A Project Report on

Emotion Recognition_

By

Aditya Sable

Atharva Vaidya

Shubham Pawar

Under the Guidance of

Prof. Jaya Gupta

Department of Computer Engineering
A.P. Shah Institute of Technology
G.B.Road, Kasarvadavli, Thane(W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year 2018-2019

Approval Sheet

This Project Report entitled *“Emotion Recognition”* Submitted by *“Aditya Sable”(16102020),“AtharvaVaidya”(16102043),“Shubham Pawar”(16102035)* is approved for the partial fulfillment of the requirement for the Mini Project .

Prof. Jaya Gupta
Guide

Prof. Sachin Malve
Head Department of
Computer Engineering

Place :A.P.Shah Institute of Technology, Thane
Date:11/04/2019

CERTIFICATE

This is to certify that the project entitled *“Emotion Recognition”* Submitted by *“Aditya Sable”(16102020),“Atharva Vaidya”(16102043),“Shubham Pawar”(16102035)* for the partial fulfillment of the requirement for Mini Project is a bonafide work carried out during academic year 2018-2019.

Prof. Sachin Malve

Head Department of Computer Engineering

Prof. Jaya Gupta
Guide

External Examiner(s)

1.

2.

Place : A.P.Shah Institute of Technology, Thane

Date:11/04/2019

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified a idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

——————————
——————————

——————————————————

(*Aditya Sable, 16102020)*
(*Atharva Vaidya, 16102043*)
(*Shubham Pawar, 16102035*)

Date: 11/04/2019

ABSTRACT

In this paper we propose an implement a generalconvolutional neural network (CNN) building framework fordesigning real-time CNNs. We validate our models by creating a real-time vision system which accomplishes the tasks offace detection, gender classification and emotion classification

Simultaneously in one blended step using our proposed CNNarchitecture. After presenting the details of the training procedure setup we proceed to evaluate on standard benchmark sets. We report accuracies of 96% in the IMDB gender dataset and 66% in the FER-2013 emotion dataset. Along with this we also introduced the very recent real-time enabled guided back-propagation visualization technique.

Guided back-propagation uncovers the dynamics of the weight changes and evaluates the learned features. We argue that the careful implementation of modern CNN architectures, the use of the current regularization methods and the visualization of previously hidden features are necessary in order to reduce the gap between slow performances and real-time architectures.

Contents

1 Introduction 1

1.1Machine learning 1

1.1.1Supervised Learning 2

1.2 Convolutional Neural Network 2

2 Literature Review 4

2.1 Minixception Architecture 4

3 Data Preprocessing 5

4 Result 6

5 Conclusions and Future Scope 7

6 Bibliograph 8

Keywords

- CNN
- emotion recognition
- convolution neural network
- feature extraction
- feature selection
- pattern recognition

Chapter 1

Introduction

An face emotion recognition system comprises of two step process i.e. face detection (bounded face) in image followed by emotion detection on the detected bounded face. The following two techniques are used for respective mentioned tasks in face recognition system.

Haar feature-based cascade classifiers : It detects frontal face in an image well.It is real time and faster in comparison to other face detector. This blog-post uses an implementation from Open-CV.

Xception CNN Model: Mini Xception

We will train a classification CNN model architecture which takes bounded face (48*48 pixels) as input and predicts probabilities of 7 emotions in the output layer.

1.1 Machine Learning

The term Machine Learning was coined by Arthur Samuel in 1959, an American pioneer in the field of computer gaming and artificial intelligence and stated that “it gives computers the ability to learn without being explicit programmed”.

And in 1997, Tom Mitchell gave a “well-posed” mathematical and relational definition that “A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

1.1.1 Supervisied Learning

supervised machine learning is one of the most commonly used and successful types of machine learning. In this chapter, we will describe supervised learning in more detail and explain several popular supervised learning algorithms. We already saw an application of supervised machine learning in classifying iris flowers into several species using physical measurements of the flowers.

Remember that supervised learning is used whenever we want to predict a certain outcome from a given input, and we have examples of input/output pairs. We build a machine learning model from these input/output pairs, which comprise our training set. Our goal is to make accurate predictions for new, never-before-seen data. Supervised learning often requires human effort to build the training set, but afterward automates and often speeds up an otherwise laborious or infeasible task.

