Proposed Implementation for Fast Face Detection and Recognition

Oscar Chen
Dept of Elec & Comp Engineering
University of Calgary
Calgary, Canada
oscar.chen1@ucalgary.ca

Savith Jayasekera
Dept of Elec & Comp Engineering
University of Calgary
Calgary, Canada
scjayase@ucalgary.ca

Prince Okoli
Dept of Elec & Comp Engineering
University of Calgary
Calgary, Canada
prince.okoli@ucalgary.ca

***Abstract*—We propose a combined implementation using Viola-Jones feature-based detector with deep learning techniques, to detect, recognize and differentiate faces from streaming video footage. We want to leverage Viola-Jones feature based detector for fast detection and capture of human faces from video footage, and further process the facial images for recognition by treating it as a regression deep learning problem.**

***Index Terms*—deep learning, convolutional neural network, siamese neural network, face detection, facial recognition, Viola-Jones algorithm, Harr-like features.**

I. Introduction

Facial detection and recognition is a relative old area in computer vision. Convolutional neural network (CNN) has been used in the past to achieve good object detection capabilities. With large enough sample set, CNN can perform classifying and generalizing tasks with high accuracy as compared to many other techniques. However CNN is traditionally slow and unsuited for performing classifying tasks on live streaming video footage. CNN also requires large amount of images to train for each class, therefore making it unfeasible as a classifier for facial recognition. By treating facial recognition as a regression, one can assign a face to a vector and thus not requiring training on large set of images for each face. When a Viola-Jones feature detector is used in conjunction with such a CNN, we believe that fast facial recognition and differentiation can be achieved without prior training on each new test subject.

II. Related Work

1) FaceNet: A Unified Embedding for Face Recognition and Clustering https://arxiv.org/abs/1503.03832 The paper proposes a system to implement face verification and recognition efficiently at scale FaceNet creates a mapping to a Euclidean space where distance corresponds to the facial similarity. The output of the FaceNet can be used as a feature vectors for tasks such as face recognition, verification and clustering.

2) Names and faces in the news https://ieeexplore.ieee.org/document/1315253 The paper shows the results of face clustering with a face image dataset containing ambiguous or inaccurate labels. The goal is to show the results of a clustering model trained using "in the wild" images that are more realistic compared to the usual training images.

3) Face detection without bells and whistles http://rodrigob.github.io/documents/2014_eccv_face_detection_with_supplementary_material.pdf The paper proposes a face detection model based on Deformable Parts Model (DPM) that yields high performance results; the paper also discusses a face detector based on rigid templates similar to the Viola & Jones detector. Additionally, the paper discusses issues with evaluation benchmarks and proposes an improved procedure.

III. Methodology

For the purpose of fast face detection in streaming video, we want to implement a Haar Cascade feature detector trained on large set of images of people, where the facial regions have been labeled, we plan to mainly rely on the IMDB full-body images for this part of the training.

For the purpose of face recognition, we want to implement Siamese neural networks to compare and distinguish facial features. Such networks will be trained to minimize differences of the same person's face from different images, while maximize the differences of different peoples' faces. A well-train siamese neural network will be capable of distinguishing people's facial features whose images have not been previously trained on, therefore making it a perfect fit for facial recognition without having to train on each new face. To train this network, we will use IMDB cropped facial images as well as several other smaller data sets.

We will be primarily using machine learning libraries/frameworks available for Python such as Tensorflow for deep learning, OpenCV for image processing, and Spark for parallel processing.

A. Research Questions

Will this work? Maybe. Will it not work? Probably. We may try and fail, or we may not try and wonder what would have happened if we tried. To try, or not to try, that is the question which is not up for debate, without trying we will never know.

B. Data Collection and Preparation Plan

We will be mainly using the IMDB-WIKI face images with labels from a 2015 study on age prediction using deep learning. The images are labelled in MatLab format, we will export and manipulate the label data into suitable xml format.

The data labels include names, age, gender of the people in the images, as well as the coordinates of their face within the image. We do not expect to do laborious manual data labelling, but if required we will be using LabelImg.

C. Data Analysis Plan

Since the IMDB-WIKI image set comes with comprehensive data labels, we plan to use this data set for testing and evaluation purpose. The facial recognition accuracy can be treated as a classification problem and evaluated using k-fold cross validation.

IV. Expected Results

Face detection and face recognition are both well-researched areas, similar algorithms have been tried and trued in both applications. We expect this implementations to work to reasonably high accuracies in each of the two areas. The speed is the concern, we want to implement this for video stream processing, which is 30 frame per second with potentially multiple masks.

V. Discussion

A. Limitations

Motion tracking of subjects in the video stream is not being considered in this implementation. The frame-to-frame movement and change of angle of the images present important temporal information which is not being considered by the convolutional network. The frame-to-frame object correlation could be a function in the detection module, which would potentially improve the accuracy of recognition by evaluating multiple frame of the same subject.

B. Possible Extensions [Future Work]

Parallel implementations of age, behavioural prediction, and other pattern recognizing models is possible in the same framework, as well as a recurrent neural network to assist in analyzing temporal data captured as the subjects move and their face angles change in a video stream, which could provide further accuracy improvement in face recognition.