IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 8 Issue 8, August 2021
ISSN (Online) 2348 – 7968 | Impact Factor (2020) – 6.72
www.ijiset.com

Facial Expression Detection Using Convolution Neural Network Method

[1]Archana dubey, [2]Rohini Deshpande
[1]Student, K.J.Somaiya College of Engineering,Mumbai-400077
Archana.sd@somaiya.edu

[2]Professor, K.J.Somaiya College of Engineering,Mumbai-400077
rohinideshpande@somaiya.edu

Abstract

In this project, we have developed convolutional neural networks (CNN) for a facial expression recognition. The goal is to classify each facial image into one of the seven facial emotion categories. We trained CNN models with different depth using gray-scale images from the Kaggle website. we develop our models in python and and exploited Graphics Processing Unit (GPU) computation in order to expedite the training process. we trained a novel CNN model with the combination of raw pixel data and Histogram of Oriented Gradients (HOG) feature. To reduce the overfitting of the models, we utilized different techniques including dropout and batch normalization in addition to L2 regularization. We applied cross validation to determine the optimal hyper-parameters and evaluated the performance of the developed models by looking at their training histories. We also present the visualization of different layers of a network to show what features of a face can be learned by CNN models.

Keywords

CNN,GPU, HOG, Oerfitting, regularization.

1. Introduction

Facial expressions are essential to human social communication, as this communication is both verbal and non-verbal. Facial expressions are one aspect of non-verbal communication, as the face expresses prominent signals of communication, which includes eye contact. Other aspects of non-verbal communication are gestures and body language. It is easy for humans to notice and understand faces and facial expressions. There are several problems related to this issue, such as the detection of an image segment as an actual face, due to occlusions or illumination, as well as variations in head poses, extraction of facial expression information, facial landmark detection, or classification of expression. Facial Expression Recognition (FER) is an active research area in the field of Artificial Intelligence and applied in vast domains, such as security, monitoring and law enforcement, marketing and entertainment, e-learning and medicine, emotionally intelligent robotic interfaces, or social humanoid robots. Various fields, like data analytics, psychological research, social gaming, and others that include human-computer interactions, can benefit from the ability to recognize facial expressions automatically. The feature extraction part consists of four convolutional layers, each two followed by max-pooling layer and rectified linear unit (ReLU) activation function.

They are then followed by a dropout layer and two fully-connected layers. The spatial transformer (the localization network) consists of two convolution layers (each followed by max-pooling and ReLU), and two fully-connected layers. The objective of this paper is to develop a novel architecture, from scratch, to classify images of human faces into discrete emotion categories using CNNs, also illustrate the pre-processing and feature extraction techniques to further improve the accuracy on the FER2013 dataset.

We implemented the aforementioned model in python and took advantage of GPU accelerated deep learning features to make the model training process faster.

2. Literature Review

Shekhar Singh and Fatma Nasoz proposed Facial Expression Recognition with Convolutional Neural Networks where we demonstrate the classification of FER based on static images, using CNNs, without requiring any pre-processing or feature extraction tasks. This also illustrates techniques to improve future accuracy in this area by using pre-processing, which includes face detection and illumination correction. Feature extraction is used to extract the most prominent parts of the face, including the jaw, mouth, eyes, nose, and eyebrows. Facial expressions are one aspect of non-verbal communication. n this section, we discuss the challenges and future work that could be done to further improve the test accuracy on the FER2013 dataset[1].

Shima Alizadeh and Azar Fazel proposed Convolutional Neural Networks for Facial Expression Recognition. In this project, we have developed convolutional neural networks (CNN) for a facial expression recognition task. The goal is to classify each facial image into one of the seven facial emotion categories considered in this study. We trained CNN models with different depth using gray-scale images from the Kaggle website [2].

Shan Li and Weihong Deng proposed Deep Facial Expression Recognition: A Survey where we study whole detail about facial expression in different method.deep neural networks have increasingly been leveraged to learn automatic FER. We survey the issues and applications and challenge and opportunities in field of FER System. For the state of the art in deep FER, we review existing novel deep neural networks and related training strategies that are designed for FER based on both static images and dynamic image sequences, and discuss their advantages and limitations[3].

Shervin Minaee , Amirali Abdolrashidi proposed Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network has been an active research area over the past few decades, and it is still challenging due to the high intra-class variation. In this work, we propose a deep learning approach based on attentional convolutional network, which is able to focus on important parts of the face, and achieves significant improvement over previous models on multiple datasets, including FER-2013, CK+, FERG, and JAFFE. We also use a visualization technique which is able to find important face regions for detecting different emotions, based on the classifier’s output[4].

