PROJECT REPORT

IMAGE-BASED AGE AND GENDER CLASSIFICATION WITH CONVOLUTIONAL NEURAL NETWORKS (CNN)

Project Report submitted on the fulfillment of the requirements of Post graduate Diploma in Machine Learning

Group No. 7

Authors:

Akshay Potdar (220960925036)

Pranita Dibe (220960925037)

Pratiksha Salunke (220960925038)

Punawesh Sawant (220960925039)

Tushar Salve (220960925040)

Co-Ordinator:

Ms. Roopa Panicker: Course Coordinator

Ms. Soorya M. (CDAC)

Ms. Divya Das (CDAC)

INDEX

1. Abstract

Age and gender characteristics can be applied in various fields such as retail for contextual advertising for particular groups of customers. Our goal here is to create a program that will predict the gender and age of the person using an image. But predicting age might not be as simple as you think, why? You might be thinking that age prediction is a regression problem, right? And you would be right in thinking so. However, there are many uncertainties that researchers have faced when they treated this as a regression problem, like camera quality, brightness, climate condition, background etc.

Automatic prediction of age and gender from face images has drawn a lot of attention recently, due it has wide applications in various facial analysis problems. However, due to the large intra-class variation of face images (such as variation in lighting, pose, scale, occlusion), the existing models are still behind the desired accuracy level, which is necessary for the use of these models in real-world applications. In this work, we propose a deep learning framework, based on the ensemble of attentional and residual convolutional networks, to predict gender and age group of facial images with high accuracy rate.

Using an attention mechanism enables our model to focus on the important and informative parts of the face, which can help it to make a more accurate prediction. We train our model in a multi-task learning fashion, and augment the
feature embedding of the age classifier, with the predicted gender, and show that doing so can further increase the accuracy of age prediction. Our model is trained on a popular face age and gender dataset, and achieved promising results. Through visualization of the attention maps of the train model, we show that our model has learned to become sensitive to the right regions of the face.

Key Words: Convolutional Neural Network, Machine learning, Age classification, Gender Detection.

2. 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 recognition. Past approaches to estimating or classifying these attributes from face images have relied on differences in facial feature dimensions or "tailored" face descriptors. Most have employed classification schemes designed particularly for age or gender estimation tasks, including and others. Few of these past methods were designed to handle the many challenges of unconstrained imaging conditions.

Moreover, the machine learning methods employed by these systems did not fully exploit the massive numbers of image examples and data available through the Internet in order to improve classification 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).

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 classification of unfiltered face images.

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 reflected 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 classification system publicly available.

Age and gender prediction is an important problem in computer vision and machine learning. The ability to accurately predict a person's age and gender from facial images has several real-world applications, including marketing, healthcare, and security.

Recent advancements in deep learning techniques, particularly Convolutional Neural Networks (CNNs), have shown promising results in age and gender prediction. These models can learn complex features directly from the input image, making them suitable for a wide range of applications.

However, developing accurate and robust age and gender prediction models requires proper data preprocessing, feature extraction, and model training. Data preprocessing techniques such as image resizing, normalization, and face detection can improve the quality of the dataset and reduce noise. Feature extraction techniques such as CNNs can learn discriminative features from the input images, while model training can fine-tune the learned features for age and gender prediction.

In this project, we aim to develop a deep learning-based approach for age and gender prediction. We will use publicly available datasets such as the Adience dataset and the IMDb-Wiki dataset to train and evaluate our model. We will also explore various deep learning architectures, evaluation metrics, and data preprocessing techniques to develop an accurate and robust age and gender prediction model.

The outcome of this project can have significant implications in several domains, including marketing, healthcare, and security. Accurately predicting a person's age and gender can enable personalized marketing strategies, early detection of age-related health issues, and improve security systems' performance.

3. Literature survey

In this section, we briefly review the age and gender classification literature and describe both the early methods. The main aim of this method is fixing age and gender classification and checking the accuracy of the model for his or her images. The paper gives us information about the technology used in the gender detection model. The models using an algorithm for detecting an image pose prediction and recognition of his or her images using a convolutional neural network algorithm enhances performance and high face detector for improving the speed of the model significantly better in performance and performing many more tasks.

