Abstract

Since the advent of social media, there has been an increased interest in automatic age and gender classification through facial images. So, the process of age and gender classification is a crucial stage for many applications such as face verification, aging analysis, ad targeting and targeting of interest groups. Yet most age and gender classification systems still have some problems in real-world applications. This work involves an approach to age and gender classification using multiple convolutional neural networks (CNN). The proposed method has 5 phases as follows: face detection, remove background, face alignment, multiple CNN and voting systems. The multiple CNN model consists of three different CNN in structure and depth; the goal of this difference It is to extract various features for each network. Each network is trained separately on the AGFW dataset, and then we use the Voting system to combine predictions to get the result.

CHAPTER 1

INTRODUCTION

Biometrics, is the science of analyzing the physical or behavioral characteristics of each individual that enable the authentication of their identity in a reliable manner, it offers significant advantages conventional identification methods, such as passwords and cards, are not transferable, exclusive to each person and are not lost or stolen, particularly because of biometric features. The range of biometric solutions
relies on user approval, security, cost and time for implementation...etc. Recently, face recognition has been one of the most interesting tasks in pattern recognition, many applications use this technique because the human face is considered a very rich source of information. In particular, gender and age are facial features that can be very useful for a multitude of applications, for example an automatic gender and age prediction system is used to profile customers who are interested for a product or for target advertising. The areas of age and gender classification have been studied for decades. Until detailing the methods used in this article, we will first provide a summary of the facial recognition experiments carried out by scholars, which can be grouped into tree classes of interest. Over the last decade, the rate of image uploads to the Internet has grown at a nearly exponential rate. This newfound wealth of data has empowered computer scientists to tackle problems in computer vision that were previously either irrelevant or intractable. Consequently, we have witnessed the dawn of highly accurate and effificient facial detection frameworks that leverage convolutional neural networks under the hood. One of the most critical barriers that face any system to ageestimation or age-classification is the absence of a consistent pattern of facial aging. This is due to the nature of human faces, and the stages of aging may differ from one human to another.

CHAPTER 2
LITERATURE SURVEY

2.1 Age and Gender Classification using Multiple Convolutional Neural Network

Abstract
Since the advent of social media, there has been an increased interest in automatic age and gender classification through facial images. So, the process of age and gender classification is a crucial stage for many applications such as face verification, aging analysis, ad targeting and targeting of interest groups. Yet most age and gender classification systems still have some problems in real-world applications. This work involves an approach to age and gender classification using multiple convolutional neural networks (CNN). The proposed method has 5 phases as follows: face detection, remove background, face alignment, multiple CNN and voting systems. The multiple CNN model consists of three different CNN in structure and depth; the goal of this difference It is to extract various features for each network. Each network is trained separately on the AGFW dataset, and then we use the Voting system to combine predictions to get the result.

Introduction
Age and gender play fundamental roles in social interactions. Languages reserve different salutations and grammar rules for men or women, and very often different vocabularies are used when addressing elders compared to young people. Despite the basic roles these attributes play in our day-to-day lives, the ability to automatically estimate them accurately and reliably from face images is still far from meeting the needs of commercial applications. This is particularly perplexing when considering recent claims to super-human capabilities in the related task of face recogni
tion (e.g., [48]).
Past approaches to estimating or classifying these attributes from face images have relied on differences in facial feature dimensions [29] or "tailored" face descriptors
(e.g., [10, 15, 32]). Most have employed classifification schemes designed particularly for age or gender estimation tasks, including [4] and others. Few of these past methods were designed to handle the many challenges of unconstrained imaging conditions [10]. Moreover, the machine learning methods employed by these systems did not fully Figure 1. Faces from the Adience benchmark for age and gender classifification [10].
These images represent some of the challenges of age and gender estimation from real-world, unconstrained images. Most notably, extreme blur (low-resolution), occlusions, out-of-plane pose variations, expressions and more exploit the massive

numbers of image examples and data available through the Internet in order to improve classifification capabilities.
In this paper we attempt to close the gap between automatic face recognition capabilities and those of age and gender estimation methods. To this end, we follow the successful example laid down by recent face recognition systems: Face recognition techniques described in the last few years have shown that tremendous progress can be made by the use of deep convolutional neural networks (CNN) [31].

