Utrecht University

MSc Artificial Intelligence
Faculty of Science, Information and Computing Science

Do Alzheimer's Patients Appear Younger than Their Age? A Study with Automatic Face Analysis

Author:
Abdullah Emir Zeylan

Supervisor (UU):
Prof. dr. A.A. Salah

Second reader:
Dr. Z. Yumak

Supervisor (external):
Dr. H. Dibeklioğlu

July 1, 2020

Acknowledgement

I would like to thank my supervisors Prof. Dr. Albert Ali Salah and Dr. Hamdi Dibeklioğlu, for their guidance, support and patience during the master thesis. It was a great opportunity for me to work with them.

They provided critical feedback and interesting new information about deep learning. I am very grateful to work with such people who have a lot of experience in this field.

I would like to thank Dr. Zerrin Yumak, Dr. Zeynep Tüfekçioğlu, Prof. Dr. Başar Bilgiç and Prof. Dr. Murat Emre for participating in my thesis committee.

I also would like to thank my friends for their feedback on the dissertation and support during the thesis. It was a challenging task to write a thesis, but with their help, it was possible to finish my thesis.

Finally, the moral support that I needed of my family kept me going on for finishing the thesis.

Emir Zeylan, Utrecht, 2020

Abstract

Facial age estimation is a challenging task, especially if the subjects are older. The facial appearance of human face changes due to several lifestyle factors. Estimating the age of a person with a disease based on facial images is a relatively new topic. In this thesis, we test the hypothesis that Alzheimer's patients appear younger than their real age. To analyze this, we use automatic age estimation for this task. First, we propose training and normalization regimes to improve deep learning based facial age estimation. Then we fine-tune a pre-trained ImageNet model using first the APPA-REAL database and then the UTKFace database. The experimental results show that the proposed approach predicts older faces more accurately compared to other studies, and improves the mean absolute error for the FG-NET database to 8.14 for the age group 60-69. While the tested database is too small for broad estimations, the results on the current database show promising results. Then, we run our approach on a special database from patients with Alzheimer's Disease (AD) and healthy controls, to test our main hypothesis. The database contains video recordings of 96 subjects with an age range between 64 and 87. Our findings show that residual age estimation indeed underestimates the age of AD patients significantly more than the healthy subjects.

Contents

List of Figures

List of Tables

List of Abbreviations

AAM Active Appearance Model

AD Alzheimer's Disease

ANN Artificial Neural Network

AU Action Unit

AUs Action Units

BIF Bio-inspired Features

CD Cumulative Distribution

CLBP Completed Local Binary Patterns

CNN Convolutional Neural Network

ELBP Extended Local Binary Patterns

FACS Facial Action Coding System

GEF Gradient-based Encoded features

GF Gabor Filters

GW Gabor Wavelets

HOG Histogram of Oriented Gradients

IEF Intensity-based Features

KNN k-Nearest Neighbor

LBP Local Binary Patterns

LDA Linear Discriminant Analysis

LPQ Local Phase Quantization

LTP Local Ternary Patterns

MAE Mean Absolute Error

MLP Multi-layer Perceptron

mRMR Minimal Redundancy Maximal Relevance

NN Nearest Neighbor

PBVD Piecewise Bézier Volume Deformation

PCA Principal Component Analysis

QF Quadratic Functions

ReLU Rectified Linear Unit

SIFT Scale Invariant Feature Transform

SVM Support Vector Machine

SVR Support Vector Regression

VLBP Volume Local Binary Patterns

Chapter 1

Introduction

Predicting the age of another person is an obvious thing that we do. However, sometimes we could estimate it wrong or close to their real age. Actually, why do we choose a specific age? On which factors do we base our predictions? Do features such as wrinkles have any influence or glasses, beard, hair color and what about facial expressions? As it is simple for humans to observe these facial appearances, it is a difficult task for computers. In decade years, many computer vision approaches have proposed in age estimation for different kind of applications, such as forensics [2]. However, this field tackles with problems that it faces due to some factors influencing the aging of a person. During the aging of a human face, the facial appearances changes due to living conditions and medical status, plastic surgery, scars, and make-up [98].

Automatic facial age estimation requires a large amount of information from the image data. Extracting this information is essential for the performance of an age estimation system, which depends on the quality of the extracted features. Most extracted features in facial age estimation are based on the static appearances of human face.