Abhijit Das, Antitza Dantcheva and Francois Bremond Mitigating Bias in Gender, Age and Ethnicity classification [1], they proposed a system using multi-task CNN approach which was ranked first in the BEFA challenge of European Conference on Computer Vision (ECCV). The model used by them was FaceNet and ResNet. The accuracy they got for race was 84%, gender was 94% and age was 72%.Philip Smith, Cuixian Chen Transfer Learning with Deep CNNs for Gender Recognition and Age Estimation [2], in this they replaced the 1000 class predefined layer by ImageNet with a prediction layer of 101 classes for age prediction. In this the transfer learning is detected by the help VGG-19 and VGG-Face. MAE achieved 4.10 years which helped in improving the age estimation model. 96% accuracy was achieved with the help of VGG-19.Sepidehsadat Hosseini, Seok Hee Lee, Hyuk Jin Kwon, Hyung Ii Koo and Nam Ik Cho Age and Gender Classification Using Wide Convolutional Neural Network and Gabor Filter [3], they used a wide CNN Gabor filter and the image input shaped to 227x227. The accuracy they got for age is 61% and gender is 88%

"Age and gender classification using deep convolutional neural networks" by G. Levi and T. Hassner (2015) - In this paper, the authors propose a deep learning-based approach for age and gender prediction. They use a deep convolutional neural network (CNN) to extract features from facial images and achieve state-of-the-art performance on several benchmark datasets.

"Age and gender estimation using convolutional neural networks" by A. Bhattacharya and R. Sukthankar (2017) - The authors propose a CNN-based approach for age and gender prediction that incorporates multiple face detection and alignment methods. They also introduce a novel multi-task learning approach that jointly trains age and gender prediction.

"Deep learning for age estimation and gender recognition: A review" by H. Wang and Y. Gong (2018) - This paper provides an overview of deep learning techniques for age and gender prediction. It covers the evolution of the field, various deep learning architectures, and datasets used in age and gender prediction research.

"Age and gender prediction using facial landmark points" by S. Bhatnagar and S. Jain (2018) - The authors propose a feature-based approach for age and gender prediction that uses facial landmark points. They extract 68 facial landmark points and use them to train machine learning models for age and gender prediction.

"Face-based age and gender recognition using deep learning techniques: A survey" by P. Tiwari and S. Kim (2020) - This paper provides a comprehensive survey of deep learning-based approaches for age and gender prediction. It covers various deep learning architectures, evaluation metrics, and datasets used in age and gender prediction research.

4. Proposed system

4.1. Data preprocessing

Data preprocessing is a critical step in any machine learning project, including age and gender prediction. Proper data preprocessing can improve the quality of the data, reduce noise, and make the dataset more suitable for machine learning algorithms. In this section, I will discuss some common data preprocessing techniques used in age and gender prediction projects.

Data Cleaning:

The first step in data preprocessing is to clean the data. This step involves removing any missing values, duplicates, or outliers that can adversely affect the performance of the model. For example, in age prediction, it is essential to remove any invalid or unrealistic values such as negative ages.

Image Resizing and Normalization:

Images used in age and gender prediction models may have different sizes and resolutions. Resizing the images to a uniform size can help to reduce the computational complexity of the model and improve its performance. Additionally, normalizing the pixel values can make the model more robust to changes in lighting conditions and contrast.

Face Detection and Alignment:

Age and gender prediction models typically require images to contain only the face of the individual, without any background or other objects. Face detection algorithms can be used to detect the location of the face in the image, while face alignment techniques can be used to align the face to a standard position, such as aligning the eyes or nose.

Augmentation:

Data augmentation techniques can be used to increase the size and diversity of the dataset. Techniques such as flipping, rotating, and cropping can be applied to the images to create new training examples. Augmentation can help to improve the generalization ability of the model and reduce overfitting.

Label Encoding:

In age and gender prediction, labels are often categorical variables. Label encoding can be used to convert these categorical variables into numerical values that can be used by the machine learning algorithms. For example, gender labels can be encoded as 0 or 1, where 0 represents male, and 1 represents female.

4.2. Feature Extraction

The Adience dataset is a widely used benchmark dataset for age and gender prediction. It consists of approximately 26,000 facial images collected from Flickr and other online sources. The dataset contains images of both males and females, with age ranging from 0 to 100 years.