We demonstrate similar gains with a simple network architecture, designed by considering the rather limited availability of accurate age and gender labels in existing face data sets.
We test our network on the newly released Adience benchmark for age and gender classifification of unfifiltered face images [10]. We show that despite the very challenging nature of the images in the Adience set and the simplicity of our network design, our method outperforms existing state of the art by substantial margins. Although these results provide a remarkable baseline for deep-learning-based approaches, they leave room for improvements by more elaborate system designs, suggesting that the problem of accurately estimating age and gender in the unconstrained settings, as reflflected by the Adience images, remains unsolved.
In order to provide a foothold for the development of more effective future methods, we make our trained models and classifification system publicly available

A CNN for age and gender estimation
Gathering a large, *labeled* image training set for age and gender estimation from social image repositories requires either access to personal information on the subjects appearing in the images (their birth date and gender), which is often private, or is tedious and time-consuming to manually label. Data-sets for age and gender estimation from real-world social images are therefore relatively limited in size and presently no match in size with the much larger image classifification data-sets (e.g. the Imagenet dataset [45]). Overfifitting is common problem when machine learning based methods are used on such small image collections. This problem is exacerbated when considering deep convolutional neural networks due to their huge numbers of model parameters. Care must therefore be taken in order to avoid overfifitting under such circumstances.

3.1. **Network architecture**
Our proposed network architecture is used throughout our experiments for both age and gender classifification. It is illustrated in Figure 2. A more detailed, schematic diagram of the entire network design is additionally provided in Figure 3. The network comprises of only three convolutional layers and two fully-connected layers with a small number of neurons. This, by comparison to the much larger architectures applied, for example, in [28] and [5]. Our choice of a smaller network design is motivated both from our desire to reduce the risk of overfitting as well as the nature
Figure 3. Full schematic diagram of our network architecture. Please see text for more details. of the problems we are attempting to solve: age classification on the Audience set

requires distinguishing between eight classes; gender only two. This, compared to, e.g., the ten thousand identity classes used to train the network used for face recognition in [48].

All three color channels are processed directly by the network. Images are fifirst rescaled to 256 × 256 and a crop of 227 × 227 is fed to the network. The three subsequent convolutional layers are then defined as follows.

1. 96 filters of size 3×7×7 pixels are applied to the input in the first convolutional layer, followed by a rectified linear operator (ReLU), a max pooling layer taking the maximal value of 3 × 3 regions with two-pixel strides and a local response normalization layer [28].

2. The 96 × 28 × 28 output of the previous layer is then processed by the second convolutional layer, containing 256 filters of size 96 × 5 × 5 pixels. Again, this is followed by ReLU, a max pooling layer and a local response normalization layer with the same hyper parameters as before.

3. Finally, the third and last convolutional layer operates on the 256 × 14 × 14 blob by applying a set of 384 filters of size 256 × 3 × 3 pixels, followed by ReLU and a max pooling layer. The following fully connected layers are then defined by:

4. A first fully connected layer that receives the output of the third convolutional layer and contains 512 neurons, followed by a ReLU and a dropout layer.

5. A second fully connected layer that receives the 512- dimensional output of the first fully connected layer and again contains 512 neurons, followed by a ReLU and a dropout layer.

6. A third, fully connected layer which maps to the final classes for age or gender.

Finally, the output of the last fully connected layer is fed to a soft-max layer that assigns a probability for each class. The prediction itself is made by taking the class with the maximal probability for the given test image.

3.2. **Testing and training**

Initialization. The weights in all layers are initialized with random values from a zero mean Gaussian with standard deviation of 0.01. To stress this, we do not use pretrained models for initializing the network; the network is trained, from scratch, without using any data outside of the images and the labels available by the benchmark. This, again, should be compared with CNN implementations used for face recognition, where hundreds of thousands of images are used for training [48].