Mean absolute error (MAE) is mostly used in age estimation systems. A recent study by Han et al. found that humans can estimate the age with MAE of 7.2–7.4 years [38]. Several automatic age estimation has estimated the age with MAE using various approaches. Automatic age estimation from facial images used the K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) to estimate the age of the human face [62]. The results for KNN and SVM were MAE of 8.43 and 5.71, respectively. Another automatic age estimation approach based on deep learning achieved MAE of 2.96 [15].

AE based on facial images is still a challenging problem and further from

real acceptance and practical satisfactory [28]. So far, an extensive attention has been paid by researchers on AE and many approaches have been proposed to solve this problem. Mainly, many aspects which make AE as a challenging problem:

A challenge of age estimation is estimating the age of patients with Alzheimer's disease (AD). The problem is that the increased variance in living conditions affects the age of AD patients. Thus, it is more difficult to estimate the age of older people. Although there are other risk factors for AD patients as well [14]. A study in sex shows that women have more risk in AD after the age of 85 years old [5]. Other studies researched the relation of drugs with AD. As drugs are risk factors of dementia, those studies show that taking the Nonsteroidal anti-inflammatory drugs (NSAIDs) significantly reduces the risk of Alzheimer's disease [45, 106, 26]. Besides those factors, AD affects the facial expression of the human face. Patients with AD cannot express emotions well like healthy persons [9]. They struggle with expressing their emotions due to impairments of nerves. The patients will lose their cognitive capacity and even their verbal communication ability.
s The current automatic age estimation approaches, humans as well, make more mistakes on elderly subjects. Most studies focus on impairments in face recognition in AD patients [41]. However, image-based facial age estimation for patients with Alzheimer's disease have not been researched for age estimation problems.

In this thesis, we will analyze the age of Alzheimer's patients. Additionally, applications are too narrow for estimating age from human faces with a disease in a broad context. We aim to investigate the possibility that patients with Alzheimer's disease seem younger than healthy subjects.

1.1 Research questions

The goal of this thesis is to implement a state-of-the-art facial age estimation technique to estimate the age of AD patients and healthy subjects. We use a Convolutional Neural Network (CNN) architecture to extract features and classify the age into range of 0-100. The following research questions summarizes the the goal of the research.

(i) How can we adapt a deep learning technique for automatic age estimation?

(ii) How can we estimate the age of Alzheimer patients by static images and compare the estimation to healthy subjects? and how well do they

perform?

(iii) Can we improve the state-of-the-art automatic age estimation for estimating the age for elderly?

1.2 Contributions

This research targets the following contributions:

(i) The state-of-the-art of age estimation for the elderly has been improved and extended.

(ii) An automatic image-based facial age estimation for Alzheimer's patients has been conducted.

(iii) Proposed training and normalization regimes that improved the deep learning based facial age estimation.

(iv) Published a paper to IPTA Conference "Do Alzheimer's Patients Appear Younger than Their Age? A Study with Automatic Facial Age Estimation" [107]

1.3 Overview

The thesis is organized in six chapters. The first chapter introduces the research problem and brief reviews existing work of age estimation. Chapter 2 describes theory behind age estimation and face analysis. Chapter 3 provides an overview of the implemented frameworks that constitute this thesis. Chapter 4 shows the conducted experiments with results. Chapter 5 discusses the limitation of our approach. Finally, chapter 6, gives an conclusion to our work and presents the future work.

Chapter 1 introduces information about the research field, problem statement and states the principal aims for this research. The main contributions are also described.

Chapter 2 informs the reader about the theoretical background behind age estimation. This chapter provides information to understand the human face such as facial appearances, features and expressions. This chapter also describes a range of different feature extraction and classification

techniques. Further, analysis of Alzheimer's is also proposed in this chapter. Finally, a brief review of different literature's in facial age estimation is discussed.

Chapter 3 introduces the methodology of the automatic age estimation framework, including pre-processing, feature extraction, classification and statistical analysis.

Chapter 5 gives an overview of the conducted experiments of automatic age estimation. Further, results of the conducted experiments are presented.

Chapter 6 discusses the limitations of our work.

Chapter 7 presents the conclusion and future work.

Chapter 2

Background

This chapter addresses the theoretical background of age estimation and describes various methods and techniques that are relevant for facial age estimation. A general facial age estimation pipeline includes face pre-processing, feature extraction and classification. Figure 2.1 illustrates this general pipeline for an age estimation framework with CNN.

