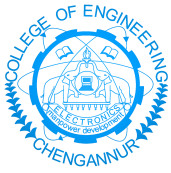
A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing

03CS6902 Mini Project
Design Report

CHN20CSIP08 Vishnu Vinod
vishnuvinod499@gmail.com
M. Tech. Computer Science (Image Processing)



Department of Computer Engineering
College of Engineering Chengannur
Alappuzha 689121
Phone: +91.479.2165706
http://www.ceconline.edu
hod.cse@ceconline.edu

Abstract

Facial emotion recognition is the process of detecting human emotions from facial expressions. The human brain recognizes emotions automatically and software has now been developed that can recognize emotions as well. This technology is becoming more accurate all the time, and will eventually be able to read emotions as well as our brains do. The facial expression recognition, as an important means of intelligent human-computer interaction, has a broad application background. It has been applied in the fields of assistant medicine, distance education, interactive games and public security. The facial emotion recognition extracts the information representing the facial expression features from the original input facial expression images through computer image processing technology, and classifies the facial expression features according to human emotional expression, such as happiness, surprise, aversion and neutrality. In the recent years, the development of facial expression recognition technologies has been rapid and many scholars have contributed to the development of facial expression recognition. Although the CNN algorithm has made some progress in the field of facial expression recognition, it still has some shortcomings, such as too long training time and low recognition rate in the complex background. To avoid the complex process of explicit feature extraction in traditional facial expression recognition, a facial expression recognition method based on CNN and image edge detection is proposed.

Contents

Chapter 1

Introduction

Human emotions can be classified as fear, contempt, disgust, anger, surprise, sad, happy, and neutral. These emotions are very subtle. Facial muscle contortions are very minimal and detecting these differences can be very challenging as even a small difference results in different expressions. Also, expressions of different or even the same people might vary for the same emotion, as emotions are hugely context dependent. While we can focus on only those areas of the face which display a maximum of emotions like around the mouth and eyes, how we extract these gestures and categorize them is still an important question. Neural networks and machine learning have been used for these tasks and have obtained good results.

1.1 Proposed Project

1.1.1 Problem Statement

Computers and other electronic devices in our daily lives will become more user-friendly if they can adequately interpret a person's facial expressions, thereby improving human-machine interfaces. The goal of this project is to develop a framework which can predict, from the grayscale picture of a person's face, which emotion the facial expression conveys.

1.1.2 Proposed Solution

Here a face emotion recognition system based on image processing and convolutional neural network is proposed. The facial expression recognition extracts the information representing the facial expression features from the original input facial expression images through computer image processing technology, and classifies the facial emotion features according to human emotional expression, such as happiness, surprise, aversion and neutrality using a convolutional neural network.

Chapter 2

Project Design

The proposed system has two stages facial emotion data pre-processing and facial emotion recognition using convolutional neural network. In the first stage the input facial image will be processed by using some image processing techniques to extract emotion features and in the second stage the facial image will be classified to the specific emotion based on the features using convolutional neural network developed.

2.1 Facial Emotion Data Preprocessing

Because the original pictures of facial expressions have complex background, different sizes, different shades and other factors, a series of image pre-processing processes have to be completed before facial expressions are input into the network for training. Firstly, we locate the face in the image and cut out the face image. Then, we normalize the face image to a specific size. Next, we equalize the histogram of the image to reduce the influence of illumination and other factors. Finally, we extract the edge of each layer of the image in the convolution process. The extracted edge information is superimposed on each feature image to preserve the edge structure information of texture image.

2.1.1 Face Detection And Location

This method uses the Haar-like to extract facial features, and uses an integral graph to realize fast calculation of Haar-like features, and screens out important features from a large number of Haar-like features. Then, we use the Adaboost algorithm to train and integrate the weak classifier into a strong classifier

Haar-like feature

Haar-like features are digital image features used in object recognition. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums.

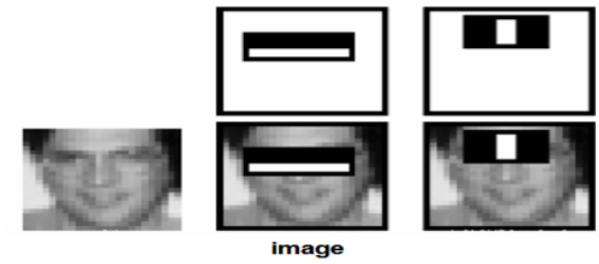


Figure 2.1: Haar Feature

Integral Images

An Integral Image is an intermediate representation of an image . Value for location (x, y) on the integral image equals the sum of the pixels above and to the left (inclusive) of the (x, y) location on the original image . This intermediate representation is essential because it allows for fast calculation of rectangular region.