Target values for training are represented as sparse, binary vectors corresponding to the ground truth class of the number of classes (two for gender, eight for the eight age classes of the age classification task), containing 1 in the index of the ground truth and 0 elsewhere Network training. Aside from our use of a lean network architecture, we apply two additional methods to further limit the risk of overfifitting. First we apply dropout learning [24] (i.e. randomly setting the output value of net work neurons to zero). The network includes two dropout layers with a dropout ratio of 0.5 (50% chance of setting a neuron's output value to zero). Second, we use data augmentation by taking a random crop of 227 × 227 pixels from the 256 × 256 input image and randomly mirror it in each forward-backward training pass. This, similarly to the multiple crop and mirror variations used by [48]. Training itself is performed using stochastic gradient decent with image batch size of fififty images. The initial learning rate is e−3, reduced to e−4 after 10K iterations. Prediction. We experimented with two methods of using the network in order to produce age and gender predictions for novel faces:

• Center Crop: Feeding the network with the face image, cropped to 227 × 227 around the face center.
• Over-sampling: We extract fifive 227 × 227 pixel crop regions, four from the corners of the 256 × 256 face image, and an additional crop region from the center of the face. The network is presented with all five images, along with their horizontal reflections. Its final prediction is taken to be the average prediction value across all these variations. We have found that small misalignments in the Audience images, caused by the many challenges of these images (occlusions, motion blur, etc.) can have a noticeable impact on the quality of our results. This second, over-sampling method, is designed to compensate for these small misalignments, bypassing the need for improving alignment quality, but rather directly feeding the network with multiple translated versions of the same face.

4. Experiments

Our method is implemented using the Caffe open-source framework [26]. Training was performed on an Amazon GPU machine with 1,536 CUDA cores and 4GB of video memory. Training each network required about four hours, predicting age or gender on a single image using our network requires about 200ms. Prediction running times can conceivably be substantially improved by running the network on image batches.

4.1. **The Audience benchmark**

We test the accuracy of our CNN design using the recently released Adience benchmark [10], designed for age and gender classifification. The Adience set consists of images automatically uploaded to Flickr from smart-phone devices. Because these images were uploaded without prior manual fifiltering, as is typically the case on media webpages (e.g., images from the LFW collection [25]) or social websites (the Group Photos set [14]), viewing conditions in these images are highly unconstrained, reflflecting many of the real-world challenges of faces appearing in Internet images. Adience images therefore capture extreme variations in head pose,

lightning conditions quality, and more. The entire Adience collection includes roughly 26K images of 2,284 subjects. Table 1 lists the breakdown of the collection into the different age categories. Testing for both age or gender classification is performed using a standard five-fold, subject-exclusive cross-validation protocol, defined in [10]. We use the in-plane aligned version of the faces, originally used in [10]. These images are used rather than newer alignment techniques in order to highlight the performance gain attributed to the network architecture, rather than better preprocessing. We emphasize that the same network architecture is used for all test folds of the benchmark and in fact, for both gender and age classification tasks. This is performed in order to ensure the validity of our results across folds, but also to
demonstrate the generality of the network design proposed here; the same architecture performs well across different, related problems. We compare previously reported results to the results computed by our network. Our results include both methods for testing: center-crop and over-sampling (Section 3).

4.2. Results

Table 2 and Table 3 presents our results for gender and age classification respectively. Table 4 further provides a confusion matrix for our multi-class age classification results. For age classification, we measure and compare both the accuracy when the algorithm gives the exact age-group classification and when the algorithm is off by one adjacent age-group (i.e., the subject belongs to the group im
mediately older or immediately younger than the predicted group). This follows others who have done so in the past, and reflflects the uncertainty inherent to the task – facial features often change very little between oldest faces in one age class and the youngest faces of the subsequent class. Both tables compare performance with the methods described in [10]. Table 2 also provides a comparison with [23] which used the same gender classification pipeline of [10] applied to more effective alignment of the faces; faces in their tests were synthetically modifified to appear facing forward.

Evidently, the proposed method outperforms the reported state-of-the-art on both tasks with considerable gaps. Also evident is the contribution of the over-sampling approach, which provides an additional performance boost over the original network. This implies that better alignment (e.g., frontalization [22, 23]) may provide an additional boost in performance. We provide a few examples of both gender and age misclassifications in Figures 4 and 5, respectively. These show that many of the mistakes made by our system are due to extremely challenging viewing conditions of some of the Adience benchmark images. Most notable are mistakes caused by blur or low resolution and occlusions (particularly from heavy makeup). Gender estimation mistakes also frequently occur for images of babies or very young children where
obvious gender attributes are not yet visible.