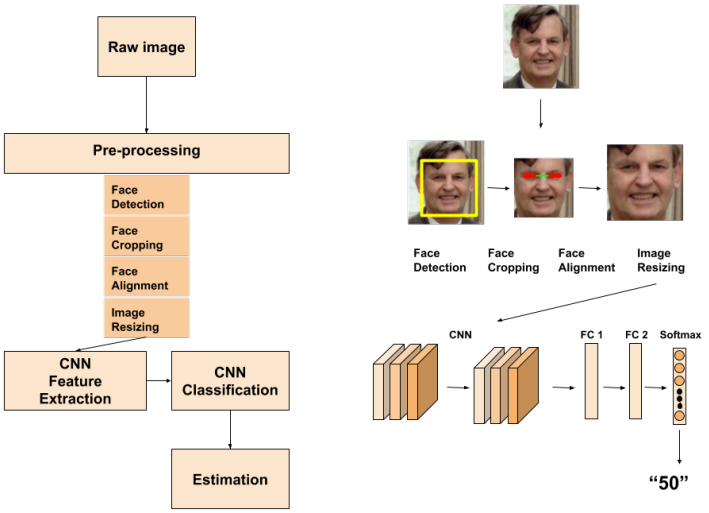


Figure 2.1: A general age estimation pipeline for CNN. The raw image is inserted; The image is pre-processed with face detection (e.g. Viola-Jones), cropping and alignment (e.g. Facial landmark) methods, then resized to a specific dimension for CNN; The CNN is used for feature extraction and classification purposes; The output of the CNN is an age that is estimated from the facial image.

This chapter is separated in three sections: (1) Understanding the human face, (2) The general process of a facial age estimation framework, and (3) Analysis of the human face, which includes facial expressions and face analysis of the pathology.

2.1 A process of the aging a human face

Aging of a human face is a slow process that includes changes in bones, fat mass, and skin texture [113, 16]. A face of a young person is slightly shorter and the skin patterns as well compared to an older person, because of the growth and development of the face during the aging [104]. The shape development, which is the growth of the craniofacial skeleton, is the most noticeable change of the human face [2]. A craniofacial change remodels the cranial bone horizontally and vertically. During the aging, the size of the face increases horizontally and decreases vertically. The cranial bones remodel around the age of 40-50 years and facial lines are increasing in depth. These changes are affecting the appearance of the face and its expressions

[70, 87]. The positions of the nose and chin are influenced by the length of the dental arch when it is decreasing. Besides the craniofacial growth, there appear changes in skin texture in the later stages of aging. The changes in skin texture do not change greatly compared with craniofacial growth. The most known skin patterns that occur on the face are wrinkles and pigmentation [82]. Fat tissue and muscular changes are some of the factors that are responsible for wrinkles. The mimetic muscles control the facial expressions which create wrinkles on the face [44].

Figure 2.2: Aging of a human face. *From Changes Associated with the Aging Face by Oren Friedman* [27].

A facial age of a human can be affected by many factors, besides the biological factors [98] and environmental factors [2]. Factors like lifestyle, natural, and psychological are separated into intrinsic and extrinsic categories [52]. The intrinsic factors are internal such as changes in bone structure and genetic influences over time [113]. The extrinsic factors are external factors influencing facial aging, such as drugs, smoking, and exposure to ultraviolet radiation [99]. Human facial skeletons have lifelong and continuous changes [70, 87].

2.2 Facial features

Features contain a piece of information that is relevant for solving problems related to specific tasks such as classifying ages from facial images. Age estimation based on static images categorizes facial appearance into various

kinds of features. In this section, we will discuss the static appearance features that exist on the face. Those features can be extracted with feature descriptors which are described in section 2.3.2.

Features are of importance for detection problems in computer vision. They affect the classification performance of the age estimation problem. Features may influence facial appearances like the age, ethnic origin, gender, facial hair, light, expression and more of a human face [102]. Especially, features like the eyes, nose, mouth, chin, and the forehead are primarily detected in face detection and valuable for facial age estimation. For example, large eyes make a person appear more empathetic [77]. The appearance features can be separated into geometric and appearance-based features [8]. Geometric features are features such as the shape of the mouth, nose, or eyes. Appearance-based features are features which are attempted to extract information about spatial patterns of light and dark in the face image.

Facial features can be categorized into local, global, and hybrid features. Local features are features such as wrinkles on the face, skin deformations, and geometric features. These features are robust to illumination [84]. Global features contain information such as identity, expression, gender, appearance, and the shape of a face. Those features represent the face model for the aging process. Local and global features can be combined into hybrid features [97, 33].

