Facial Expression Recognition

Akash Kadel, Brenton Arnaboldi, Seda Bilaloglu
Center for Data Science
New York University
New York, NY 10011
ak6201@nyu.edu, ba1303@nyu.edu, sb3923@nyu.edu

1 Introduction - Initial Idea

We plan to build a classifier to predict facial expressions. The dataset we will use consists of over 35,000 48x48 pixel grayscale images of faces. Images (faces) in the dataset can fall into one of seven categories: 0) Angry, 1) Disgust, 2) Fear, 3) Happy, 4) Sad, 5) Surprise, and 6) Neutral. The dataset, among other things, was used in a Kaggle competition in 2013 (Fer2013). The top-performing model in the Kaggle competition achieved a test-set accuracy of 71% using an ensemble of convolutional neural nets followed by a multi-class SVM (rather than the softmax function) at the top layer (Tang, 2016). Because this competition took place in 2013, we are interested in testing new model architectures (introduced after 2013) on this dataset. **In particular, we would like to test the effectiveness of the Inception model for this task.**

2 Literature Review

No	Model Description	Architecture	Depth	Year	Accuracy
1	CNN Ensemble (VGG + ResNet) [10]	CPCPCPCPCPFF	10	2017	75.8 %
2	Ensemble of modern deep CNNs [8]	CPCPCPCPFF	8	2016	75.2 %
3	Deep CNN + gender, expression, pose, age-related attributes [14]	CPNCPNCPCFF	6	2015	75.10 %
4	Ensemble CNN with face alignment information [4]	CPCPCPFF	5	2016	73.73 %
5	Hybrid CNN– SIFT Aggregator [1]	CCPCCPCCP	9	2016	73.4 %
6	CNN with random perturbations [13]	PCCPCCPCFFF	8	2015	72.2 %
7	Hierarchical Deep CNN [5]	CPCPCPFF	5	2016	72.2 %
8	CNN [9]	CPCPCPCPFF	8	2017	71.9 %
9	CNN + complexity perception classification [3]	CPCCPCCPCCPCCPFFF	17	2018	71.35 %
10	DL + SVM [12]	CPCPFF	4	2013	71.2 %
11	Deep CNN [7]	CPCPIIPIPFFF	11	2015	66.4 %
12	CNN Ensemble [6]	CPCPCPCPFFF	9	2016	65.03 %
13	CNN + Histogram of Oriented Gradients [2]	CPF	2	2017	64 %

Table 1: Summary of previous architectures used for facial expression recognition task

Facial expressions are crucial for human interactions. Humans are capable of recognizing various emotions of happiness, sadness, anger, surprise, disgust, and fear accurately by looking at person's face. Facial expression recognition is one of the popular and challenging problems in computer vision

and it's important when designing proper human-computer interaction, behavioral science research, etc.

Over the last 5 years different groups tried various architectures as summarized in Table-1. Among these, ensemble deep CNN architectures (proposed by Savoiu et al) achieved the highest accuracy on the Fer2013 dataset.

Google's inception layer [11] was proposed to approximate sparse networks to having deep network without over fitting or computational problems. Given that using the Inception layer in Deep Neural Network has had remarkable results in the past, we aim exploiting inception layers in our architecture. Pramerdorfer et al [8] tested inception and reported test accuracy of 71.6%. Also, Mollahosseini et al. [7] included inception layers in their ensemble architecture which resulted in 66.4%.

3 Baseline Analysis

So far, we have implemented i) a very simple feed-forward neural network and ii) a simple CNN. The code for these models can be found on our GitHub page here. Our results have not been great so far – both models are achieving roughly 25% accuracy, which is right around the baseline, as the most common emotion in the data ("Happy") covers 25% of images. As such, we have significant modeling work to do over the next few weeks.

References

[1] Mundher Al-Shabi, Wooi Ping Cheah, and Tee Connie. Facial expression recognition using a hybrid CNN-SIFT aggregator. *CoRR*, abs/1608.02833, 2016.

[2] Shima Alizadeh and Azar Fazel. Convolutional neural networks for facial expression recognition. *CoRR*, abs/1704.06756, 2017.

[3] T. Chang, G. Wen, Y. Hu, and J. Ma. Facial Expression Recognition Based on Complexity Perception Classification Algorithm. *ArXiv e-prints*, February 2018.

[4] 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 *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1499–1508, June 2016.

[5] Bo-Kyeong Kim, Jihyeon Roh, Suh-Yeon Dong, and Soo-Young Lee. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. *Journal on Multimodal User Interfaces*, 10(2):173–189, 2016.

[6] K. Liu, M. Zhang, and Z. Pan. Facial expression recognition with cnn ensemble. In *2016 International Conference on Cyberworlds (CW)*, pages 163–166, Sept 2016.

[7] Ali Mollahosseini, David Chan, and Mohammad H. Mahoor. Going deeper in facial expression recognition using deep neural networks. *CoRR*, abs/1511.04110, 2015.

[8] Christopher Pramerdorfer and Martin Kampel. Facial expression recognition using convolutional neural networks: State of the art. *CoRR*, abs/1612.02903, 2016.

[9] D. V. Sang, N. Van Dat, and D. P. Thuan. Facial expression recognition using deep convolutional neural networks. In *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*, pages 130–135, Oct 2017.

[10] Alexandru Savoiu and James Wong. Recognizing facial expressions using deep learning. *CoRR*, abs/1704.06756, 2017.

[11] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.

[12] Yichuan Tang. Deep learning using support vector machines. *CoRR*, abs/1306.0239, 2013.

[13] Zhiding Yu and Cha Zhang. Image based static facial expression recognition with multiple deep network learning. November 2015.

[14] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Learning social relation traits from face images. *CoRR*, abs/1509.03936, 2015.