Impact Factor: 6.252

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 2, May 2022

Emotion Recognition using Machine Learning

Gaurav Sheth[1], Sahil Pataskar[2], Siddhesh Kakade[3], Prof. Priyanka Khalate[4]

Students, Department of Computer Engineering[1,2,3]
Faculty, Department of Computer Engineering[4]
Sinhgad Inst. Smt. Kashibai Navale College of Engineering, Vadgaon Bk. Pune, Maharashtra, India
Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract: *One of the most universal ways that people communicate is through facial expressions. Facial expression recognition plays a crucial role in the area of human-machine interaction. Automatic facial expression recognition system has many applications including, but not limited to, human behavior understanding, detection of mental disorders, and synthetic human expressions. Recognition of facial expression by computer with high recognition rate is still a challenging task. In this paper, we take a deep dive, implementing multiple deep learning models for facial expression recognition (FER). Our goals are twofold: we aim not only to maximize accuracy, but also to apply our results to the real-world. Additionally, we showcase a mobile website which runs our FER models on-device in real time.*

Keywords: Machine Learning, Deep Neural Networks, WebApplication

I. Introduction

Facial expressions are fundamentally important in human communication. Although recognizing basic expressions under controlled conditions (e.g. frontal faces and posed expressions) is a solved problem with 98.9% accuracy, distinguishing basic expressions in natural conditions is still challenging due to variations in head pose, illumination, and occlusions [1-2].

However, with the advent of deep learning in the recentdecade, FER technology under natural conditions has achievedremarkable accuracy in categorizing emotions from facial images, exceeding human level performance. This has allowedfor the development of ground breaking applications in sociablerobotics, medical treatment, driver fatigue surveillance, and many other human-computer interaction systems [1].

In this paper, our goals were not only to better understand and improve the performance of emotion recognition models, but also to apply them to real world situations. We took several approaches from recentpublications to improve accuracy, including transfer learning, data augmentation, class weighting, adding auxiliary data, andensembling. In addition, we analyzed our models through erroranalysis and several interpretability techniques.

II. Related Work

FER2013 was designed by Goodfellow et al. as a Kaggle competition to promote researchers to develop betterFER systems. The top three teams all used CNNs trained discriminatively with image transformations [3]. The winner, Yichuan Tang, achieved a 71.2% accuracy by using the primal objective of an SVM as the loss function for training and additionally used the L2-SVM loss function [4]. This was a new development at the time and gave greatresults on the contest dataset.
There is a wealth of existing research in the FER domain.

In particular, a recent survey paper on FER by S. Li and W.Deng sheds light on the current state of deep-learning-basedapproaches to FER [1]. Another paper by Pramerdorfer andKampel [2] describes the approaches taken by six current state-of-the-art papers and ensembles their networks to achieve 75.2% test accuracy on FER2013, which is, to our knowledge, the highest reported in any published journal paper.

Among the six papers, Zhang et al. achieved the highestaccuracy of 75.1% by employing auxiliary data and additional features: a vector of HoG features were computedfrom face patches and processed by the first FC layer of the CNN (early fusion). They also employed facial landmark registration, suggesting its benefits even in challenging conditions (facial landmark extraction is inaccurate for about15% of images in the FER dataset) [5]. The paper with the second highest accuracy by Kim et al. utilized face registration, data augmentation, additional features, and ensembling [6].

DOI: 10.48175/568

From graduate community at Stanford, we also found reports from recent CS229 and CS230 projects on FER useful asreference [7,8].

III. Data and Pre-Processing

Pre-processing generally means the transformationsthat are applied to our dataset before feeding it to the MachineLearning/Deep Learning algorithm. Data Preprocessing is a technique that is used to convert the raw data into a cleandata set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

For pre-processing purposes we have used different Deep learningand Machine Learning libraries like Keras and Numpy. FER is a well-studied field with numerous available datasets. We used FER2013 as our main dataset. FER2013 Dataset FER2013 is a well-studied dataset and has been used in ICML competitions and several research papers. It is one of the more challenging datasets with human-level accuracy only at 65±5% and the highest performingpublished works achieving 75.2% test accuracy. Easilydownloadable on Kaggle, the dataset's 35,887 contained images are normalized to 48x48 pixels in grayscale. FER2013 is, however, not a balanced dataset, as it contains images of 7 facial expressions, with distributions of Angry (4,953), Disgust(547), Fear (5,121), Happy (8,989), Sad (6,077), Surprise (4,002), and Neutral (6,198) [3].