Sandeep Kumar Ramani Proposed Facial Expression Detection using Neural Network for Customer Based Service. In this paper, the images of customer's face are used to detect the facial expression in them to enhance the customer based services. The convolution neural network using the VGG-16 architecture is used as the deep learning model which extracts essential features in an image and enables us to recognize the expression of the customer. A comparison of the two proposed model was made based on the speed and accuracy and found the computational resources tradeoffs required [5].

3. Design and Concept

The design of the system that will be created has several stages of detection of facial expressions as shown in Fig.1

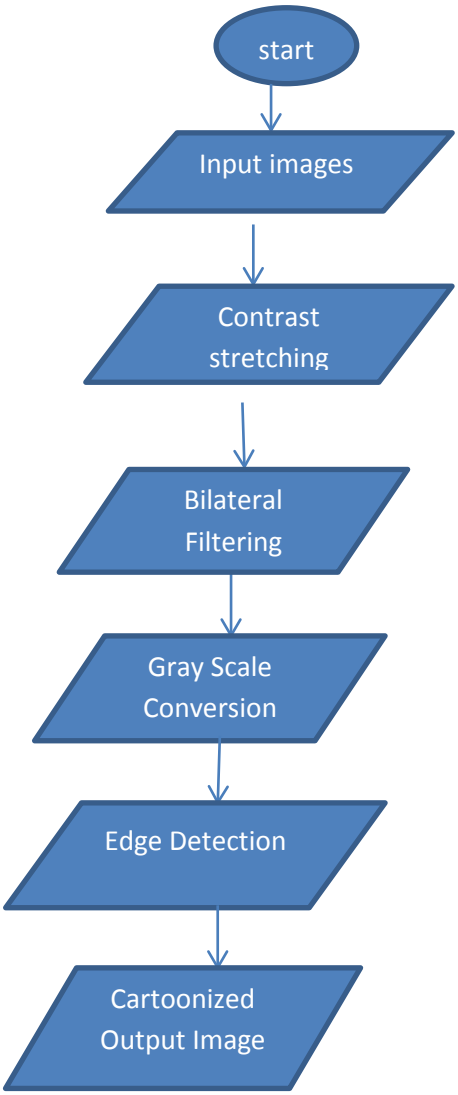


Fig.1 Flow Chart of Expresssion detection

Input Images : Take a input images from Google Or any Images

Preprocessing :

1) Grayscale to change the color image RGB (Red Green Blue) which interprets bits between 0 - 255 converted to gray scale. Changing grayscale luminosity is by multiplying the value specified by the RGB value, then doing the addition. The

formula applied:

Grayscale = (R*0,213 + G*0,715 + B*0,072)

Where the RGB value represents the number of Red Green Blue values in an image.

2) Resize the photo to change the scale of the photo to be a small size to facilitate the computation process of 48 x 48 pixels, change the size to make it easier in inputting the image and convolution into several features until it reaches 1x1 output so it cannot be extracted anymore.

3) Saturation is the intensity of hue by changing the scale of the image will experience a decrease in quality, so it is necessary to saturate to change the image that is to be more stable with higher intensity.

R = 0,213f * invSat; G = 0,715f * invSat; B = 0,072f * invSat;

Feature Extraction :

TensorFlow TensorFlow is a library for deep learning, especially neural networks with many layers and various open source topologies developed by Google. The complex problems in the image recognition feature with this framework accelerate the training image process.

Convolutional Neural Network (CNN) Convolution is form of matrix for filtering cases. This CNN approach combines 3 main architectures , local receptive fields, shared weight in the form of filter dan spatial subsampling in the form pooling [13]. Some layers in filtering are Convolutional Layer, Poling Layer dan Fully Connected Layer. The CNN architecture is illustrated in Fig.2

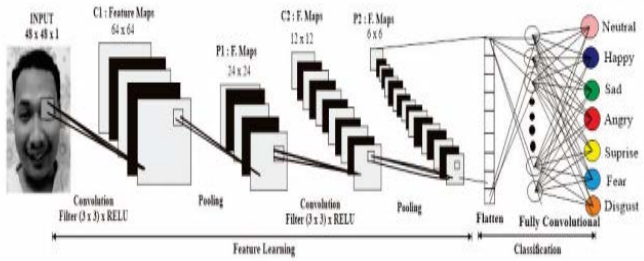


Fig.2 Convolutional Neural Network Architecture

The parameters in the Convolution Layer have several parameters, namely kernel skipping factors and connection factor. The kernel in CNN always shifts to the area in the input image,

while Skipping factor is the number of pixels that shift in the kerne .The size of the output on the map.

Formula :

$$a_j=(\frac{max}{n\,x\,n}\,a_i{}^{n\,x\,n}\mu(n,n))$$

Information :

$a_j = value\ dari\ pooling\ map$
$a_j = value\ dari\ input\ map$
$\mu(n,n) = window\ function$

The following is a sample process from Max Pooling with a size of 2 x 2

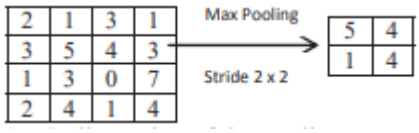


Fig 3. Illustration of the Pooling Layer Process

4. Methodology

We developed CNNs with variable depths to evaluate the performance of these models for facial expression recognition.