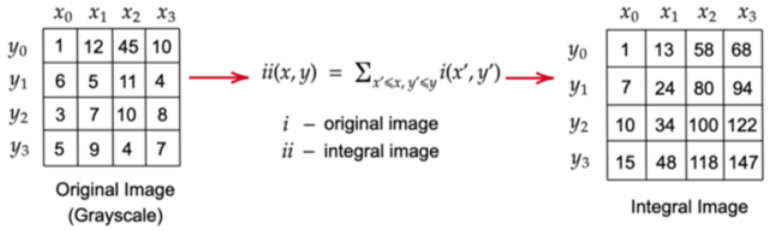


Figure 2.2: Integral Image

Haar Cascade Classifier

This is basically a machine learning based approach where a cascade function is trained from a lot of images both positive and negative. Based on the training it is then used to detect the objects in the other images. Based on the training it is then used to detect the face from the images. And using certain program codes(like OpenCV) we can extract the face from the input image as an image.

Adaboost Algorithm

AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique that is used as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights to incorrectly classified instances. Boosting is used to reduce bias as well as the variance for supervised learning. It works on the principle where learners are grown sequentially. Except for the first, each subsequent learner is grown from previously grown learners. In simple words, weak learners are converted into strong ones. Adaboost algorithm also

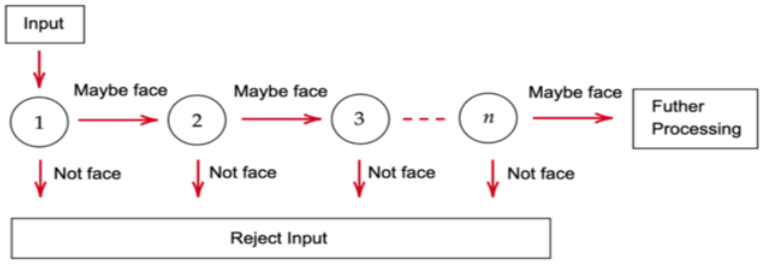


Figure 2.3: Haar Classifier

works on the same principle as boosting, but there is a slight difference in working. Let's discuss the difference in detail.

2.1.2 Scale Normalization

Through normalization, the input image is scaled to 128*128 size. Let point (x, y) in the original picture be normalized and mapped to point x' , y' . The mapping is as follows

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Where sx represents the scaling ratio of the image in the direction of x axis and sy represents the scaling ratio of the image in the direction of y axis. In the process of image scaling, bilinear interpolation algorithm is also needed to fill the image.

Bilinear interpolation algorithm

Bilinear Interpolation is a resampling method that uses the distance weighted average of the four nearest pixel values to estimate a new pixel value. The four cell centers from the input raster are closest to the cell center for the output processing cell will be weighted and based on distance and then averaged.
A, B, C and D are the four points around the pixel (x, y). The corresponding gray values are g (A), g (B), g (C), g (D). To get the gray value of point (x, y) and calculate the gray value of points E and F, the formula is as follows:
g (E) = (x - xD) (g (C) - g (D)) + g (D)
g (F) = (x - xA) (g (B) - g (A)) + g (A)
xA and xD are the abscissa of point A and point D, respectively. The gray scale formula of (x, y) is as follows: g (x, y) = (y - yD) (g (F) - g (E)) + g(E) where yD represents the ordinates of CD points.

2.1.3 Gray Level Equalization

In the actual image acquisition process, it is easy to be affected by illumination, shadows and other factors. It is necessary to average the gray level of the image to enhance the contrast of the image. The Histogram Equalization (HE) method is used to process images. Histogram equalization is a technique for adjusting image intensities to enhance contrast.

Histogram Equalization

Histogram Equalization is a computer image processing technique used to improve contrast in images. It accomplishes this by effectively spreading out the most frequent intensity values, i.e. stretching out the intensity range of the image. This method usually increases the global contrast of images when its usable data is represented by close contrast values. This allows for areas of lower local contrast to gain a higher contrast.The basic idea is to transform the histogram of the original graph into a uniform distribution form. If the gray level of the gray image is Ł, the size