2.3 Face processing

This section describes the general approach of face processing methods which includes pre-processing, feature extraction, and classification. These are described in detail in the following subsections 2.3.1, 2.3.2 and 2.3.3, respectively.

2.3.1 Pre-processing

The general age estimation frameworks start with a pre-processing step. Pre-processing is an approach for images that aims to improve the image data for suppressing unwilling distortions or enhances some image features important for further processing [94]. Some examples are the geometric transformations of images such as rotation, scaling, and translation.

Age estimation frameworks have to pre-process the image data before extracting features to increase the performance of the model. The pre-processing step in age estimation usually relates to face localization, tracking, and registration to remove the variability due to changes in pose and illumination. A normalization, alignment, and histogram equalization can be applied to deal with these problems. Also, a dataset that contains noisy images, which affects the performance of the system should be removed. Several factors can influence face detection, which makes a very challenging task to detect faces.

The factors that influence the image and face detection are described below:

Pose A face can be rotated due to the camera. The pose of the face may rotate in θ degrees, where objects can occlude the parts of the face like eyes or mouth.

Presence of appearances Appearances as beards and glasses may present on the face. The features of the will be occluded by these appearances.

Facial expression The appearance of faces are directly affected by a person's facial expression.

Occlusion A group of faces may partially occlude other faces. Although, other objects may occlude as well.

Image orientation In addition to the rotation of faces, images can also be rotated. Face detection algorithms may have difficulty to detect faces from rotated images.

Image conditions External factors such as lightning and camera characteristics may affect the appearance of a face in the image.

Facial images should be normalized to solve the mentioned factors. Basically, normalization is a process that will change the pixel values and facial images will be prepared for feature extraction. A Linear Normalization will "stretch" the intensity range of an image as

$$I_n = (I - x_{min})\frac{y_{min} - y_{max}}{x_{max} - x_{min}} + y_{min}, \tag{2.1}$$

where x_{min} and x_{max} are the minimum and maximum pixel intensity and y_{min} and y_{max} the new pixel intensity values. I_n and I are normalized pixel and intensity of original pixel, respectively.

A histogram equalization can be applied to improve the contrast of the image. It simply changes the pixel intensity values to a specific range of values. It usually increases the global contrast of images when its usable data is represented by close contrast values. Areas with lover local contrast will gain a higher contrast.

Given a facial image, where intensity is described as i and the probability density function is P_x. The number of possible intensity value, L, is frequently 256. So

$$P_x = \frac{n_i}{N}, 0 \leq i \leq L, \tag{2.2}$$

where n_i is number of pixels with intensity i and N total number of pixels. Then a cumulative distribution function, C_x is then defined from P_x as

$$CDF_x = \sum_{j=0}^{i} P_x(j) \tag{2.3}$$

The following equation is transforming the input image's intensity r_i to into output image's intensity s_i

$$S_i = T(R_i) = Round(CDF_x \times (L1)), \tag{2.4}$$

rounded down to the nearest integer.

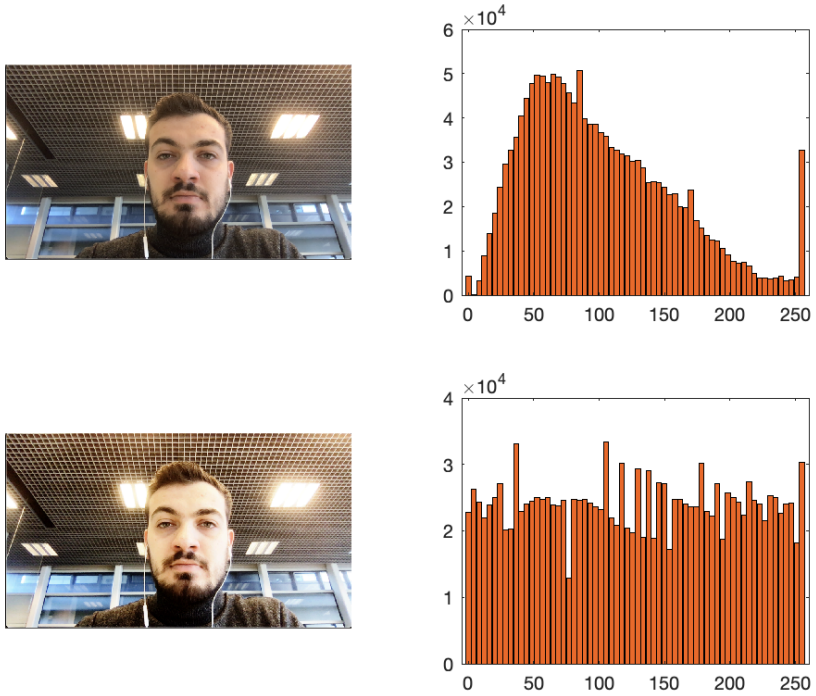


Figure 2.3: An image and corresponding histogram achieved using histogram equalization.