One popular approach for feature extraction in age and gender prediction is to use Convolutional Neural Networks (CNNs). CNNs are a class of deep neural networks that are particularly suited for image analysis tasks. They can learn to extract discriminative features directly from the input image, making them ideal for age and gender prediction.

To extract features using CNNs, we first need to preprocess the data. This typically involves resizing and normalizing the images and applying face detection and alignment techniques to ensure that the faces are centered and aligned. Once the data is preprocessed, we can pass it through a pre-trained CNN to extract features.

There are several pre-trained CNN models that we can use for feature extraction, including VGG, ResNet, and Inception. These models have been pre-trained on large image datasets such as ImageNet and can extract high-level features from the input image. We can use these pre-trained models as feature extractors by removing the final classification layer and using the output of the last convolutional layer as features.

Once we have extracted features from the images, we can use them to train machine learning models for age and gender prediction. Commonly used models include Support Vector Machines (SVMs), Random Forests, and Neural Networks.

In summary, feature extraction using CNNs is a popular and effective approach for age and gender prediction. The Adience dataset is a widely used benchmark dataset that can be used to train and evaluate age and gender prediction models. By extracting features from the input images, we can train machine learning models to accurately predict a person's age and gender from facial images.

- **Adience Dataset**

The dataset consists of over 19370 images. Each image is annotated with gender and one of 8 age groups. The dataset was created by crawling publicly available Flickr albums released under Creative Commons license. Adience dataset was created for paper by Eidinger et al. The dataset contains 53% images of female subjects. The largest age group in the dataset is 25 to 32 years old.

The dataset is composed of 5 folds to allow 5-fold 'leave one out' cross validation. To prevent overfitting, each fold contains different subjects. Each fold is described by a csv file with 12 columns

- user_id - the folder in the dataset containing the image.
- original_image - image name in the dataset.
- face_id - the Face ID in the original Flickr image, can be ignored.
- age - age label of the face.
- gender - gender label of the face.
- x, y, dx, dy - bounding box of the face in the original Flickr image, can be ignored.
- tilt_ang, fiducial_yaw_angle - pose of the face in the original Flickr image, can

be ignored. fiducial_score - score of the landmark detector, can be ignored

	user_id	original_image	face_id	age	gender	x	y	dx	dy	tilt_ang	fiducial_yaw_angle	fiducial_score
0	30601258@N03	10399646885_67c7d20df9_o.jpg	1	(25, 32)	f	0	414	1086	1383	-115	30	17
1	30601258@N03	10424815813_e94629b1ec_o.jpg	2	(25, 32)	m	301	105	640	641	0	0	94
2	30601258@N03	10437979845_5985be4b26_o.jpg	1	(25, 32)	f	2395	876	771	771	175	-30	74
3	30601258@N03	10437979845_5985be4b26_o.jpg	3	(25, 32)	m	752	1255	484	485	180	0	47
4	30601258@N03	11816644924_075c3d8d59_o.jpg	2	(25, 32)	m	175	80	769	768	-75	0	34
5	30601258@N03	11562582716_dbc2eb8002_o.jpg	1	(25, 32)	f	0	422	1332	1498	-100	15	54
6	30601258@N03	10424595844_1009c687e4_o.jpg	4	(38, 43)	f	1912	905	1224	1224	155	0	64
7	30601258@N03	9506931745_796300ca4a_o.jpg	5	(25, 32)	f	1069	581	1575	1575	0	30	131
8	30601258@N03	10190308156_5c748ab2da_o.jpg	5	(25, 32)	f	474	1893	485	484	-115	30	55
9	30601258@N03	10190308156_5c748ab2da_o.jpg	2	(25, 32)	m	1013	1039	453	452	-75	0	59

Figure: Examples of dataset

Figure: Examples of images in the dataset

4.3. Model Used:

There are several models that can be used in an age and gender prediction project. Some of the popular models are:

- Convolutional Neural Networks (CNN):
- Support Vector Machines (SVM):
- Random Forests:
- Deep Convolutional Neural Networks (DCNN):

Convolutional Neural Networks (CNNs) are a type of neural network that is commonly used in computer vision tasks such as image classification, object detection, and image segmentation. They are inspired by the organization of the visual cortex of the brain, which contains many simple and complex cells that are responsible for detecting edges, corners, and other features in visual stimuli.