Given a face image, it is clear that not all parts of the face are important in detecting a specific emotion, and in many cases, we only need to attend to the specific regions to get a sense of the underlying emotion.

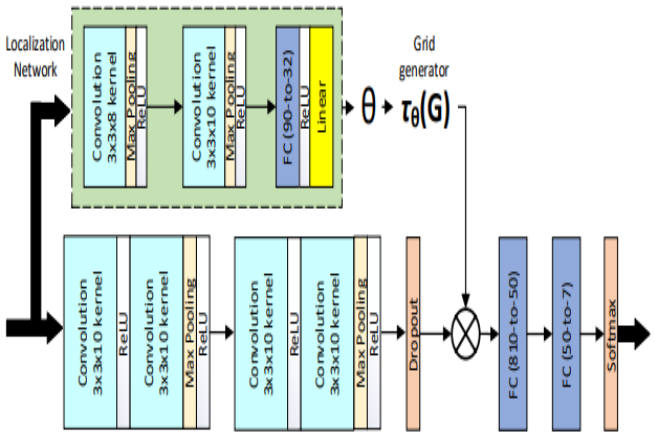


Fig.4 The Proposed model architecture

Based on this observation, we add an attention mechanism, through spatial transformer network into our framework to focus on important face regions. Figure 3 illustrates the proposed model architecture. The feature extraction part consists of four convolutional layers,

each two followed by max-pooling layer and rectified linear unit (ReLU) activation function. They are then followed by a dropout layer and two fully-connected layers. The spatial transformer (the localization network) consists of two convolution layers (each followed by max-pooling and ReLU), and two fully-connected layers. After regressing the transformation parameters, the input is transformed to the sampling grid T(θ) producing the warped data. The spatial transformer module essentially tries to focus on the most relevant part of the image, by estimating a sample over the attended region. One can use different transformations to warp the input to the output, here we used an affine transformation which is commonly used for many applications.

This model is then trained by optimizing a loss function using stochastic gradient descent approach (more specifically Adam optimizer). The loss function in this work is simply the summation of two terms, the classification loss (cross-entropy), and the regularization term .

5. Dataset and Features

we used a dataset provided by Kaggle website, which consists of about 37,000 well structured 48 × 48 pixel gray-scale images of faces.

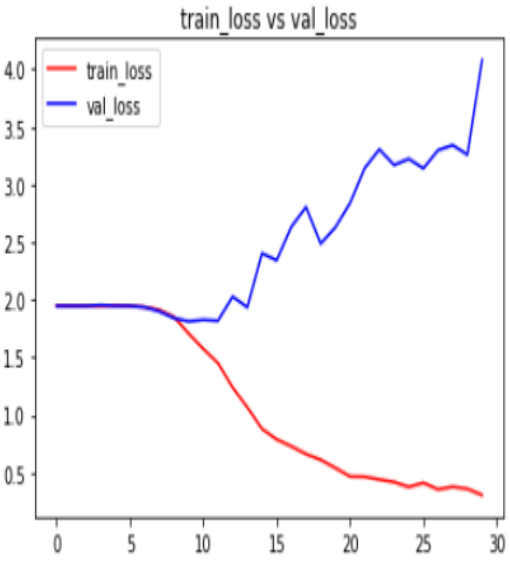
The images are processed in such a way that the faces are almost centered and each face occupies about the same amount of space in each image. Each image has to be categorized into one of the seven classes that express different facial emotions. These facial emotions have been categorized as: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, and 6=Neutral. There are about 29,000 training images, 4,000 validation images, and 4,000 images for testing.