The preprocessing step is simply changing the input data and returns a new output that can be used for face detection and feature extraction. The next section will discuss the face detection.

2.3.1.1 Face detection

Face Detection is usually finding faces and facial regions such as the eyebrows, mouth, nose, forehead, and the iris in a facial image. Viola-Jones is one of the first and well-known algorithms for face detection. This algorithm is proposed by Paul Viola and Michael Jones [101]. The main idea of this algorithm is to locate the human face without regard to its size, background, and situation. The pipeline of this algorithm includes four phases: Haar feature selection, creating an integral image, AdaBoost training, and cascading classifiers. One of the advantages of this algorithm is that it can detect faces in real-time. The algorithm has a high detection accuracy for detecting faces. Another advantage is that it can train a cascade of classifiers for reducing

the detection time. The Viola-Jones face detection is most effective in frontal facial images. Due to multiple detections of the same face, we might get some overlapping of the bounding boxes.

Haar-like features

The Haar-like features are used together with integral images representation. As shown in Figure 2.4, there are four types of Haar-like features. These features have different contrasts and are applied to the facial feature to figure out whether the facial feature is present or not. For example, the eyes are darker than the upper cheeks, nose, and forehead. The Haar-like features are divided into two blocks: black and white, where the amount is determined by the sum of pixels within this area.

Figure 2.4: Haar-like features: (1) and (2) Edge features; (3) Line feature; (4) Four-rectangle feature.

Integral Image

Integral Image is a technique that calculates the blocks of Haar-like features in an image. The Integral Image is also known as the summed area table. It can efficiently compute the sum of the black and white areas. Each pixel equals to the entire sum of all pixels of left and top neighbors of the concerned pixel. A pixel at location (x,y) = $\sum LP_xy + TP_xy$, where LP and TP are left and top neighbours, respectively. For example, the integral image from figure 2.5 calculated the pixels in cell ABCD as $A - B - C + D$, also known as

$$II(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y'), \tag{2.5}$$

where $II(x, y)$ the pixel value of the original image.

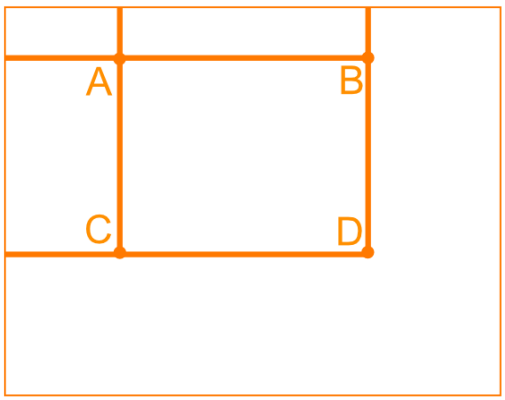


Figure 2.5: The integral image.

AdaBoost

The Viola-Jones algorithm uses an AdaBoost learning technique. This learning technique creates one strong classifier from a specific amount of weak classifiers. These weak classifiers work slightly better than a random classifier. The weak classifiers, each feature is considered as one weak classifier, are trained in a sequence, where at each step the current weights are used to produce new weights for the next step. At each step, a threshold is used that separates the positive and classified examples. A weak classifier is described as

$$h(x.f.p,\theta) = \begin{cases} 1 & \text{pf}(x) > \text{ p } \theta \\ 0 & \text{otherwise} \end{cases}, \tag{2.6}$$

where x, is a $M \times M$ pixel sub-window, f an applied feature, p polarity and θ threshold, which decides if a face needs to be classified as positive or negative face.

Cascade classifier

Cascade classifiers is a process where strong classifiers are cascaded. Every stage contains a strong classifier, where sub-windows with wrong faces are rejected and correct faces shifted to next classifier.

