Deep Learning-Based Prediction of Age and Gender from Facial Images

Venkata Srinivasu Veesam1, Suban Ravichandran1, Rama Mohan Babu Gattram2

1 Department of Information Technology, Annamalai University, Chidambaram 608002, Tamil Nadu, India
2 Department of Computer Science & Engineering (AI&ML), R.V.R. & J.C. College of Engineering, Guntur 522019, Andhra Pradesh, India

Corresponding Author Email: vasuveesam@gmail.com

https://doi.org/10.18280/isi.280421

Received: 24 June 2023
Revised: 19 August 2023
Accepted: 23 August 2023
Available online: 31 August 2023

Keywords: automatic gender classification, computer vision, artificial intelligence, Deep Convolutional Neural Network (DCNN), face recognition, deep learning

1. INTRODUCTION

The automatic prediction of age and gender from facial images has been a topic of considerable interest among computer vision researchers, due to its significance in human interactions [1-3]. This burgeoning interest is primarily driven by increasing commercial demands for gender and age classification, utilizing digital images and videos, thereby spurring active research in the field. Notable application domains for gender recognition include human-computer interaction, surveillance technologies, content indexing and retrieval, biometric identification, and personalized advertising [4].

It is essential for systems or robots engaged in the domain of human-computer interaction to accurately recognize and confirm human genders, thereby leveraging specialized data to optimize system performance. A system capable of categorizing the gender and age of a user could potentially provide a personalized service tailored to individual users. Moreover, gender-focused monitoring may aid in ascertaining threat levels for a specific gender, given that gender information can be automatically gathered beforehand [5]. The potential of gender sorting to enhance user experiences in video games and mobile applications is indeed noteworthy. Gender information can be employed to present preferred gaming characters or content to individuals, given the divergent interests observed between men and women in video games. The implementation of a gender classification system facilitates efficient demographic data collection for demographic studies. It is, however, challenging to implement gender and age classification of digital face images due to variations in illumination, pose, expression, and occlusion.

Although distinguishing between male and female facial features is a routine task for humans, it poses a significant challenge for computers. For successful identification, machines require key attribute-based data. By employing certain distinguishing traits, machines can discern a person's gender based on facial images.

In this study, the goal is to investigate and analyze gender-specific characteristics of facial features. It is widely acknowledged that distinct variations exist between the facial attributes of females and males [6]. For instance, the eyes of females are often smaller and oval in shape, with eyebrows that have a wider and more arched area between the eyes and the eyebrows. Conversely, males typically have larger and rectangular-shaped eyes, coupled with a relatively short and straight distance between the eyes and the eyebrows. These gender-specific differences in facial attributes warrant further exploration into the underlying genetic and hormonal factors contributing to these variations.
The current study aims to bridge the gap between age and gender estimation methodologies and automatic facial recognition technologies, following the efficient paradigm established by existing facial recognition systems. Recent descriptions of facial recognition technologies have demonstrated that deep convolutional neural networks can significantly evolve [7].

This study underscores the importance of utilizing accurately annotated datasets derived from social image databases containing sensitive individual information, including age, to ensure the reliability and relevance of the analysis [8].

The paper is structured as follows: Section 2 provides an overview of the background and previous work in this field, discussing how these insights have informed the current study. In Section 3, a detailed description of the technical intricacies underpinning the proposed approach is provided. Section 4 outlines the experiments conducted and the corresponding results. Finally, Section 5 concludes the paper, presenting the findings and highlighting potential areas for future research.

2. LITERATURE SURVEY

This section encompasses a comprehensive review of the literature focusing on the evolution of face and age identification techniques, the application of deep learning methods, and the aims of the current study. The subsequent sections delve into the literature related to facial detection algorithms employed by various researchers.

2.1 Facial detection and identification

Facial detection and identification serve as critical components in every face recognition system, necessitating rapid and precise function. Object recognition techniques have facilitated the development of face recognition strategies. One such strategy, the R-CNN, employs a micro-Computer Neural Network known as the Region Proposal Network (RPN) [9]. The detection and labeling of facial landmarks are integral to various facial tasks including face verification, face characteristics inference [10, 11], and face recognition [12, 13].

