Project Proposal

More details on the data and algorithms at each part
exactly the inputs outputs and experiments

Description

Facial expressions are a crucial component of humans' non-verbal communications and social interactions. Having the ability to detect faces and classify these expressions correctly and efficiently can result in a wide variety of useful applications such as psychological analysis, forensics (lie detection), medical diagnosis, impact of advertisements and so on. Advanced machine learning techniques can be employed to determine facial expressions.

In this project, we will implement a system that could detect faces in the image and classify facial expressions associated with them. The project will be developed three parts: Image denoising, face detection and emotion classification. The input will be images of one or more people and the output will be the labels of facial expressions for each person in the frame. Labels will be of seven emotions: anger, disgust, fear, happy, sad, surprise and neutral. Some tests will be implemented and the result would be discussed.

Implementation

Part 1: This part is image denoising. Generally, for many data, will be used in later parts, may be very noisy, and it is very hard to be used in face detection and emotion classification. Thus, we will first denoising the picture, which has very low PSNR. Then, pipe to the next part, which may increase the success rate in later parts. We are trying to use the technique described in "Image Super-Resolution Using Deep Convolutional Networks" [1], http://arxiv.org/abs/1501.00092.

Part 2: Face Detection [4].
Viola-Jones object detection framework will be used in this part. This framework is invented by Paul Viola and Michael Jones in 2003. It's a widely used and fast enough method. The input would be dataset from CBCL Face Database #1 - MIT Center For Biological and Computation Learning and the output would be face images . The general idea would be: First, creating integral images for train sets and use Haar Feature selection to extract the weak features; Second, using Adaboost algorithm to select the best features and train cascade classifiers; Third, using trained cascade classifiers on test set to detect the faces.

Part 3: Facial expression classification can be achieved by using convolutional neural networks. In this project, we are going to modify canonical VGG convolutional networks for image classification as Sang. D and et al [2] proposed. Image preprocessing and image augmentation steps will be undertaken before training the network. The network will be trained separately on labeled FERC - 2013 dataset. Finally, we will report the performance of the network on both a subset of FERC - 2013 dataset and CK+ dataset.

Task Allocation

Part 1 - Cheng Zeng
Part 2 - Chen Chang
Part 3 - Zwe Naing

Datasets

1. FERC - 2013 dataset - This dataset is provided by Kaggle facial expression competition. The dataset includes 35887 gray images of 48x48 resolution. The faces are not posed. The labels include anger, disgust, fear, happy, sad, surprise and neutral.
2. CBCL Face Database #1 - MIT Center For Biological and Computation Learning
   This is a database of faces and non-faces that mentioned in [4], this dataset has been used extensively at the Center for Biological and Computational Learning at MIT. It is freely available for research use.
   Training set: 2,429 faces, 4,548 non-faces
   Test set: 472 faces, 23,573 non-faces

Reference Papers

[1]Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang, "Image Super-Resolution Using Deep Convolutional Networks", http://arxiv.org/abs/1501.00092

[2]Sang, D., Van Dat, N., & Thuan, D. (2017). Facial expression recognition using deep convolutional neural networks. Knowledge and Systems Engineering (KSE), 2017 9th International Conference on, 130-135.

[3]Zeng, Zhang, Song, Liu, Li, & Dobaie. (2018). Facial expression recognition via learning deep sparse autoencoders. Neurocomputing, 273, 643-649.

[4]Viola, P. & Jones, M.J. (2004). Robust Real-Time Face Detection. International Journal of Computer Vision (2004) 57: 137.