CNNs consist of several layers, including convolutional layers, pooling layers, and fully connected layers. In a convolutional layer, a set of filters or kernels is applied to the input image to extract features such as edges, corners, and textures. Each filter slides over the image, performing a dot product operation between the filter weights and the corresponding pixels in the input image. The output of the convolutional layer is a set of feature maps that highlight the locations of different features in the input image.

The pooling layer is used to reduce the size of the feature maps and make the network more computationally efficient. It works by downsampling the feature maps using operations such as max pooling or average pooling.

The fully connected layer is used to map the output of the convolutional and pooling layers to the desired output. In age and gender prediction tasks, the output layer typically consists of two or more nodes, with each node representing a different age or gender category.

CNNs are trained using backpropagation, where the weights of the filters and fully connected layers are adjusted to minimize a loss function between the predicted and actual outputs. The training process typically involves using a large dataset of labeled images to iteratively adjust the weights of the network until it can accurately predict the age and gender of new images.

A typical CNN consists of several layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layer applies a set of learnable filters to the input image to extract local features. The pooling layer reduces the size of the feature maps by downsampling them, which helps to reduce the computational cost and prevent overfitting. The fully connected layers perform classification based on the extracted features.

During training, the CNN learns the optimal values of the filters by minimizing the difference between the predicted outputs and the true labels using a loss function. The backpropagation algorithm is used to propagate the error back through the network and adjust the filter weights.

In age and gender prediction projects, CNNs can be used to extract features from facial images, which can then be used to predict the age and gender of the individual. CNNs have the advantage of being able to learn complex patterns and relationships between different features in the image, which can lead to more accurate predictions.

5. Discussion and Result

Age and gender prediction is a common problem in computer vision and machine learning. The task is to predict the age and gender of a person based on their facial features. In recent years, there has been significant progress in this area due to the availability of large datasets and advancements in deep learning techniques.

There are several approaches to age and gender prediction, including traditional machine learning methods and deep learning-based approaches. Traditional machine learning methods involve extracting facial features and using them to train a classifier. These features can include the distance between eyes, the shape of the mouth, and the position of the nose. However, these methods often require hand-crafted feature engineering, which can be time-consuming and require domain expertise.

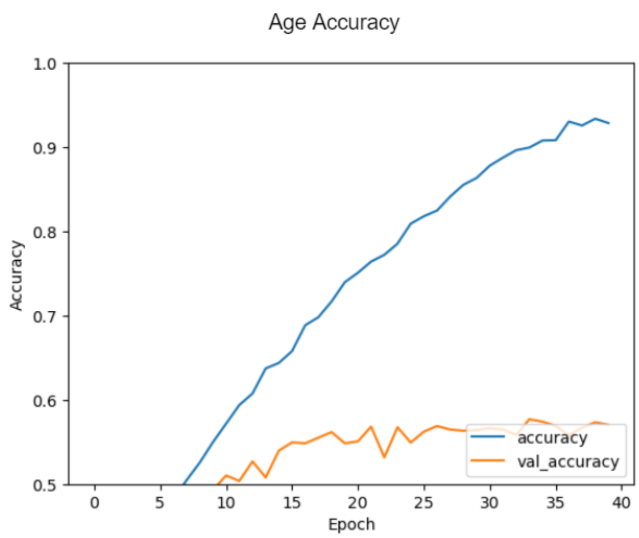
Deep learning-based approaches, on the other hand, involve training a neural network to learn features directly from the input image. Convolutional neural networks (CNNs) have been shown to be effective at this task. The input image is passed through a series of convolutional layers, which extract features at different levels of abstraction. The features are then passed through fully connected layers to predict the age and gender.