2.2 Age identification

Recent years have witnessed a surge in interest towards automatically extracting age-related features from facial images, leading to the proposal of numerous strategies. Similar types of analyses were conducted by Fu et al. [14] and Han et al. [15] for age estimation from facial images. Initial methods for age estimation relied on calculating ratios between various extents of facial traits [16]. Following the localization of facial features such as the nose, eyes, mouth, chin, and ears, faces were categorized into different age groups based on customized rules.

A study by Ramanathan and Chellappa [17] utilized a similar methodology to model age progression in individuals under 18. However, these techniques are not suitable for images taken in uncontrolled settings, like those typically found on the internet, due to the requirement for accurate localization of facial features. Another line of research describes the aging process as a subspace or a manifold [18, 19], but these techniques require closely matched input images and are limited to research observations from a few closed image sets (e.g., UIUC-IFP-Y [20, 21], FG-NET [22], and MORPH [23]).

Diverse image descriptors were introduced by Gao et al. for age classification, using a Gabor feature and Fuzzy-LDA technique [24, 25]. Age estimation also employed Biologically-Inspired Features (BIF) and several manifold-learning approaches [26, 27]. Choi et al. used a Support Vector Machines-based hierarchical age classifier (SVM) to categorize an age-class on the input image, followed by a support vector regression for specific age estimation [28-30]. Attributes like Gabor and Local Binary Patterns (LBP) were also utilized. The current methodology outperforms these previously proposed approaches, which were only effective on small or constrained levels of age approximation.

2.3 Gender identification

In 2002, Sun et al. [31] established a gender classification method using biological search for eigen-feature selection. Each image was characterized as a feature vector using Principal Component Analysis (PCA), followed by feature selection using a genetic algorithm, and the system was trained with neural networks. Tivive and Bouzerdoum proposed a gender recognition system in 2006 using an inhibitory convolutional neural network for shunting [32]. Yang et al. [33] conducted an automatic face gender classification study in 2006 utilizing LVM and LDA for sorting. In 2006, Phung and Bouzerdoum developed a classification framework using a novel network, the pyramidal neural network, or PyraNet [34].

A rapid gender classification method was proposed by Ozbudak et al. [35] in 2010 using two-dimensional Discrete Wavelet Transform to break down facial images, and PCA for gender identification. In 2012, Kumari and Majhi adopted the information maximization method to extract features from facial images for a new gender detection system [36]. These features were post-processed using back-propagation neural networks (BPNN) and radial basis function neural network (RBFNN). Nayak and Lakshmi in 2013 proposed a new approach using neural networks that involved preprocessing the image before classification [37]. A multi-layered neural network model was used for categorization.

2.4 Deep learning methods

Deep learning is an artificial intelligence (AI) function that aims to emulate human brain learning through the acquisition of representations. To train software to detect an object, it is necessary to provide it with a large volume of categorized object images. The first deep learning technique employed in machine learning was the deep neural network, which had drawbacks such as overfitting and prolonged training periods. The inclusion of Boltzmann machines (RBMs) and a Deep Belief Network alongside deep neural networks enhanced the performance of the DNN.

2.5 Overview of gender classification methods

A myriad of techniques has been proposed within the sphere of gender classification studies, encompassing methods that rely on facial features, attire, stride, iris, hand morphology, and hair. These methodologies can be broadly bifurcated into two categories: the appearance-based approach and the non-appearance-based approach.
2.5.1 Appearance-Based Approach

The appearance-based approach, a method that leverages visual cues, glean features from three distinct categories: clothing features, dynamic body features, and static body features. The appeal of this approach lies in its ability to harness visual richness and its non-intrusive nature, which renders it applicable across a diverse range of domains. Static body characteristics encompass the form of the body [38], the facial structure [39-41], brows [42], hands [43], and fingernails. Dynamic body features pertain to the individual’s physical activities and movements, such as gait and other motions. However, it is imperative to acknowledge the limitations of this approach, notably its sensitivity to variable conditions and vulnerability to adversarial attacks.