Conclusions

Though many previous methods have addressed the problems of age and gender classification, until recently, much of this work has focused on constrained images taken in lab settings. Such settings do not adequately reflflect appearance variations common to

the real-world images in social websites and online repositories. Internet images, however, are not simply more challenging: they are also abundant. The easy availability of huge image collections provides modern machine learning based systems with effectively endless training data, though this data is not always suitably labeled for supervised learning. Taking example from the related problem of face recognition we explore how well deep CNN perform on these tasks using Internet data. We provide results with a lean deep-learning architecture designed to avoid overfifitting due to the limitation of limited labeled data. Our network is “shallow” compared to some of the recent network architectures, thereby reducing the number of its parameters and the chance for overfifitting. We further inflflate the size of the
training data by artifificially adding cropped versions of the images in our training set. The resulting system was tested on the Adience benchmark of unfifiltered images and shown to signifificantly outperform recent state of the art. Two important conclusions can be made from our results. First, CNN can be used to provide improved age and gender classifification results, even considering the much smaller size of contemporary unconstrained image sets labeled for age and gender. Second, the simplicity of our model implies that more elaborate systems using more training data
may well be capable of sub

Acknowledgments

This research is based upon work supported in part by the Offifice of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA 2014-14071600010. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the offifficial policies or endorsements, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright annotation thereon.
-

2.2 Human Age And Gender Classification using Convolutional Neural Network

Abstract

Pattern recognition and automatic classification are very active research areas, their main objectives are to develop intelligent systems able to achieve efficiently learning and recognizing objects. An essential section of these applications is attached to biometrics, which is used for security purposes in general. The facial modality as a fundamental biometric technology has become increasingly important in the field of
research. The goal of this work is to develop a gender prediction and age estimation system based on convolutional neural networks for a face image or a real-time video. In this paper, three CNN network models were created with different architecture (the number of filters, the number of convolution layers...) validated on IMDB and WIKI dataset, the results obtained showed that CNN networks greatly improve the performance of the system as well as the accuracy of the recognition.

INTRODUCTION

Biometrics, is the science of analyzing the physical or behavioral characteristics of each individual that enable the authentication of their identity in a reliable manner, it offers significant advantages over conventional identification methods, such as passwords and cards, are not transferable, exclusive to each person and are not lost or stolen, particularly because of biometric features. The range of biometric solutions
relies on user approval, security, cost and time for implementation...etc. Recently, face recognition has been one of the most interesting tasks in pattern recognition, many applications use this technique because the human face is considered a very rich source of information. In particular, gender and age are facial features that can be very useful for a multitude of applications, for example an automatic gender and age prediction system is used to profile customers who are interested for a product or for
target advertising. The areas of age and gender classification have been studied for decades. Until detailing the methods used in this article, we will first provide a summary of the facial recognition experiments carried out by scholars, which can be
grouped into tree classes of interest. [1]. Local strategies extract facial characteristics by focusing on the key points of the face, such as nose, mouth, eyes, which will offer more information. Global approaches their concept is to use the full surface of the face as a source of information independent of local characteristics such as eyes, lips ... etc. Hybrid techniques they merge all sorts of strategies, theoretically giving the best of all [1] [2]. Many techniques were applied for gender prediction from
face images, Hui-Cheng Lian et al [2] proposed gender recognition taking into account both form and texture information from facial images. This last is divided into small regions, from which local binary histograms are extracted and concated into a single vector representing the facial image, then the support vector machine (SVM) is applied for gender prediction. Jing Wu et al [3] proposed a gender classification using the form shading (SFS) and multi layers perceptron (MLP)
was applied for this study by Golomb et al [4]. Khan et al [5] applied classifier reinforcement in particular adaboost for gender prediction during this period. Among age prediction studies, Yamaguchi et al [6] confirmed that differences between the characteristics of an adult's face and a child's include the length of the face and the ratio of each side. Burt and Perrett [7] studied the age estimate based on the use of average faces of people between 25 and 60 years of age, they used approach that
generalize facial texture and shape, also Ueki and Coll [8] reported a method of classifying age groups by linear discriminating analysis. Although the SVM has been tested for age classification several times [9]. Kwon and Lobo [10] defined a method for classifying input images into one of three age groups: child, young and old using texture information. However, almost all previous research has been based on craniofacial development method and analysis of skin wrinkles. During the last few years, a convolution neural network centered on deep learning [11], according to the powerful ability to estimate and extract features to enhance the precision of image classification, state-of-the-art achievements have been achieved in large areas. In this paper, we will first introduce the basic structure of the convolutional neural network. Next, we will

describe models CNN for training data to classify gender and age, then we will present the results obtained using a model trained by these data,
and finally, our conclusion.