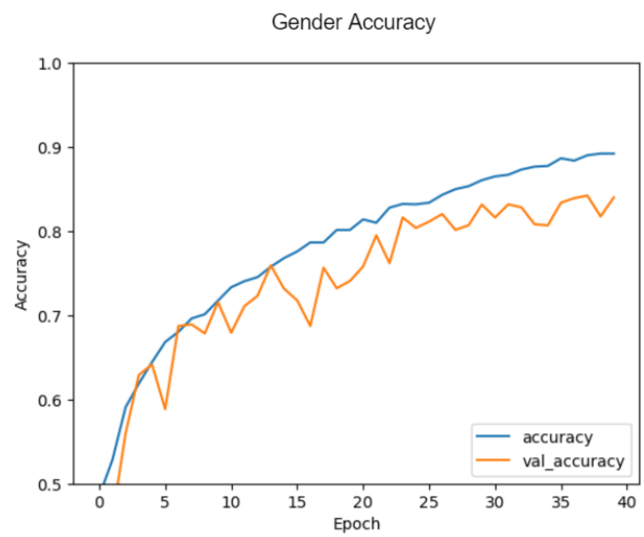
In the first,we evaluate our method for classifying a person to the correct age. We train our network to classify face images into eight age group classes and report the performance of our classifier on the Adience dataset. We also evaluate our method for classifying face images to the correct labels gender. We test the performance on the same dataset. For this task, we train our network for classification of three classes (female male and undefined) and report the result accuracy. We assess our solution for the age and gender classifications on Adience benchmark dataset.

The purpose is to predict whether a person's gender is within a precise age range. Based on gender and age classification using CNN. We tackled the classification of age group and gender of face images and posed the task as a multiclass classification problem as such train the model with a classification-based loss function as training targets.We perform data preprocessing on the training dataset image by the flatten on a one dimensional vector feature of input image for age and gender.

After successfully extracting features , these last will pass to the phase of classification by CNN , the result of the last layer CNN Based Features Extraction for Age Estimation of convolutional network is passed to three fully connected layer , finally outputs are given to the fully connected layers to minimize the classification loss on eight age classes (groups 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, and + 60 years old) , and by the sigmoid function on two gender (female and male) classes. As a result we got following :

5.1. Output obtained .





For age:

loss = 0.3179

accuracy = 0.8679

val_loss = 0.410

val_accuracy = 0.8351

For gender:

loss = 0.3479

accuracy = 0.8279

val_loss = 0.4306

val_accuracy = 0.7951

5.2. Evaluation measures used

For comparison, we also trained a model with different epochs and batch sizes. We used the following 2 variations:

1st variation:

1.1 20 Epochs for age, 256 batch size, 500 random images per batch

1.2 20 Epochs for gender, 256 batch size, 500 random images per batch

2nd variation:

2.1 40 Epochs for age, 512 batch size, 1000 random images per batch

2.2 40 Epochs for gender, 512 batch size, 1000 random images per batch

As it is clearly visible, the model with more epochs gives us a better output and better accuracy.

When fewer epochs are used, the accuracy is not stabilized.

6. Conclusion

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 reflect 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 CNN performs on these tasks using Internet data.

We provide results with a lean deep-learning architecture designed to avoid overfitting 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 overfitting. We further inflate the size of the training data by artificially adding cropped versions of the images in our training set.

The resulting system was tested on the Adience benchmark of unfiltered images and shown to significantly 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 classification 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 substantially improving results beyond those reported here.

In conclusion, age and gender prediction is an important problem in computer vision and machine learning. Deep learning-based approaches, particularly CNNs, have shown promising results in this area. The choice of dataset, feature extraction method, and evaluation metrics can have a significant impact on the performance of age and gender prediction models.

7. References

[1] Abhijit Das, Antitza Dantcheva and Francois Bremond Mitigating Bias in Gender, Age and Ethnicity classification.

Link:

https://www.researchgate.net/publication/328031032_Mitigating_Bias_in_Gender_Age_and_Ethnicity_Classification_a_Multi-Task_Convolution_Neural_Network_Approach

[2] Philip Smith, Cuixian Chen Transfer Learning with Deep CNNs for Gender Recognition and Age Estimation.

Link:

https://arxiv.org/pdf/1811.07344.pdf

[3] Sepidehsadat Hosseini, Seok Hee Lee, Hyuk Jin Kwon, Hyung Ii Koo and Nam Ik Cho Age and Gender Classification Using Wide Convolutional Neural Network and Gabor Filter.

Link:

https://www.researchgate.net/publication/325495265_Age_and_gender_classification_using_wid_convolutional_neur%20al_network_and_Gabor_filter

[4] Dataset Reference: Adience Benchmark Gender And Age Classification

Link:

https://www.kaggle.com/datasets/ttungl/adience-benchmark-gender-and-age-classification