Facial Expression and Emotion detection

Abdirahman Abdullahi

*Kulliyyah of information communication and technology, International Islamic University Malaysia,*

abdomursal@gmail.com

MHD Khaled Maen

*Kulliyyah of information communication and technology, International Islamic University Malaysia,*

mk.maen93@gmail.com

***Abstract*—Facial expression and Emotion detection are playing an important role in today technology. Facial expression and emotion detection can be used in many application such as in online learning can be useful for the academic purpose, In this paper, we aimed to experiment on a dataset which is image-based, we have downloaded a dataset from Kaggle library, and then we used CNN (conventional neural network) as our algorithm. This paper provides a basic knowledge of the facial expression and emotion detection using Deep learning.**

1. Introduction

Facial expression and emotion detection are essential aspects in human communication that help people to understand each other, not only human but also the machine must understand our emotions in terms to help us to discover certain emotion that is difficult from human to recognise it. However, people infer the emotional states of other people, such as joy, sadness, and anger, using facial expressions and emotion detection. it is obvious the study of facial expression has become the centre of attention over the past decades in affective computing and computer animations.

1. Related Work
2. *Deep Learning*

Deep learning is a machine learning subfield that deals with algorithms based on the structure and function of the brain called artificial neural networks. In other words, it mirrors the brain's functioning. Deep learning algorithms are similar to the structure of the nervous system in which each neuron connects and passes information. Deep learning models work in the layers and three layers of a typical model at least. Each layer accepts and passes the information from the previous layer to the next layer.

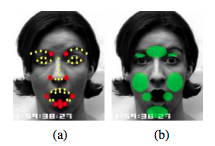
1. *Convolutional Neural Network*

Convolutional Neural Networks are very similar to ordinary Neural Networks, they are made up of neurons that have learnable weights and biases. The architectures of ConvNet make the explicit assumption that the inputs are images which will be encoded to certain properties in the architecture. This makes the forward function more efficient to implement and reduces the number of parameters in the network considerably.

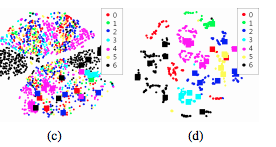
Convolutional Neural networks allow computers to see, meaning that Convnets is used to recognize images by transforming the original image into a class scoring through layers. CNN was inspired by the visual cortex. Every time we see something, a series of layers of neurons get activated, and each layer will detect a set of features such as lines, edges. The high level of layers will detect more complex features in order to recognize what we saw.

1. *Previous Work*

this paper [1] represented a reduction effort on facial expression using deep learning which is regarded as automatically extract useful features from raw data. [1] used two different deep learning models, The first deep network called DTAN that extracts temporal appearance features from image sequences. However, the other deep network used called DTGN which to extract temporal geometry features from temporal facial landmark points. These two models combined together using integration method in terms to boots the performance of facial expression recognition. the DTAN in the first layer have the ability to obtain the difference between the input frames. Furthermore, the important landmark points extracted by DTGN were also shown.

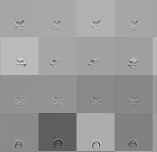


Visualization of DTGN in figure (a) and (b):



 Visualization of DTAN in figure (c) and (d)

 In this work [2] showed that convolutional neural networks (CNNs) can achieve strong performance and it showed that which part of face exactly has been influenced the CNNs prediction. In this paper [2] trained the CNNs with zero-bais on both quantitatively and qualitative to do emotion recognition that is able to perform a model with high-level features. Quantitatively showed which part of the face has most discriminate information. Meanwhile, quality correlates the numerical activations of the visualized filter using facial movement FAU labels.



Visualization of spatial patterns that activate 10 selected filters

1. Methodology

In order to get a model with satisfying results, we setup an environment with a powerful computational resources and large datasets.

* 1. *Environment*

GoogleCloudPlatform **(**GCP**)** is a cloud computing service provided by Google which provides a lot of products including cloud computing.

Google Compute Engine (GCE) is the Infrastructure as a Service (IaaS) enables users to setup Virtual Machines (VMs)  can be accessed via the developer console, command line interface (CLI) and RESTful API [5].

The VM instance of choice  run on an Ubuntu operating system version 16.04. Alongside, the VM instance is powered by 8 Graphics Processing Units (GPUs) of type NVIDIA Tesla K80  which engineered to boost throughput in real-world applications by 5-10x times, compared to Central Processing Unit (CPU).

* 1. *Dataset*

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression in to one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral) as shown in Fig. 1.

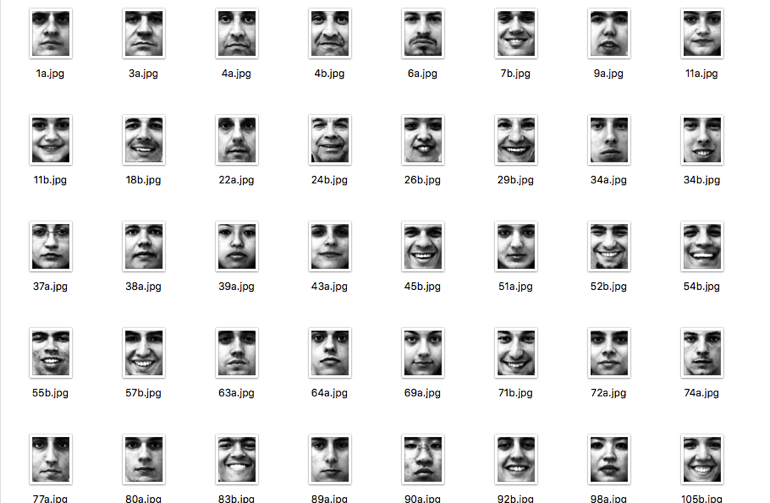


Fig. 1 A sample images of the dataset

* 1. *Algorithm*

The algorithm used in this experiment is 10 Convolutional neural network layers, each layer followed by batch normalization layer, activation function pooling layer and drop-out layer. The first two convolutional layers compose of 16 filters size of (7x7), for second two layers 32 filters size of (5x5), the third two layers 64 filters size of (3x3), followed by another two layers with 128 filters size of (3x3) then one layer with 256 filters of size(3x3), for the final convolutional layer number of filters is equals to number of classes and all convolutional layers have the same padding. The drop-out value was relatively big since after each step 50% of neurons being drop to overcome overfitting problem. Back to the activation function used in this experiment rectified linear unit(ReLU) except for the last layer which is SoftMax.

1. Results

REFERENCE

1.   H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, “Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition,” 2015 IEEE International Conference on Computer Vision (ICCV), 2015.

2.   “ICCV 2015 Area Chairs,” *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015.