2.5.2 Non-Appearance-Based Approach

The non-appearance-based method, on the other hand, exploits behavioral characteristics and biological measurements. This approach provides the advantages of a deeper contextual understanding and resilience against changes in appearance. The non-appearance approach extracts features derived from biological and social network data. Biological data sources include biometrics such as voice [44], fingerprint, iris [45], empathetic speech [46], and biological signals, including EEG [47, 48], DNA sequences, and ECGs. Despite the benefits, it is important to note the drawbacks, such as limited suitability for certain visual recognition tasks and potential computational complexities.

2.5.3 Deep Convolution Neural Networks

At the heart of Deep Convolution Neural Networks (DCNNs) lies the convolutional layer [49, 50]. Convolution, a mathematical operation, is performed to intertwine data from two sources. DCNNs are capable of transcending the limitations of both appearance and non-appearance methods, making them a versatile choice for tasks requiring the capture of rich visual features while maintaining a contextual understanding. An extensive literature survey led to the identification of DCNN methodologies as promising techniques for age and gender prediction. Harr-like features were found to struggle with images captured at a significant distance, an issue that is mitigated with the employment of the DCNN technique. The study utilized a cascaded approach, with the region of interest (ROI) playing a pivotal role. The classifier was fed an ROI image as input for the analysis. The intent was to train the model using faces of 500 females and 500 males, aiming to identify the face within the document.

3. PROPOSED WORK

In our example, the input data is Convoluted using a convolution filter to produce a feature map. The created DCNN architecture is used to address the proposed gender and age recognition challenge. There are six layers in this network’s architecture made up of two fully linked layers and five convolutional layers. Deep neural network technology is used by DCNN to carry out feature extraction and classification tasks. The suggested work extracts age and gender traits from the face using DCNN, and the results are encouraging.

3.1 Age approximation using caffe

Caffe is a CNN system that allows researchers and other individuals to create sophisticated neural networks and train them without having to write a lot of code. Figure 1 shows about the architecture of Caffe Network. Creating a large dataset for the convolution neural network’s algorithm to train on is a laborious and time-consuming task for age estimation. The dataset must be accurately labelled and come from a social image database that contains the subjects' private information, such as age.

In this research we proposed a framework which includes pre-processing, feature selection, feature extraction, gender and age categorization as shown in Figure 2.

3.2 Implementation of caffe

3.2.1 Sensing

Using specific sensors, such as a camera (pictures, videos) [8, a recorder (audio), physiological measures (EEG, ECG), and information from social networks, the effective raw data needed to classify gender must first be obtained (e.g., Facebook posts, Tweets, blogs)]. The gender classification process uses a variety of methodologies depending on the gathered features.

Figure 1. Caffe network architecture

![Figure 1. Caffe network architecture](image)

Figure 2. Proposed frame work

![Figure 2. Proposed frame work](image)
3.2.2 Pre-processing

To increase the quality of raw data, pre-processing is a crucial step that comprises the normalisation of major identification of signals and extraction of the informative region, filling in the gaps, amputation of noise, and face recognition. The correct signal pre-processing technique removes unwanted information from the raw data, improving the proof of identity precision rate while having no effect on the quality of feature extraction.

3.2.3 Feature extraction

Feature extraction finds the essential elements of the pre-processed data and uses them as an input factor for the classification algorithm. The feature extraction algorithm lowers the amount of data by extracting the characteristics that are important for classification from the pre-processed data.

The ideal characteristics ought to be easy to compute, trustworthy, recognisable, and condition-insensitive. In the subsequent stage, these collected features will be processed by the classifier, who will then do the classification.

