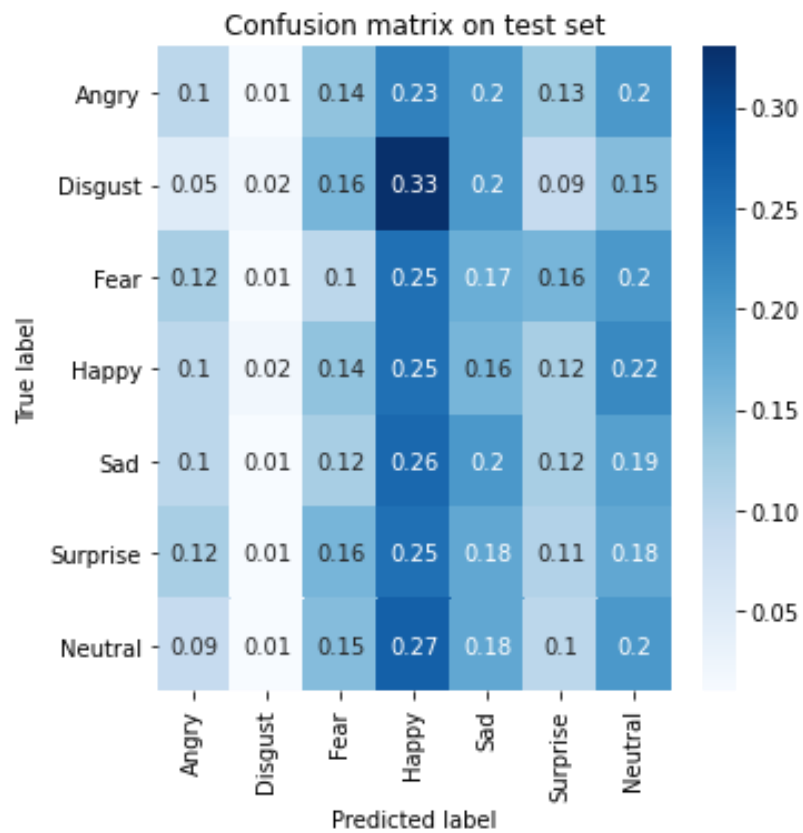
Fine-tuned VGG16 model

Even though VGG16 only has 16 layers it is a very complex model and has lots of parameters to train to. To fine tune vgg16 we froze all the layers in the pretrained model and added two fully-connected(Dense) and two dropout layers as output layers. We used Adam Optimizer with 0.0001 learning rate and trained the model for 100 epochs, after training we achieved 69.02% accuracy on the test dataset, 4% increase from the base model.

Results

We achieve accuracy of 65.03% on our base model and accuracy of 69.02% on our fine-tuned vgg16 model. This accuracy is 6% lower than previously achieved highest accuracy of 75.1%. But achieved accuracy is still in the range of human-level accuracy.

Model	Test accuracy
Human-Level	65±5%
Tang [4]	71.2%
Zhang [8]	75.1%
Vanilla Model	65.03%
Fine-tuned vgg16	69.02%



Confusion matrix of test set on fine-tuned vgg16 model

Project code is available [here](#).

Conclusion and Future Work

This project's purpose was to get a better understanding of facial recognition systems in natural conditions and try to improve the currently available facial expressions recognition systems. We explored a vanilla shallow CNN model and pre-trained network based on VGG16. We used data augmentation to increase relevant features. Some papers achieved high accuracy up to 75% by ensembling several models, but we mostly focused on facial features extraction.

Zhang et al. [8] used a HOG feature descriptor to extract facial features and used them as an input in his CNN model. About 15% of the time facial landmark detection is incorrect in this method for the FER2013 dataset. So, we tried using current facial landmark detection and facial feature extraction models, but current models do not work very well with small images and side facing images. We tried to train a [new model](#) based on Yin Guobing's model [9] that can work with side facing images and small images. We took six famous facial landmark datasets (300vw, 300w, ibug, helen, afw and lfpw) and extracted all the faces from images and created a new dataset of size 48 resized to 128. But because of limited time we could not complete the model. In future we want to work on the facial registration and facial feature extraction part of the project.

References

- [1] E. Sariyanidi, H. Gunes, and A. Cavallaro, “Automatic analysis of facial affect: A survey of registration, representation, and recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 37, no. 6, pp. 1113–1133, 2015.
- [2] M. V. B. Martinez, “Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition,” in *Advances in Face Detection and Facial Image Analysis*. Springer, 2016, pp. 63–100.
- [3] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, “Challenges in representation learning: A report on three machine learning contests,” *Neural Networks*, vol. 64, pp. 59–63, 2015.
- [4] Yichuan. Tang, “Deep Learning using Support Vector Machines,” in *International Conference on Machine Learning (ICML) Workshops*, 2013
- [5] Z. Yu and C. Zhang, “Image based static facial expression recognition with multiple deep network learning,” in *ACM International Conference on Multimodal Interaction (MMI)*, 2015, pp. 435–442.
- [6] A. Mollahosseini, D. Chan, and M. H. Mahoor, “Going Deeper in Facial Expression Recognition using Deep Neural Networks,” *CoRR*, vol. 1511, 2015.
- [7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going Deeper with Convolutions,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [8] Z. Zhang, P. Luo, C.-C. Loy, and X. Tang, “Learning Social Relation Traits from Face Images,” in *Proc. IEEE Int. Conference on Computer Vision (ICCV)*, 2015, pp. 3631–3639.
- [9] Yin Guobing <https://github.com/yinguobing/cnn-facial-landmark>