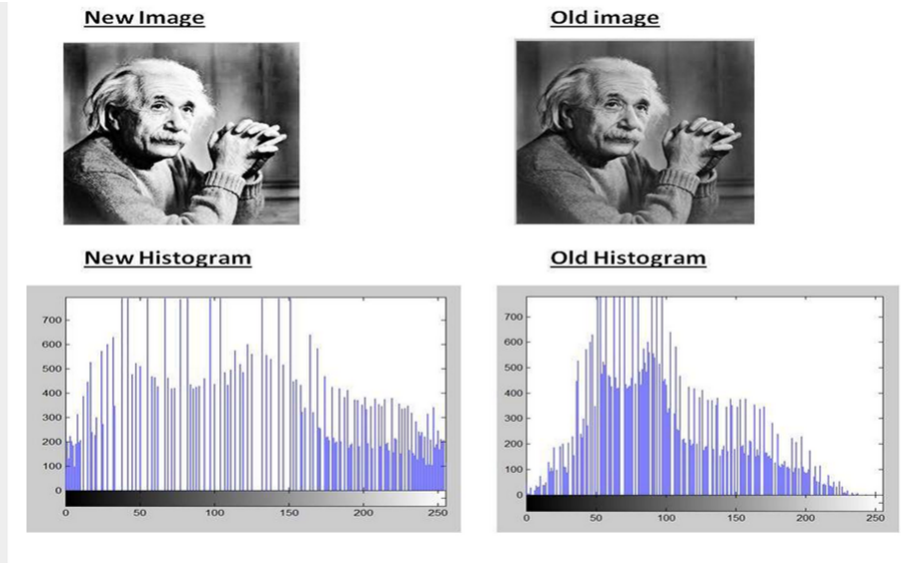


Figure 2.4: Histogram

is M×N, and the number of pixels in the r_i gray level is E, the corresponding probability of gray level occurrence is as follows:

$$P_r(r_i) = \frac{n_i}{M \times N}, \quad i = 0, 1..., L-1 \tag{2.1}$$

Subsequently the cumilative distribution function is calculates using the following equation.

$$T(r_i) = \sum P_r(r_j), \quad i = 0, 1..., L-1 \tag{2.2}$$

Finally the image histogram is averaged using the following mapping relations:

$$e_j = INT[(e_{max} - e_{min})T(r) + e_{min} + 0.5], \quad j = 0, 1..., L-1 \tag{2.3}$$

2.1.4 Image Edge Detection

Kirsch edge operator is used to extract image edge information. The Kirsch operator or Kirsch compass kernel is a non-linear edge detector that finds the maximum edge strength in a few predetermined directions. The operator takes a single kernel mask and rotates it in 45 degree increments

$$N = \begin{vmatrix} +5 & +5 & +5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{vmatrix} \quad NW = \begin{vmatrix} +5 & +5 & -3 \\ +5 & 0 & -3 \\ -3 & -3 & -3 \end{vmatrix} \quad W = \begin{vmatrix} +5 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & -3 & -3 \end{vmatrix}$$

$$SW = \begin{vmatrix} -3 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & +5 & -3 \end{vmatrix} \quad E = \begin{vmatrix} -3 & -3 & +5 \\ -3 & 0 & +5 \\ -3 & -3 & +5 \end{vmatrix} \quad NE = \begin{vmatrix} -3 & +5 & +5 \\ -3 & 0 & +5 \\ -3 & -3 & -3 \end{vmatrix}$$

$$S = \begin{vmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ +5 & +5 & +5 \end{vmatrix} \quad SE = \begin{vmatrix} -3 & -3 & +5 \\ -3 & 0 & +5 \\ -3 & +5 & +5 \end{vmatrix}$$

Figure 2.5: Kernels Used

through all 8 compass directions: N, NW, W, SW, S, SE, E, and NE. Each pixel in the image is operated with these 8 kernels (convolution) . The edge magnitude of the Kirsch operator is calculated as the maximum magnitude across all directions

Convolving mask over image

Place the center of the mask at each element of an image. Multiply the corresponding elements and then add them , and paste the result onto the element of the image on which you place the center of mask.