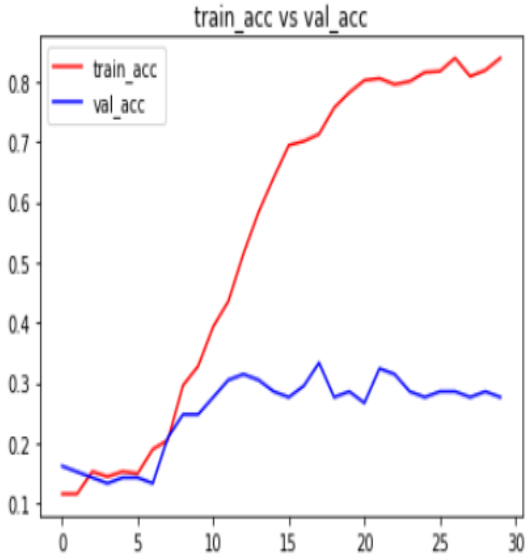
After reading the raw pixel data, we normalized them by subtracting the mean of the training images from each image including those in the validation and test sets. For the purpose of data augmentation, we produced mirrored images by flipping images in the training set horizontally. In order to classify the expressions, mainly we used the features generated by convolution layers using the raw pixel data. As an extra exploration, we developed learning models that concatenate the HOG features with those generated by convolutional layers and give them as input features into Fully Connected (FC) layers.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 128, 128, 6)	456
activation (Activation)	(None, 128, 128, 6)	0
max_pooling2d (MaxPooling2D)	(None, 64, 64, 6)	0
conv2d_1 (Conv2D)	(None, 64, 64, 16)	2416
activation_1 (Activation)	(None, 64, 64, 16)	0
max_pooling2d_1 (MaxPooling2	(None, 32, 32, 16)	0
conv2d_2 (Conv2D)	(None, 6, 6, 120)	48120
activation_2 (Activation)	(None, 6, 6, 120)	0
dropout (Dropout)	(None, 6, 6, 120)	0
flatten (Flatten)	(None, 4320)	0

Validation losses and Accuracy :




train_acc vs val_acc
train_acc
val_acc
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
15
20
25
30

Results :

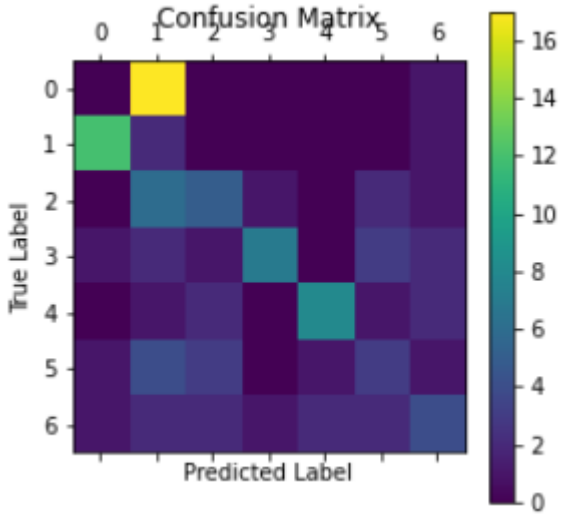

prediction = ANGRY
prediction = FEAR
prediction = NEUTRAL


prediction = NEUTRAL
prediction = SURPRISE
prediction = SURPRISE


prediction = DISGUST
prediction = FEAR
prediction = SURPRISE

Confusion Matrix :

FER-2013 dataset is more challenging than other facial expression recognition datasets we used. Besides the intra-class variation of FER, another main challenge in this dataset is the imbalance nature of different emotion classes. Some of the classes such as happiness and neutral have a lot more examples than others. We used the entire 28,709 images in the training set to train the model, validated on 3.5k validation images, and report the model accuracy on the 3,589 images in the test set. We were able to achieve an accuracy rate of around 70.02% on the test set. The confusion matrix on the test set of FER dataset is shown in Figure. As we can see, the model is making more mistakes for classes with less samples such as disgust and fear. [0:Anger,1:digust,2:fear,3:happy,4:neutral,5:sad,6:surprise]



Conclusion :

We developed CNN's for facial expression recognition problemand evaluated their performance using different techniques. This paper proposes a new framework for facial expression recognition using an attentional convolutional network. We believe attention is an important piece for detecting facial expressions, which can enable neural networks with less than 10 layers to compete with (and even outperform) much deeper networks for emotion recognition. We also provided an extensive experimental analysis of our work on four popular facial expression recognition databases, and showed promising results.

we discuss the challenges and future work that could be done to further improve the test accuracy on the FER2013 dataset. We will also include pre-processing and feature extraction techniques, discussed in the technical work section, in order to achieve better accuracy.

Reference :

[1] Shekhar Singh and Fatma Nasoz based on "Facial Expression Recognition with Convolutional Neural Networks" 2020 IEEE conference.

[2] Shima Alizadeh and Azar Fazel “Convolutional Neural Networks for Facial Expression Recognition” 2017.

[3] Shan Li and Weihong Deng “Deep Facial Expression Recognition: A Survey” 2020 IEEE.

[4] Shervin Minaee , Amirali Abdolrashidi “Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network” 4 feb 2019.

[5] Sandeep Kumar Ramani “Facial Expression Detection using Neural Network for Customer Based Service” 2nd International Conference on Computer, Communication, and Signal Processing (ICCCSP 2018).

[6] Chao Liu, “Face Emotion Classification and Recognition using Convolutional Neural Network”.

[7] Vera Wati, Kusrini, Hanif Al Fatta, “Real Time Face Expression Classification Using Convolutional Neural Network Algorithm” 2019 International Conference on Information and Communications Technology (ICOIACT).