Emotion Detection from Facial Images

Sivasangaran V[1] **Abhijeet Jawalekar Rajendra**[2]

Abstract

Emotion recognition is a very important topic that has many real-world applications. In this paper, classification is performed on the publicly available FER-2013 dataset. We propose a combination of CNN and SVM to classify the emotions in the dataset. We do different experiments on different models and compare their performance. Our proposed model has a potential to outperform the traditional CNN models.

1. Technical Details

The FER-2013 dataset contains 35,887 facial grayscale 48x48 images with the images stored in CSV format. The dataset has 28,709 images for training, 3,589 images as public test (which we used for validation), and another 3,589 images (which we used for testing) as private test. Our model is based on the architecture shown in Fig-1. We have run experiments using different combinations of the following techniques – Batch Normalization (BN), Global Average Pooling (GAP), freezing of learning layers, and oversampling, PCA with DNN. These experiments are based on our benchmark research paper [1]. The experiments have been run for 30 epochs. The training was customized to use early stopping. The experiments were run with the SGD optimizer using batch size of 32, learning rate of 0.001 and momentum of 0.9. Our benchmark research paper [1] used optimizers Adam, SGD and SWATS but obtained best results with SGD and the said parameters. Our benchmark research paper used the VGG-16 architecture, but we have run the same experiments using a custom architecture because training the VGG-16 took a long time. However, we have run few experiments using the VGG-16 architecture also and have shown some results related to it as well.

2. Results

Fig-1 summarizes the architecture used and Table-1 summarizes the results obtained. The results show that our proposed model outperforms the traditional models.

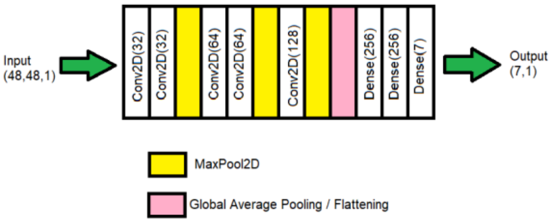

Figure-1: CNN Architecture

3. Novel Contributions

Sometimes having hard classification of face emotions is quite difficult and confusing among some categories so, we thought of predicting soft classes rather than predicting hard label about emotion. We extracted probabilities from last softmax layer and predicted hard label only if model has confidence more than 0.55 about that prediction and predicted two either or labels for confidence between (0.35-0.55) and predicted not known for other test data.[2]

SVMs are very popular because of the so-called “kernel trick”. SVMs also don’t have too many hyperparameters. CNNs are very efficient in feature extraction. So, we thought of extracting features using CNN and feeding it into an SVM[1]. Since XGBoost is also a very popular ensemble-based classifier and is known to outperform many other classifiers[2]. Hence, we also experimented classifying using the extracted features with XGBoost.[1] In these tables, A and B represent the CNN with the first 2 and first 4 Convolutional layers inactive respectively.

Table-1: Results of CNN Model (Values show accuracy in %)[3]

Balanced, No GAP, No BN	CNN	51.49
Balanced, No GAP, No BN	CNN/Dense	11.53
Balanced, No GAP, No BN	CNN/Softmax	22.84
Imbalanced, GAP, BN	CNN	55.17
Imbalanced, No GAP, No BN	CNN	54.16
Imbalanced, No GAP, No BN	CNN/A	54.03
Imbalanced, No GAP, No BN	CNN/B	24.94

The results for our proposed models[3] are summarized below:

Table-2: Results of CNN+SVM+XGB, PCA+X Models[3]

CNN	56.48%
CNN+SVM	**60.68%**
CNN+XGB	**60.35%**
CNN, VGG16	60.46%
CNN+SVM, VGG16	**62.80%**
PCA (D=784) + DNN	45.7%
PCA (D=512) + SVM	43.4%

2.1 Results of Soft Classification Model

Soft classification method for prediction predicted a single class for 76% of total test samples with 72.2% of accuracy and predicted 2 either or labels for 20% of test samples with 60.6% of accuracy and 4% as don’t know.

[1] Contributor: Sivasangaran V
[2] Contributor: Abhijeet Jawalekar Rajendra
[3] A complete table of results is available at [3]

References

1. Kusuma Negara, I Gede Putra & Jonathan, Jonathan & Lim, Andreas. (2020). *Emotion Recognition on FER-2013 Face Images Using Fine-Tuned VGG-16*. Advances in Science, Technology and Engineering Systems Journal. 5. 315-322. 10.25046/aj050638, available at https://www.researchgate.net/publication/347069053_Emotion_Recognition_on_FER-2013_Face_Images_Using_Fine-Tuned_VGG-16

2. Tinaqi Chan, Carlos Guestrin, *XGBoost: A Scalable Tree Boosting System*, available at https://arxiv.org/pdf/1603.02754.pdf

3. Our full tabulation of results available at https://docs.google.com/spreadsheets/d/1mv5khwK3ZJgTIA3tjcSjgpjEraXSKkcSCHjWOoofiyw/edit?usp=sharing

4. Savita Ahlawat, Amit Choudhary, *Hybrid CNN-SVM Classifier for Handwritten Digit Recognition*, Procedia Computer Science, Volume 167, 2020, Pages 2554-2560, ISSN 1877-0509, available at https://www.sciencedirect.com/science/article/pii/S1877050920307754

5. Setthanun Thongsuwan, Saichon Jaiyen, Anantachai Padcharoen, Praveen Agarwal, ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost, Nuclear Engineering and Technology, Volume 53, Issue 2, 2021, Pages 522-531, ISSN 1738-5733, available at https://doi.org/10.1016/j.net.2020.04.008

6. Matrixb. (2018, March 13). *CNN + SVM + XGBoost. Kaggle*. Retrieved April 25, 2022, from https://www.kaggle.com/code/matrixb/cnn-svm-xgboost/notebook

7. Xyashanandx. (2022, March 10). Human emotion recognition using CNN. Kaggle. Retrieved April 25, 2022, from https://www.kaggle.com/code/xyashanandx/human-emotion-recognition-using-cnn

Tools Used

Tensorflow and Keras – for neural networks

NumPy, Pandas, Matplotlib and Seaborn – for data analysis and visualization

Scikit Learn – for SVM, generation of classification report

XGBoost – for XGBoost

IMB Learn – for Oversampling

Pickle – for saving model files