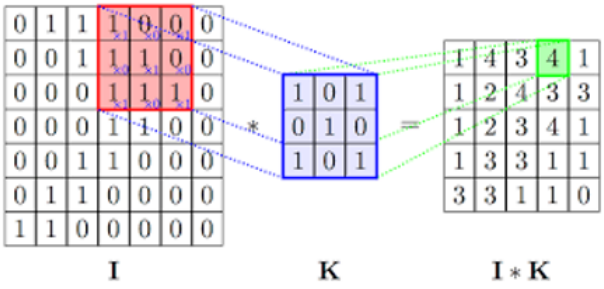


Figure 2.6: Convolution of Mask

2.2 Convolutional Neural Network For Emotion Recognition

Image classification is the process of labeling images according to predefined categories. The process of image classification is based on supervised learning. An image classification model is fed a set of images within a specific category. Based on this set, the algorithm learns which class the test images belong to, and can then predict the correct class of future image inputs, and can even measure how accurate the predictions are. Convolutional neural networks are composed of multiple layers of artificial neurons. Artificial neurons, a rough imitation of their biological counterparts, are mathematical functions that calculate the weighted sum of multiple inputs and outputs an activation value. The behavior of each neuron is defined by its weights. When fed with the pixel values, the artificial neurons of a CNN pick out various visual features. Training of CNN with facial emotion images is very important. FER-2013 dataset is going to use for training and testing the CNN.

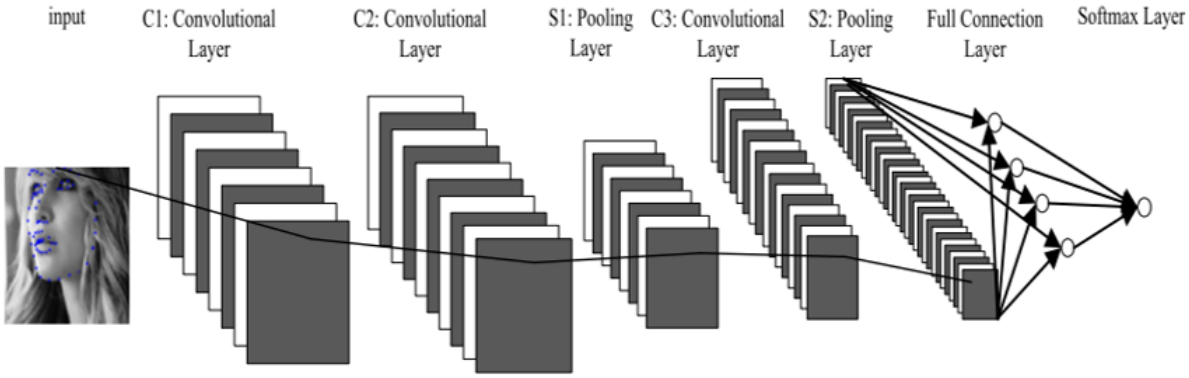


Figure 2.7: Convolutional Neural Network

2.3 Hardware & Software Requirements

Processor	: Intel Core i5 7th Gen 3.2GHz
Supporting Software	: Python
RAM	: 16GB
Graphics Card	: 6GB NVIDIA GeForce GTX 1060 GPU
Operating System	: Any Operating System

Chapter 3

Project Progress

Below are the work done so far:

1. Studied the reference papers
2. Analyzed the relevance and importance of the project
3. Studied the methodologies used in the project
4. Identified the hardware and software tools for the project implementation
5. Studied the basics of python packages going to be used
6. Collected the face emotion data set
7. Made the design

3.1 Work Schedule

Work assigned for period (26-04-2021 to 01-05-2021)

- Identify suitable project domain

Work assigned for period (02-05-2021 to 11-05-2021)

- Identify different project topics based on the domain selected

Work assigned for period (12-05-2021 to 22-05-2021)

- Identify project topic based on proper journals and other references

Work assigned for period (23-05-2021 to 31-05-2021)

- Initial conformation and selection of guide

Work assigned for period (01-06-2021 to 10-06-2021)

- Study of references and literature survey for the project

Work assigned for period (11-06-2021 to 20-06-2021)

- Detailed study of methods in the project for image processing and neural network development

Work assigned for period (01-08-2021 to 10-08-2021)

- Prepare the design of the project

Work assigned for period (11-08-2021 to 20-08-2021)

- Design presentation
- Obtaining conformation to start implementation

Work assigned for period (21-08-2021 to 31-08-2021)

- Start implementation
- Complete the implementation of the facial emotion image pre-processing methods

Work assigned for period (01-09-2021 to 10-09-2021)

- Complete the development and training of convolutional neural network
- Complete the testing and performance analysis of convolutional neural network
- Analyze overall performance of the project
- Prepare final presentation

References

[1] Hongli Zhang, Alireza Jolfaei, and Mamoun Alazab. A face emotion recognition method using convolutional neural network and image edge computing. IEEE Access, 7:159081–159089, 2019

[2] An Chen, Hang Xing, and Feiyu Wang. A facial expression recognition method using deep convolutional neural networks based on edge computing. IEEE Access, 8:49741–49751, 2020