Introduction to Computer Vision Final Project Report

Group 66: Isaac Li, Yifan Wang, Zeshu Zhao
University of Virginia, Charlottesville, VA 22904
[il5fq, yw5ma, zz8ec]@virginia.edu

Abstract

Face recognition is both an active research area of computer vision and a widely-used technology in our lives. In our project, we first use an existing face detection system, Multi-task Cascaded Convolutional Neural Networks (CNN), to extract human faces [3]. We then train a face recognition and emotion detection CNN based on two public datasets including [6, 7]. Depending on the results of the CNNs, we replace their faces with corresponding Emojis. Eventually, our face detection and expression recognition algorithm is expected to produce accurate Emoji replacement for human faces in real time.

1. Introduction

Emotion detection is becoming increasingly important in the past decade. It is becoming a trending key in many social network application and medical appliances, such as expressing feelings through emoji and diagnosing mental diseases. It is challenging for computers to understand facial expression of human beings, which requires the computer to not only read and locate the position of the face, while taking different body postures and environment into consideration, but also accurately and reliably analyze facial expression. Thus in order to achieve such goals, our facial emotional detection algorithm is divided into three stages, preprocessing, extracting features and analyzing based on trained datasets. [1] Raw datasets normally include images of human faces with their respective facial expressions identified. Since facial detection algorithms are sufficiently developed, an already available algorithm is ued for preprocessing and feature extraction.

2. Related Work

Previous works on facial detection used hierarchical knowledge-based method which consists of three levels, the higher two layers on based on mosaic images at different resolution. While the lower layer utilized improved edge-detection method. [2] Improved facial detection in unconstrained environment uses a multi-task cascaded convolution network that identify the correlation of facial detection and alignment to boost up the performance of both tasks. [3]

Early paper on static emotional detection started from full-face pictures of people exhibiting certain facial expression. The facial figures are compressed into sides and the feature pattern sets are extracted. The architecture of the early works can be simplified into two fully connected layer, and network connection feeds forward and trained with backward propagation. [4] Other research includes creating track points for faces and shoulders, then utilize dynamic and static classification to determine the specific facial expression of the picture. [5]

3. Model

As mentioned before, face detection was mainly handled by Multi-task Cascaded CNN. For the facial expression recognition part, we first propose a vanilla CNN network with two layers of CNN. Their kernel sizes are set to 5, with output channels of 16 and 32. We also add batch normalization and max pooling after each CNN layer. Finally, we add a linear layer at the end to make predictions. For ResNet, we simply use the pretrained model with small modifications to fit our input.

4. Experiments and Results

Our models were implemented with pytorch. The simple two-layer CNN model performed well on the CK+ dataset, but only achieved a testing accuracy of around 50%. This could be due to the fact that the CK+ dataset pictures are mostly front view faces, while the FER2013 dataset pictures are mostly faces of side view and our simple two-layer CNN can not deal with faces of different views. In terms of ResNet18, it has too many parameters and requires a long time to train. For the CK+ dataset, it achieved high accuracy. Although initially it didn't do well due to the high learning rate. However, for the FER dataset, ResNet18 didn't do a good job either, and as we train more

Method	Train-data	Train Acc	Test Acc
CNN	CK+	99.73%	98.75%
CNN	FER2013	67.61%	56.35%
ResNet18	CK+	98.43%	98.44%
ResNet18	FER2013	60.74%	55.06%

Table 1. Experimental results for CNN and ResNet implementations.

Figure 1. Here is a photo of our group members and a sample output from our model with our faces swapped with Emojis

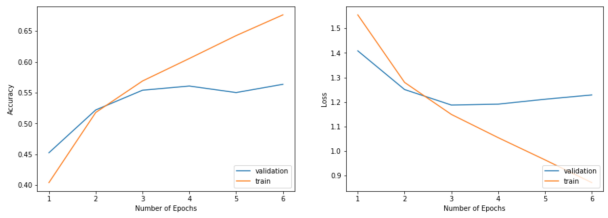

Figure 2. CNN FER Dataset Accuracy and Loss

epochs, it over-fitted on the training dataset.

Based on the figures above, we can see that the facial expression recognition part still requires quite some improvements: while all of us were smiling, only one smiling face was detected.

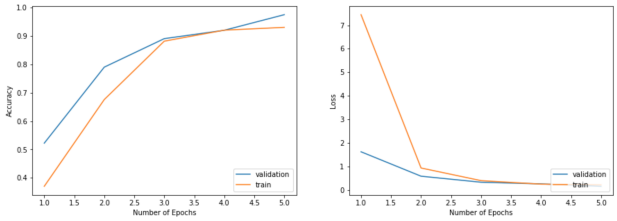

Figure 3. CNN CK+ Dataset Accuracy and Loss

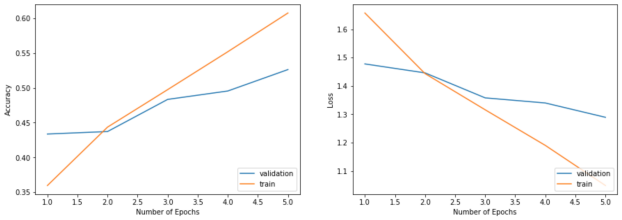

Figure 4. ResNet18 FER Dataset Accuracy and Loss

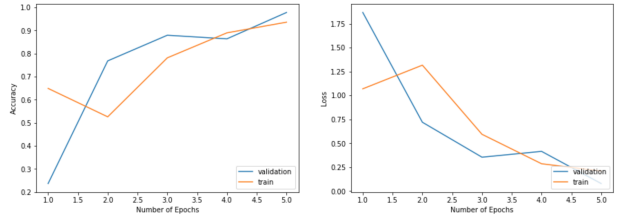

Figure 5. ResNet18 CK+ Dataset Accuracy and Loss

References

1. Samadiani, Huang, Cai, Luo, Chi, Xiang, and He, "A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data," Sensors, vol. 19, no. 8, p. 1863, 2019.

2. G. Yang and T. S. Huang, "Human face detection in a complex background," Pattern Recognition, vol. 27, no. 1, pp. 53–63, 1994.

3. Z. Cai, Q. Liu, S. Wang, and B. Yang, "Joint Head Pose Estimation with Multi-task Cascaded Convolutional Networks for Face Alignment," 2018 24th International Conference on Pattern Recognition (ICPR), 2018.

4. C. Padgety, G. Cotyrell, "Identifying Emotion in Static Face Images," Proc. of the 2nd Joint Symposium on Neural Computation. San Diego, USA: University of California, 1995.

5. S. Petridis, H. Gunes, S. Kaltwang, and M. Pantic, "Static vs. dynamic modeling of human nonverbal behavior from multiple cues and modalities," Proceedings of the 2009 international conference on Multimodal interfaces - ICMI-MLMI 09, 2009.

6. 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.

7. P. F. Lucey, J. F. Cohn, T. F. Kanade, J. F. Saragih, Z. F. Ambadar, and I. F. Matthews, "The Extended Cohn-Kanade Dataset (CK): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2010.