Predict Age, Gender and Ethnicity

(December 2020)

First A. Author, Second B. Author, Jr., and Third C. Author, Member, IEEE

[[1]](#footnote-1)

*Abstract*—These instructions give you guidelines for preparing papers for IEEE TRANSACTIONS and JOURNALS*.* Use this document as a template if you are using Microsoft *Word* 6.0 or later. Otherwise, use this document as an instruction set. The electronic file of your paper will be formatted further at IEEE. Define all symbols used in the abstract. Do not cite references in the abstract. Do not delete the blank line immediately above the abstract; it sets the footnote at the bottom of this column.

*Index Terms*— CNN, Image Analysis, Gender recognition, Age estimation, Ethnicity determination, Expression determination.

# INTRODUCTION

T

HE human face contains a variety and veracity of semantic information about a person. The prediction of identity, age, gender, ethnicity, and expression constitutes the discipline of facial biometrics that is crucial for business campaigns, security analysis, and academic interests.

Artificial neural networks offer an efficient solution to detect and classify images of different types. Convolutional Neural Network (CNN) is a special class of artificial neural network that dominates analyzing and resolving computer vision tasks. CNN provides a framework to automatically and adaptively learn spatial features in multiple layers. This network provides us with a deep learning model to process grided data, such as, the images. CNN is based on a mathematical structure describing three types of layers or building blocks: convolution, pooling, and fully connected layers. The features inherent in the images are extracted by convolutional and pooling layers. The final output is reconstructed from the extracted features in the pooling layers [1]. In a digital image, the pixel values are stored in a two-dimensional grid, and a small grid of parameter, called kernel, is applied at each image position to extract the features from them. The extracted features are passed from one layer to another to accomplish a progressively optimized pattern recognition. This process is called training. The training is performed to nullify the difference between outputs and ground truth labels using an optimization algorithm, like, backpropagation and gradient descent.

In this project, we shall implement a simplified CNN architecture to extract age, gender, ethnicity, and expression features from a set of the images of human faces. The business domain of our application is a retail clothing store where the images of the customers are extracted using the cameras at the entrance points. Those images can be analyzed with our proposed model to extract the facial biometric data of the clients. That data can, intern, be utilized by the business teams for optimizing the products or streamlining the advertisement campaigns for the business.

The reminder of this report is organized as follows: the background and related works are reviewed in Section II. The architecture and methodology are described in Section III. After that, the Section IV will concentrate on the data, results, and discussions. Conclusion and future outlook will be presented in Section V.

# Background and Related Works

The structure of a CNN architecture is divided into multiple learning stages composed of the convolutional layers, non-linear processing units, and subsampling layers A typical block diagram of a Machine Learning (ML) system is shown in Fig 1. A comprehensive survey of the recent CNN architecture has been described by Khan et al [2] that includes a concise discussion on the basic CNN components. A typical CNN architecture is made up of alternate layers of convolution and pooling followed by one or more fully connected layers at the end. The CNN performance can be optimized by introducing mapping functions, batch normalization and dropout components. The pattern learning activities are aided by activation functions.

A comparative analysis of the age estimation techniques using deep learning was done by Othmani et al [3]. Their results demonstrated the high performance of the popular CNNs frameworks against the state-of-the-art methods of automated age estimation. A joint gender, ethnicity and age estimation from 3D dataset was performed by Xia et al [4] where the authors performed a morphology-driven analysis and emphasized on the correlation between these three demographic features. A hybrid neural network model obtained by mixing CNN and Extreme Machine Learning (ELM) architectures, implemented by Duan et al [5], exhibited about 90% accuracy in predicting age and gender from MORPH-II and Adience Benchmark datasets. The solutions of age, gender and/or race determination vary in different aspects on choosing an optimal CNN architecture and training strategy. CNN depth, pretraining, mono or multi-task training strategies, and the target age encoding and loss functions were carefully tuned by Antipov et al to accomplish as good as 100% gender and age prediction accuracies [6]. The facial expressions can also be extracted from the images using the existing frameworks, such as, OpenCV, designed for computational efficiency on real-time applications [7]. In the realm of OpenCV framework, we can extract facial expressions using the Haar Cascades method. Haar Cascades are classifiers that are used to detect features (of face in this case) by superimposing predefined patterns over face segments and are used as XML files. Different pre-defined classifiers can be generated using training dataset that can be applied to the images under experiment for a facial mood or expression determination.

Graphical user interface, application

Description automatically generated

Figure 1: Layout of a multi-stage ML system.

# Architecture and methodology

TBD

# Data, Results, and discussions

TBD

# Conclusions and future outlook

TBD

# References

|  |  |
| --- | --- |
| [1] | R. Yamashita, M. Nishio, R. K. G. Do and K. Togoshi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging,* vol. 9, pp. 611-629, 2018. |
| [2] | A. Khan, A. Sohail, U. Zahoora and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review,* vol. 53, pp. 5455-5516, 2020. |
| [3] | AliceOthmani, A. R. Talebb, H. Abdelkawya and A. Hadidc, "Age estimation from faces using deep learning: A comparative analysis," *Computer Vision and Image Understanding,* vol. 196, no. 1077-3142, p. 102961, 2020. |
| [4] | B. Xia, B. B. Amor and M. Daoudi, "Joint gender, ethnicity and age estimation from 3D faces: An experimental illustration of their correlations," *Image and Vision Computing,* vol. 64, pp. 90-102, 2017. |
| [5] | M. Duan, K. Li, C. Yang and K. Li, "A hybrid deep learning CNN–ELM for age and gender classification," *Neurocomputing,* vol. 275, no. 0925-2312, pp. 448-461, 2018. |
| [6] | G. Antipov, M. Baccouche, S.-A. Berrani and J.-L. Dugelay, "Effective training of convolutional neural networks for face-based gender and age predictio," *Pattern Recognition,* vol. 72, pp. 15-26, 2017. |
| [7] | A. Sun, Y. Li, Y.-M. Huang, Q. Li and G. Lu, "Facial expression recognition using optimized active regions," *Human-centric Computing and Information Sciences,* vol. 8, no. 33, 2018. |

1. . [↑](#footnote-ref-1)