2.3.1.2 Face alignment

Face alignment is another preprocessing step, where facial landmarks are localized after face detection. After detecting the positions of the two eyes, it can be transformed into a new set of coordinates. Let (R_x, R_y) and (L_x, L_y) the corners of right and left eyes, respectively. The face region should be cropped after the transformation. An affine transformation can be used

to scale, rotate and transform the face with new coordinates (R'_x, R'_y) and (L'_x, L'_y).

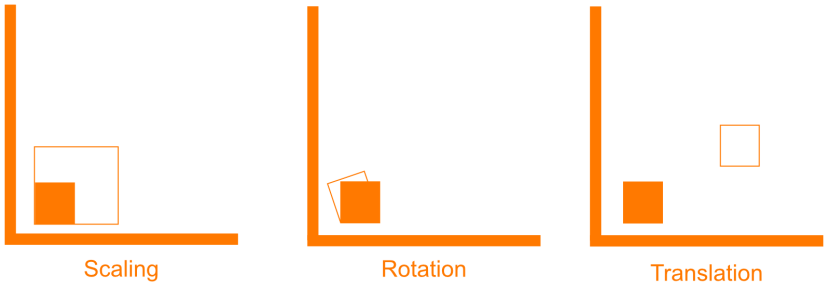


Figure 2.6: The affine transformation: from left to right, scaling, rotation, and translation.

A translation is moving an object from coordinate (x, y) to a new coordinate (x', y'). The translation factor with homogeneous coordinates $\begin{vmatrix} x' & y' & 1 \end{vmatrix}^T$ is defined as

$$\begin{vmatrix} x' \\ y' \\ 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & T_x \\ 0 & 1 & T_y \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} x \\ y \\ 1 \end{vmatrix} \tag{2.7}$$

The rotation factor is rotates the coordinate (x, y) by a specific angle θ. Rotation is defined as

$$\begin{vmatrix} x' \\ y' \\ 1 \end{vmatrix} = \begin{vmatrix} cos\theta & -sin\theta & 0 \\ sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} x \\ y \\ 1 \end{vmatrix} \tag{2.8}$$

The Scale factor changes the size with the coordinate (x, y) by a specific angle θ. The θ is a degree that will be rotated counterclockwise. Rotation is defined as

$$\begin{vmatrix} x' \\ y' \\ 1 \end{vmatrix} = \begin{vmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} x \\ y \\ 1 \end{vmatrix} \tag{2.9}$$

The transformation composited together is defined as

$$\begin{vmatrix} x' \\ y' \\ 1 \end{vmatrix} = \begin{vmatrix} S_x cos\theta & -S_x sin\theta & S_x T_x \\ S_y sin\theta & S_y cos\theta & S_y T_y \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} x \\ y \\ 1 \end{vmatrix} \tag{2.10}$$

A facial image is aligned with several steps. The first step is to detect facial landmarks in the face. Those landmarks are used to localize the regions of the face appearance, such as the eyes, mouth, jawline, and more. The facial landmarks detects 68 or 5 points in a face (Figure 2.7), based on the technique. A facial landmark approach starts with a training set of manually labeled facial landmarks [51]. The landmarks are specified at specific x and y coordinates of face regions. Then the training dataset is trained with an ensemble of regression trees to estimate the facial landmark positions. The trained facial landmark detector will detect the x and y locations of 68 or 5 landmarks. The face can be aligned with the affine transformation based on the detected locations of facial landmarks. For example, to do rotation from the affine transformation, the coordinates of the eye landmarks are used to find the angle of the center point between the eyes. Then the rotation matrix of the affine transformation is applied.

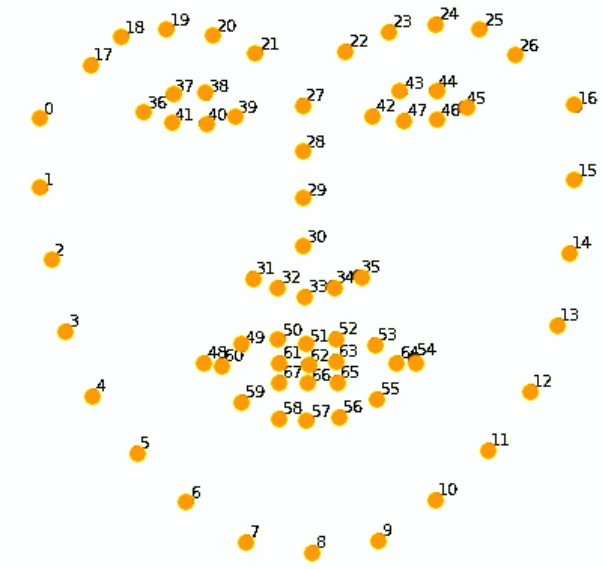


Figure 2.7: The 68 facial landmarks of human face.