EXPERIMENTAL WORK AND RESULTS

A. Experiment and result for gender prediction

The CNNs models proposed in table 1 was build using Keras which has many advantages to improve efficiency of the model. We input 2500 images of male and female separately 2000 images for train and 500 images for the test. The CNNs models were trained for 1500 epochs, after every epoch the accuracy was calculated, which is the count of predictions where the predicted value is equal to the true value, it is typically expressed as a percentage. The input is passed through a pile of convolutional and maxpooling layer, the non-linear activation function (ReLu) was
used, in output result we applied a sigmoid function as shown in table 1, for all models, RMSpro was used as an optimizer.

D. Discussion

In this work, we aimed to automate a system for gender prediction and age estimation by using CNN and deep learning techniques, first, we build three models: CNN1, CNN2, CNN3 as described in table 1, these models were trained on IMDB dataset, we noticed that the CNN 3 present best results compared to the CNN 2 and CNN 1, due to the depth of the network. In CNN 3 we used three convolutional layer but in CNN 2 and CNN 1 we used only two layers of convolution with various filter size,16 filters were used at the 1st convolution layer in CNN 1, 32 filters were applied in the 1st convolutional layer in CNN 2, when the number of the filter was large the performance of system increase. In other word, the depth of network and the number of the filters have a great influence in creating an efficient convolutional network ranking. For age estimation, we used the model CNN 3 to classify age in three categories; young (20-39 years), middle (40-59 years), old (more than 60 years), this kind of classification will eventually be useful for marketing to identify customers. After training model CNN 3, we noticed that this CNN model can obtain a perfectly acceptable result, as show in figures 6 and 7. Furthermore, the rate of classification growth with the number of epochs, this reflects that with each epoch the model
learns more information.

Fig.8. Example of age estimation form IMDB dataset.
Fig.9. Example of gender prediction form WIKI dataset.
Fig.10. Example of age and gender classification from IMDB dataset.

VII. CONCLUSION

In this article, we analyzed the implementation of deep convolutional neural network for human age and gender prediction using CNN. During this study various design was

developed for this task, age and gender classification is one of the key segments of research in the biometric as social applications with the goal that the future forecast and the information disclosure about the particular individual should be possible adequately. In this study, the main conclusion that can be drawn is that age and gender from face recognition are very popular among panels to implement an intelligent system that can achieve good and robust results in the accuracy of recognition, we employed a deep learning algorithm, as a convolutional neural network to propose a simple study contain various CNNs model in gender classification, trained in well-known datasets IMDB-WIKI, then we applied an efficient model for age estimation, the different results obtained in terms of precision, compared with those cited in the state of the art, have shown that the depth of the convolutional networks used in this work is an important factor in achieving better precision. The interpretation of the figure (4- 5-6-7) and the outcome of tables 3 and 4 was based on parameter settings in our experiment described in table 2. The proposed network provides significant precision improvements in age and gender classification, but takes considerable time to train the network to implement the correct prediction. Finally, as a perspective, an extension of this work can be envisaged by creating a face detection and recognition system based on CNNs as a feature extractor and the machine vector support as a classifier, another perspective would be the tests our approach on other facial databases showing strong variations in lighting and pose.

2.3 Convolutional Neural Network for age and gender classification

Abstract

This paper focuses on the problem of gender and age classifification for an image. I build off of previous work [12] that has developed effificient, accurate architectures for these tasks and aim to extend their approaches in order to improve results. The fifirst main area of experimentation in this project is modifying some previously published, effective architectures used for gender and age classifification [12]. My attempts include reducing the number of parameters (in the style of [19]), increasing the depth of the network, and modifying the level of dropout used. These modififications actually ended up causing system performance to decrease (or at best, stay the same) as compared with the simpler architecture I began with. This verifified suspicions I had that the tasks of age and gender classifification are more prone to over-fitting than other types of classification. The next facet of my project focuses on coupling the architectures for age and gender recognition to take adwe must make do with the nature of this problem we are apvantage of the gender-specifific age characteristics and age specifific gender characteristics inherent to images. This stemmed from the observation that gender classifification is an inherently easier task than age classifification, due to both the fewer number of potential classes and the more prominent intra-gender facial variations. By training different age classififiers for each gender I found that I could improve the performance of age classification, although gender classification did not see any significant gains.

1. Introduction