3.3 Classification algorithm

The categorization method serves as the foundation for gender identification. Appearance-based techniques and non-appearance-based approaches are the two basic groups into which classification strategies can be separated. When determining gender using the appearance-based method, only a static image or animated video is taken into account. The gender of a person is determined by their physical traits, biometric data, or data from social networks using the non-appearance-based method. Several classifiers, such as support vector machines (SVM) and K-nearest neighbours (KNN), are used with these two techniques, and Gaussian mixture models (GMM), have been used for gender classification as the algorithm for classification that is chosen.

3.3.1 Evaluation

In this step, various metrics are utilised to gauge how well the gender-based classification system is working. The system is essentially assessed for its accuracy, dependability, intrusiveness, etc. One of the most significant aspects is accuracy, or the likelihood of correctly classifying a person as male or female. Gender classification algorithms can be examined and verified using publicly accessible databases.

The basic model for the proposed model was modified VGG-Face from the original study's regular VGG model. The ability to spot some face patterns is a perk of VGG-Face. Within OpenCV, we can use the pre-trained weights and designed model structure for the Caffe framework. The pre-built Caffe model can then be loaded using OpenCV. The simplest method for detecting faces is the OpenCV Harr cascade. As was previously noted, age and gender prediction models are built on VGG and require inputs that are 224×224 shaped. Even if we resize, OpenCV loads (224,224,3) shaped images even though the Caffe model expects (1,3,224,224) shaped inputs. The OpenCV Blob from Image function in the deep neural network’s module turns read images into the desired shape. Predictions can be made by feeding pre-processed input photos into the conceptual model. To make predictions, OpenCV needs the commands set input and call forward, respectively. The gender model initially returns 101-dimensional vectors. The gender model also yields a 2-dimensional model. It is a lady if the first-dimension value is higher than the second. Conversely, if the value of the second dimension is higher than the first, then it is man.

The novelty of the work is implementation of VGG-Net architecture of Deep Convolution Neural Network (D-CNN) for age & gender prediction of an image and this work is used for future work.

4. RESULTS

In this paper, we establish certain benchmarks based on cutting-edge VGG-Net network topologies and demonstrate how gender recognition using face images can increase overall accuracy using VGG-Net architecture of Deep Convolution Neural Network (D-CNN). In the proposed study, we have used the Python language, the OpenCV, Pytorch, and TensorFlow libraries to develop code with Graphical Processing Units (GPUs), which has produced excellent results on such facial image datasets. The suggested method performs admirably, correctly classifying age and gender with less calculation time and greater accuracy. The suggested system chooses the input image from the dataset or possibly in real-time using the camera. Figure 3 displays the outcome of our testing with photos downloaded from the internet using our technique, which achieves the intended outcomes with extreme accuracy.

![Figure 3. Outcome of our testing with photos downloaded from the internet](image)

In summary, addressing the limitations of a current work involves acquiring more data, improving computational resources, validating assumptions under diverse conditions, handling imbalanced data, and broadening the domain coverage. By strategically addressing these limitations, the work's robustness, applicability, and reliability can be significantly improved.

5. CONCLUSION AND FUTURE WORK

Two of the most crucial tools for gathering information from a person are their age and gender. The information about human faces is sufficient for a range of uses. Human gender and age classification are critical for reaching the right audience. Gender classification for computer vision application is analysed and improvement in gender classification accuracy is presented in this research work. From this study, we can conclude with two important
conclusions. The gender classification in computer vision is a highly challenging task due to variations in illumination, expression, pose, age, scales and occlusion. First, despite the limited availability of age and gender-tagged photos, the results of gender and age detection can be improved using CNN. Second, by employing additional training data and more complex systems, the system's performance can be slightly increased. In summary, DCNNs like VGG are architectures used for various computer vision tasks, while Caffe is a framework used to build and train neural networks. The efficiency of a model depends on the specific architecture, framework used to build and train neural networks. The increased. In summary, DCNNs like VGG are architectures for complex systems, the system's performance can be slightly increased. In summary, DCNNs like VGG are architectures used for various computer vision tasks, while Caffe is a framework used to build and train neural networks. The efficiency of a model depends on the specific architecture, framework used to build and train neural networks. The increased.

REFERENCES