2.3.2 Feature extraction

This chapter will discuss several feature extraction techniques that can be applied after face pre-processing. A feature extractor algorithm output a feature vector that represents the characteristics of an object in the image. The Following algorithms are common in age estimation and those will be explained: Local Binary Pattern (LBP), Biologically Inspired Features (BIF), Intensity-based Features (IEF), Gabor Filters, Local Directional Pattern (LDP), Histogram of Oriented Gradients (HOG), Scale Invariant Feature Transform (SIFT). There are also some feature reduction and selection algorithms that will be discussed. Feature reduction in image concerns with the dimensions. For example, when you want to reduce the resolution of an image, you also reduce its dimension. So, you downsample the image and pick only the important pixels and throw the others. The following feature reduction algorithms are discussed: Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), and Minimal-redundancy-maximal-relevance (MRMR).

Local Binary Pattern

The local binary pattern is a popular texture descriptor proposed by Ojala et al [74]. This descriptor computes a local representation of texture. The basic principle behind this idea is that each pixel will be compared to the center pixel in a 3x3 neighborhood. Let (x,y) be the position of the center pixel. A binary pattern of 3x3 neighborhood will be created by comparing each pixel, counterclockwise, started from left top pixel x_{-1}, y_{+1}. When intensity is $(x_{-1}, y_{+1}) < (x, y)$ the binary pattern pixel at x_{-1}, y_{+1} becomes 0, otherwise 1. The neighborhood contains a total of 8 pixels with 256 possible combinations that will return a label, known as LBP-Code. The LBP is more robust against variations in pose or illumination.

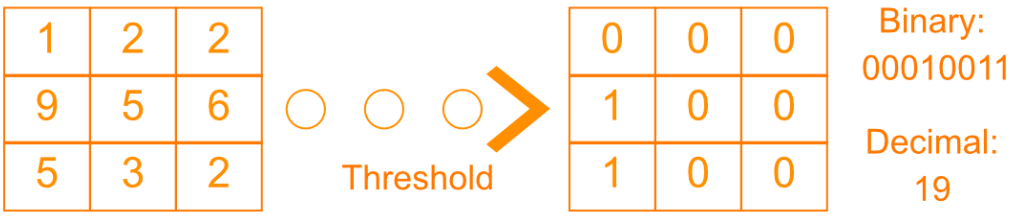


Figure 2.8: Local Binary Pattern with LBP-Code.

The LBP has recently extended with several variations such as Extended LBP (ELBP) [42, 43], Completed LBP (CLBP)[36], Volume LBP (VLBP)

[112, 37] and combinations with other feature descriptors as SIFT [40] and Gabor Wavelets [110]. The ELBP is an extension of the original LBP. The ELBP used a variable neighborhood to align an arbitrary number of neighbors on a circle in a specific radius. A point in the circle that does not correspond to the image coordinates will be interpolated. The basic LBP approach uses bilinear interpolation, where the values are interpolated by the average of its closest neighbors. Another extension of the LBP is using uniform patterns, where a binary pattern contains at most two bitwise transitions [75]. The following patterns are 01110000 and 11001111 are uniform. The uniform patterns correspond only to 58 labels.

The methodology for using LBP based on face images is proposed by Ahonen et al. where the face image is divided into local regions [1]. Each region extracts the LBP features which are concatenated to form a global description of the face. A histogram will be created by calculating the LBP-Code for every pixel in the image. The local descriptions will be combined into a global descriptor.

Biologically Inspired Features
Biologically Inspired Features for age estimation are proposed by Guo et al [35]. Basically, the BIF consists of two layers, where the first layer, S1, convolves an array of Gabor filters at different orientations and scales. This layer can be defined as

$$g(x, y) = exp(-\frac{x^2 + \gamma^2 y^2}{2\zeta^2}) \times cos(\frac{2\pi}{\chi}x), \tag{2.11}$$

where x = xcosθ + ysinθ and y = xsinθ + ycosθ are the rotations of the Gabor filters for angle θ varies between 0 and π. The orientation θ varies from 0 to π uniformly with different intervals, such as 4, 6 and so on.

