TOPIC: COMPARATIVE ANALYSIS OF MODEL FOR MULTI-TASK LEARNING FOR AGE GROUP ESTIMATION AND GENDER RECOGNITION USING FACIAL FEATURES

MODELS: CONVOLUTIONAL NEURAL NETWORK (CNN) AND SUPPORT VECTOR MACHINE (SVM)

Akinrele Taye Oluwabusayo

Computer Science Department

Email: [tayeakinrele@gmail.com](mailto:tayeakinrele@gmail.com)

ABSTRACT

Human facial image processing has been an active and interesting research issue for years. Since human faces provide a lot of information, many topics have drawn lots of attentions and thus have been studied intensively.

**CHAPTER ONE**

**INTRODUCTION**

* 1. **BACKGROUND OF STUDY**

The human face conveys much information, which people have an astonishing ability to extract, analyze and decipher (*Ali Maina, et al, 2016*). The main characteristic feature of the human being is the face (*Sayantani, 2015*), which exhibits different emotions that can be determined and easily predicted by the several facial expressions. Just by glancing at a person’s face, one can estimate or predict the age and gender of that person. Identifying human faces and modeling the distinguishing features of human faces that contribute most toward face recognition are some of the challenges faced by computer vision and psychophysics researchers (*M. R. Dileep & Ajit Danti, 2018*).

Age and gender are significant properties regarded as a crucial biological characteristics, which plays a fundamental role in human social interaction. The human face contains a wide range of information for gender perception and age estimation. (*Ke Zhang et al, 2019*). Classification of age and gender is an important visual task for human beings, since many social interactions critically depend on the correct age and gender perception.

As technologies such as visual surveillance and human computer interaction evolve, computer vision systems for age and gender classification plays an important role in our lives, it is therefore not surprising that a lot of researches has been done to investigate age and gender classification from face perception in humans, proposing various methods in order for a machine to attain human level of accuracy (*Md. Nurul & Emon Kumar, 2018*). Predictably, it is challenging for machines to identify these visual information, since discriminative feature extraction is easily affected by various factors like large variations in facial gestures, lighting, background, etc. Perceived age and gender classification is a topic with a high application potential in areas like surveillance, face recognition, video indexing and dynamic marketing surveys.

This research attempts to compare two multi-task learning models Convolutional Neural Network (CNN) and Support Vector Machine (SVM), in terms of performance while classifying humans into age groups and gender. The proposed approach treats gender prediction and age estimation as a classification problem, adopting datasets which composes of images labeled for age and gender. It uses eight classes of age group (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+) and classifies gender into one of the two sex labels, Male (M) and Female (F).

* 1. **RESEARCH MOTIVATION**

The primary motivation behind this research is due to the observation that the amount of data available for the study of computer vision problem, can have an enormous impact on the machine capabilities developed to solve it (*Prajakta, Dr. G. S. Sable, 2018*). In view of this, this approach makes use of different datasets to train two multi-task learning models Convolutional Neural Network (CNN) and Support Vector Machine (SVM), comparing their performances, in other to generate an accurate result in the prediction of gender and age groups from input images. Some datasets that will be used in this paper includes IMDB-Wiki dataset (523,051 face images), Adience dataset (34,795 images) and UTKFace dataset (over 20,000 images).

This research aims to develop a web-based assessment system that will solve the problem of collection of data from manual sources like mobile phones or computer.

* 1. **RESEARCH OBJECTIVES**

The specific objectives of this research are to:

1. Compare these two models of multi-task learning Convolutional Neural Network (CNN) and Support Vector Machine (SVM).
2. Design a CNN and SVM model to jointly recognize the gender and age group of a person through an input image.
3. Compare the models using Adience image datasets.
4. Design and implement a web-based model evaluation system.
   1. **RESEARCH METHODOLOGY**

This research proposes an effective method for human gender and age prediction/classification from the given facial images using the Convolutional Neural Network and Support Vector Machine models. The methodology includes: data preparation, data pre-processing,

* + 1. **Data Preparation**

This is the process of transforming raw data so that data scientists and analytics can run it through machine learning algorithms to uncover insights or make predictions (*Wiki*). In this research, datasets are needed for training and testing the models. The organization of these dataset is also very essential since each of them must follow the same pattern of age group classification group (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+) and gender classification (Male and Female).

In this research, the following datasets are used to train the CNN and SVM models:

1. **IMDB-WIKI dataset**: this is the largest publicly available dataset of face images with gender and age labels for training, it is composed of 2 datasets that were created by the same authors, so they both have similar structure in terms of available metadata. It provides pretrained models for both age and gender prediction containing 460,723 images in IMDB and about 62,328 in Wikipedia, giving a total of 523,051 facial images. (*Rasmus Rothe et al, 2015*)
2. **Adience Dataset**: is composed of pictures taken by camera from smartphones or tablets. The images of this dataset capture extreme variations, including exreme blur (low-resolutions), occlusions, out-of-plane pose variations, expressions. The entire Adience dataset contains 34,795 images of 2,284 subjects. (*Jia-Hong Lee, et al, 2018*). Each image is annotated with gender and one of 8 age groups.
3. **UTKFace dataset**: is a large-scale face dataset with long age span (ranging from 0 to 116 years old). The dataset consists of over 20,000 face images with annotations of age, gender, and ethnicity. The images cover large variation in pose, facial expression, illumination, occusion, resolution (*Kaggle.com*). (*UTKFace, 2018*).
   * 1. **Data Pre-processing**

Before the face can be classified into its appropriate age group and gender, some preprocessing is needed because the classifiers are usually sensitive to imperfection of the image. Therefore, the face image is processed to obtain a transformed face image to increase the quality of the face image, retaining the important characteristics.

In this stage, brightness and contrast are normalized, the face image geometric features are improved and also the image size (number of pixels) is reduced (*Zofia Stawska, 2016*).

Pre-processing includes three steps: resizing the image, converting to gray scale and noise reduced image. The input image is resized and the color images are converted to gray scale.

* + 1. **Feature Extraction**

This is considered to be an indispensable process which is needed to be followed and implemented to accomplish the task of age and gender detection and its classification.

Extraction of features from images can be achieved by adapting the following approaches:

1. **Geometry-based Approach**: which uses geometric information such as features relative positions and sizes of the face components as a features measure. It requires finding the face characteristic points like nose, mouth, eyes, ears or hair. These points are called fiducial points.
2. **Template-based Approach:** in which previously designed standard face pattern template is used to match with the located face components. It works on the image pixels that were previously transformed on the local or global level. At the local level, image can be divided into lower windows or specific face regions such as mouth, nose or eyes. This method preserves natural geometric relationships which can be used as a naïve feature.
   * 1. **Model Implementation**

A Convolutional Neural Network and a Support Vector Machine model is implemented to perform the gender and age classification. The gender has two classes- male and female and age has eight classes- [(0 – 2), (4 – 6), (8 – 12), (15 – 20), (25 – 32), (38 – 43), (48 – 53), (60 – 100)].