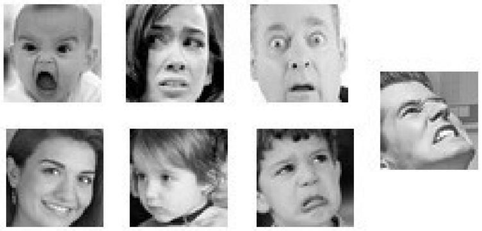


Figure 1: Images from each emotion class in the FER2013 dataset.

IV. MODELS

4.1 Baseline Model

In order to better understand the problem, we decided to first try to tackle this problem from scratch, building a vanilla CNN using four 3x3x32 same-padding, ReLU filters, interleaved with two 2x2 MaxPool layers, and completed with a FC layer and softmax layer. We also added batchnorm and 50% dropout layers to address high variance and improve our accuracy from 53.0% to 64.0%.

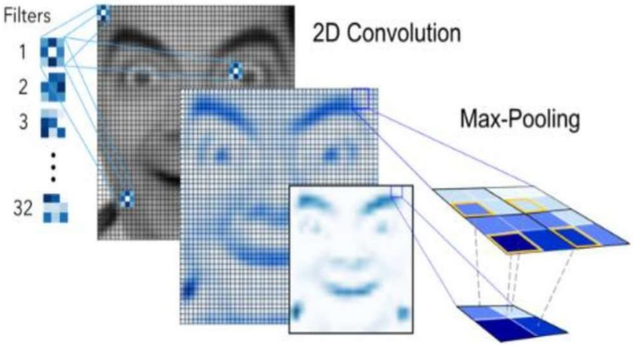


Figure 2: Convolution Neural Networks

4.2 Transfer Learning

Since the FER2013 dataset is quite small and unbalanced, we found that utilizing transfer learning significantly boosted the accuracy of our model. We explored transfer learning, using the Keras VGG-Face library and each of ResNet50, VGG16 and MobileNet as our pre-trained models.

To match the input requirements of these new networks which expected RGB images of no smaller than 197x197, we resized and recolored the 48x48 grayscale images in FER2013 during training time.

V. Methods

5.1 Auxiliary Data & Data Preparation

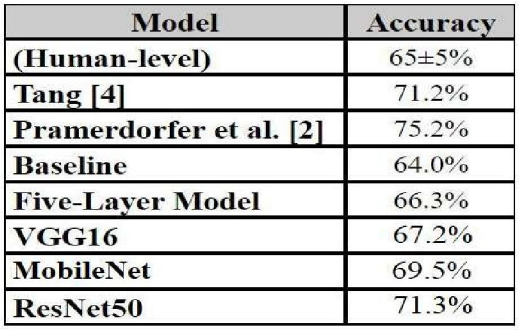
Although several FER datasets are available online, they vary widely in image size, color, and format, as well as labeling and directory structure. We addressed these differences by simply partitioning all input datasets into 7 directories (one for each class). During training, we loaded images in batches from disk (to avoid memory overflow) and utilized Keras data generators to automatically resize and format the images.

5.2 Data Augmentation

We researched and experimented with commonly used techniques in existing FER papers and achieved our best results with horizontal mirroring, ±10 degree rotations, ±10% image zooms, and ±10% horizontal/vertical shifting.

VI. Results and Analysis

Table 1 shows the accuracies our best models achieved on the FER2013 private test dataset. Results from Tang [4] (the Kaggle competition winner) and Pramerdorfer et al. [2] (the highest published accuracy) are also depicted.



Model	Accuracy
(Human-level)	65±5%
Tang [4]	71.2%
Pramerdorfer et al. [2]	75.2%
Baseline	64.0%
Five-Layer Model	66.3%
VGG16	67.2%
MobileNet	69.5%
ResNet50	71.3%

Table 1: Network parameters and results on FER2013 private test set.

Given the high complexity of transfer learning models and relatively small size of our datasets, we also experienced overfitting while training. Shown in Figure 5, although we added 50% dropout for the last three layers, our ResNet50 transfer learning model quickly overfit to the training set with dev accuracy starting to flatten after only 30 epochs.