1.2 Convolutional Neural Network

When it comes to Machine Learning, Artificial Neural Networks perform really well. Artificial Neural Networks are used in various classification task like image, audio, words. Different types of Neural Networks are used for different purposes, for example for predicting the sequence of words we use Recurrent Neural Networks more precisely an LSTM, similarly for image classification we use Convolution Neural Network. In this blog, we are going to build basic building block for CNN.

Before diving into the Convolution Neural Network, let us first revisit some concepts of Neural Network. In a regular Neural Network there are three types of layers:

1. **Input Layers:** It’s the layer in which we give input to our model. The number of neurons in this layer is equal to total number of features in our data (number of pixels incase of an image).
2. **Hidden Layer:** The input from Input layer is then feed into the hidden layer. There can be many hidden layers depending upon our model and data size. Each hidden layers can have different numbers of neurons which are generally greater than the number of features. The output from each layer is computed by matrix multiplication of output of the previous layer with learnable weights of that layer and then by addition of learnable biases followed by activation function which makes the network nonlinear.
3. **Output Layer:** The output from the hidden layer is then fed into a logistic function like sigmoid or softmax which converts the output of each class into probability score of each class.

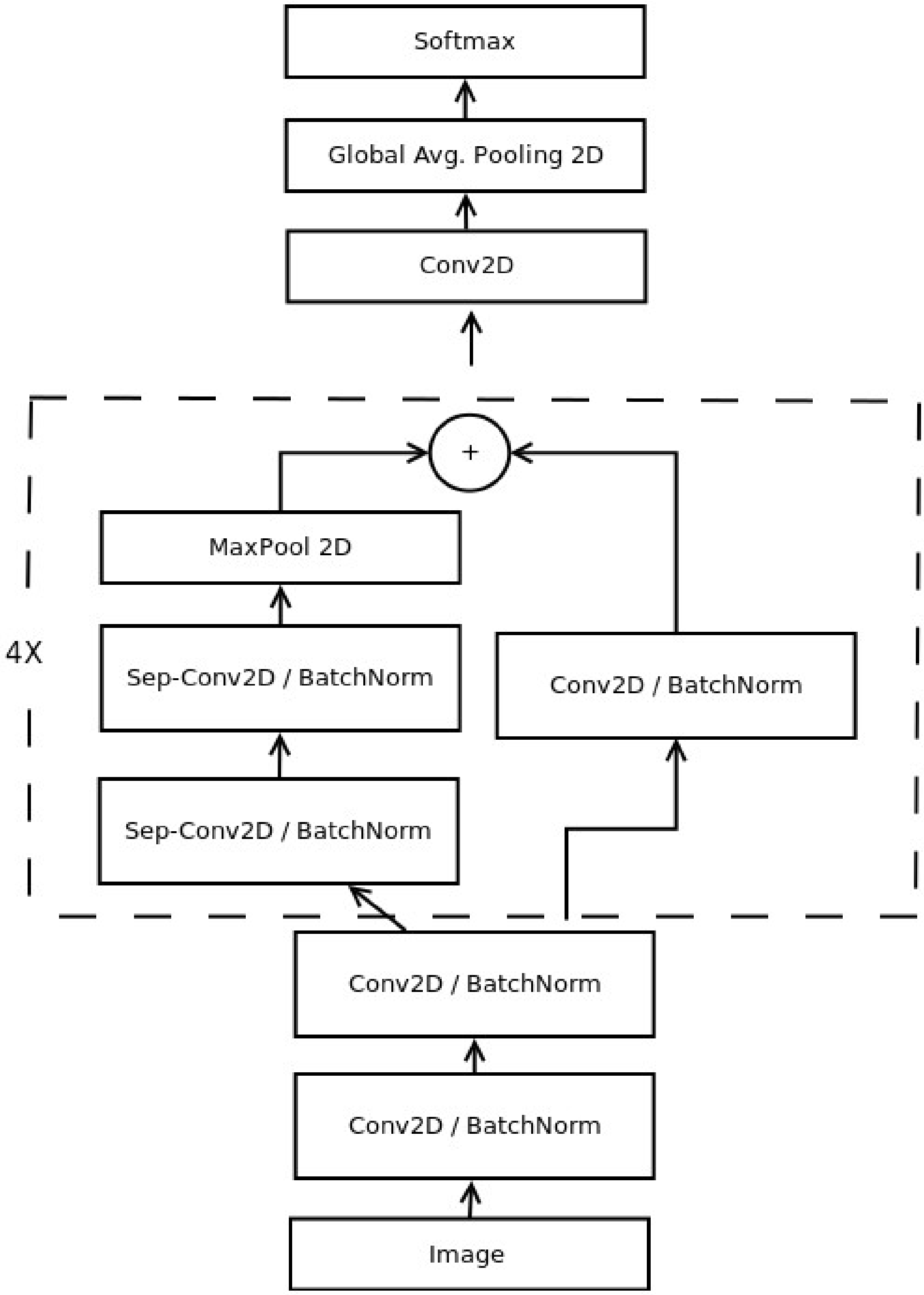
Chapter 2

Literature Review

A. Image and voice are the most direct and most natural channels that people acquire information. If achievements in these two fields are applied on robots that can greatly improve the intelligence of the machine. In practice, in image recognition and speech recognition we will encounter the feature selection problem. Common image features are composed of color feature, texture feature, shape feature, spatial relations characteristics. The commonly used features in speech recognition are MFCC (Mel Frequency Cepstrum Coefficient), prosodic features, sound quality characteristics and acoustic features. Sometimes in order to acquire a better final result, these characteristics also be integrated appropriately

B. CNN is a specially designed multi-layer perceptron to identify two-dimension shapes. Therefore dimensional information retained in waveform points is effectively utilized by CNN. CNN model due to its characteristics of adaptive feature extraction, it is applied for image recognition and emotion recognition in voice signals. In the emotional speech recognition, based on the test of two classic characteristics of the speech signal, we propose that directly use waveform points to characterize the emotional speech signals. It neither loss information, but also take advantage of the natural correlation information between the waveform to identify emotion. In image recognition, SVM and CNN models are used for image recognition. And we compare the recognition result before and after PCA