6.1 Error Analysis

For error analysis, we targeted confusion matrix cells with high misclassifications. One interesting example was an image labeled fear that was classified by our model as angry (29%), fear (28%), and sad (26%), similar to mispredictions by humans on the same image. We also investigated the high misclassification of sad images as neutral, where one of the subjects was misclassified on all of his sad images. Recognizing this as a class imbalance problem, we overcame it by collecting more sad images for our web app dataset.

6.2 Challenges

It is worth mentioning that because our models exceeded human-level accuracy, error analysis was particularly challenging for some misclassifications, such as the fear image discussed prior.

Additionally, because emotions are highly subjective, Bayes error is high and it is often the case that an image can have multiple interpretations [17].

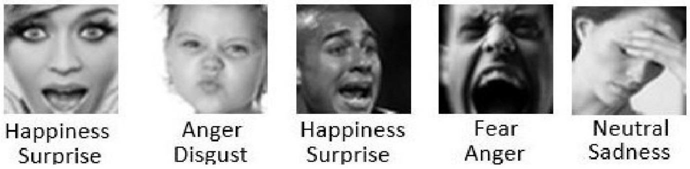


Figure 3: FER2013 images with two possible labels.

DOI: 10.48175/568

VII. Conclusion and Future Works

When we started this project, we had two goals, namely, to achieve the highest accuracy and to apply FER models to the real world. We explored several models including shallow CNNs and pre-trained networks based on SeNet50, ResNet50, and VGG16. To alleviate FER2013's inherent class imbalance, we employed class weights, data augmentation, and auxiliary datasets. We also found through network interpretability that our models learned to focus on relevant facial features for emotion detection. Additionally, we demonstrated that FER models could be applied in the real world by developing a website with real-time recognition speeds. We overcame data mismatch issues by building our own training dataset and also tuned our architecture to run on-device with minimal memory, disk, and computational requirements.

References

[1]. S. Li and W. Deng, "Deep facial expression recognition: A survey," arXiv preprint arXiv:1804.08348, 2018.

[2]. Pramerdorfer, C., Kampel, M.: Facial expression recognition using convolutional neural networks: state of the art. Preprint arXiv:1612.02903v1, 2016.

[3]. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., "Challenges in representation learning: A report on three machine learning contests," in International Conference on Neural Information Processing. Springer, 2013, pp. 117–124.

[4]. Y. Tang, "Deep Learning using Support Vector Machines," in International Conference on Machine Learning (ICML) Workshops, 2013.

[5]. 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.

[6]. B.-K. Kim, S.-Y. Dong, J. Roh, G. Kim, and S.-Y. Lee, "Fusing Aligned and Non-Aligned Face Information for Automatic Affect Recognition in the Wild: A Deep Learning Approach," in IEEE Conf. Computer Vision and Pattern Recognition (CVPR) Workshops, 2016, pp. 48–57.

[7]. Quinn M., Sivesind G., and Reis G., "Real-time Emotion Recognition From Facial Expressions", 2017

[8]. Wang J., and Mbuthia M., "FaceNet: Facial Expression Recognition Based on Deep Convolutional Neural Network", 2018

[9]. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn–Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, San Francisco, CA, USA, Jun. 2010, pp. 94–101

[10]. Brechet P., Chen Z., Jakob N., Wagner S., "Transfer Learning for Facial Expression Classification" Available: https://github.com/EmCity/transfer- learning-fer2013

[11]. Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. "SMOTE: Synthetic Minority Over-sampling Technique". JAIR 16 (2002), 321-357.

[12]. A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going Deeper in Facial Expression Recognition using Deep Neural Networks," CoRR, vol. 1511, 2015.

[13]. 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

[14]. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient- based localization," in 2017 IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[15]. Kapishnikov, A., Bolukbasi, T., Viégas, F. and Terry, M., "XRAI: Better Attributions Through Regions", 2019.

[16]. Minaee S., Abdolrashidi A., "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network", 2019

[17]. E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in Proceedings of the 18th ACM International Conference on Multimodal Interaction. ACM, 2016, pp. 279–283.

[18]. He H., Bai Y., Garcia E. A., and Li S. "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 2008, pp. 1322-1328.

[19]. A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," IEEE Transactions on Affective Computing, vol. PP, no. 99, pp. 1–1, 2017.