MiniXception Architecture



Chapter 3

Data Pre-processing

We have divided the project in four parts :

1. Data pre-processing

2. Feature Engineering

3.Training the model

4.Testing the model

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm.

Data Preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

Chapter 4

Result

Results of the real-time emotion classification task in un-seen faces can be observed. Our complete real-time pipeline including: face detection and emotion classification.

Total Parameters: 58423

Trainable Parameters: 56951

Non Trainable Parameters: 1472

Achieved Accuracy (on Mini Xception Model): 60%

Maximum Achieved Accuracy (on Mini Xception Model): 65%-66%

Maximum Achieved Accuracy (on FER(2013)): 98% by Resnet Model

Chapter 5

Conclusions and Future Scope

Machine learning models are biased in accordance to their training data. In our specific application we have empirically found that our trained CNNs for gender classification are biased towards western facial features and facial accessories. We hypothesize that this misclassfications occurs since our training dataset consist of mostly western: actors, writers and cinematographers as observed.

Furthermore, the use of glasses might affect the emotion classification by interfering with the features learned. We believe that uncovering such behaviours is of extreme importance when creating robust classifiers, and that the use of the visualization techniques such as guided back-propagation will become invaluable when uncovering model biases.

Bibliography

[1] François Chollet. Xception: Deep learning with depthwise separable convolutions. CoRR, abs/1610.02357, 2016.

[2] Andrew G. Howard et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. CoRR, abs/1704.04861, 2017.

[3] Dario Amodei et al. Deep speech 2: End-to-end speech recognition in english and mandarin. CoRR, abs/1512.02595, 2015.

[4] Ian Goodfellow et al. Challenges in Representation Learning: A report on three machine learning contests, 2013.

[5] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 315–323, 2011.

Acknowledgement

We have great pleasure in presenting the report on *Emotion Recognition* **.**We take this opportunity to express our sincere thanks towards our guide **Prof. Jaya Gupta** Department of Computer Engineering , APSIT thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of project.

We thank **Prof. Sachin Malve** Head of Department, Computer Engineering, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Student Name1:Aditya Sable
ID:16102020

Student Name2:Atharva Vaidya
ID:16102043

Student Name3:Shubham Pawar
ID:16102035