The second layer, C1, takes the maximum values within a local spatial neighbourhood. Guo et al. introduces the standard deviation operation in creating the second layer and making the number of bands and orientations adaptive to the data. The second layer is defined as

$$SD = \sqrt{\frac{1}{\widehat{N}_S^2} \sum_{i=1}^{\widehat{N}_S^2} (F_i - \overline{F})^2}, \tag{2.12}$$

which reveals the variability in the data within a neighborhood $\widehat{N}_S^2$ of S1 units, where F_i is the maximum value of two consecutive S1 units in the

same scale band S (there are 8 bands in total but using different filters) at pixel index i,

$$F_i = max(x_i^j, x_i^{j+1}), \tag{2.13}$$

where x_i^j and x_i^{j+1} are the filtered values with scales j and j + 1 at position i. $\overline{F}$ is the mean value of the filtered values within the neighbourhood $\widehat{N}_S^2$. The pooling of the "MAX" operation over two consecutive scales increases the tolerance to 2D transformations, such as scale changes with a small amount. The "MAX" operation merges two filtered images using filters of the same orientation but different scales into one, and then the "STD" operation is performed on the merged image within a neighbourhood $\widehat{N}_S^2$.

Intensity-based Features

Intensity-based Features is an intensity-based texture descriptor to capture skin texture [3]. Each facial region is divided into two code sets. One code set is learning from the sampled vectors which are extracted from the whole face and the other from one face patch. The vectors are encoded and constructed with the frequency histograms. The intensities for each neighbor pixel are sampled in ring-based pattern to form a low-level feature vector. Although, the orientation histogram of local gradients in the neighborhood are extracted due to local sensitivity to image noise and illumination variations like wrinkles.

Gabor Filters

Gabor filters can be used to detect edges in images. It is commonly used in edge detection and feature extraction. The Gabor filter gives the highest response at edges and at points where texture change. It works as a bandpass filter for local spatial frequency distribution. A 2D Gabor filter is usually considered as a Gaussian kernel function modulated by a sinusoidal plane wave. The 2D Gabor function is defined as:

$$g_{\lambda,\theta,\varphi,\sigma,\gamma}(x,y) = exp(-\frac{x^2+\gamma^2y^2}{2\sigma^2})cos(2\pi\frac{x}{\lambda}+\varphi), \tag{2.14}$$

where x = $xcos(\theta) + ysin(\theta)$ and y = $-xsin(\theta) + ycos(\theta)$.

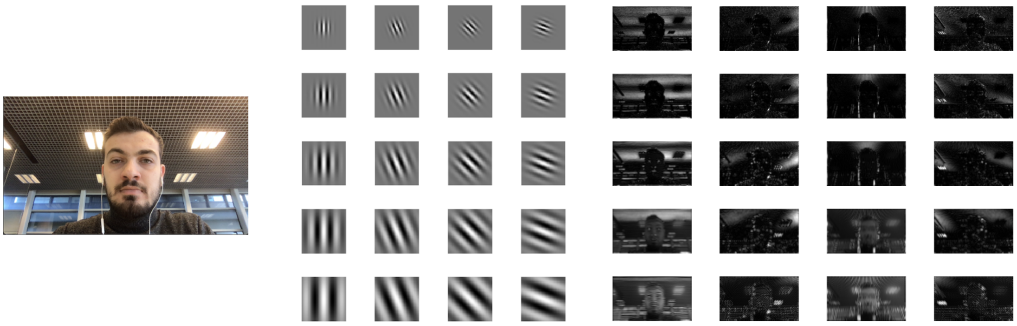


Figure 2.9: Left input image, center cropped image and right Gabor filters.

Local directional patterns

Local Directional Pattern was proposed by Jabid et al. [47] The LDP is robust against noise and illumination. An 8-bit binary code is computed by comparing different edges in different orientations. The LDP may use several edge detectors like Kirsch edge detector [54], Prewitt edge detector [81] and Sobel edge detector [93]. The Kirsch edge detector performs more accurately than Prewitt and Sobel edge detectors, which considers all eight neighbours [60].

Histogram of Oriented Gradients

Histogram of Oriented Gradients is a feature descriptor that is used for detection [18]. The objective is that occurrences of gradient orientation in localized parts of an image play important roles. The idea behind HOG is that the appearances and shapes of local objects within an image can be well described by the distribution of intensity gradients as the votes for dominant edge directions. The algorithm of HOG includes gradient computation, orientation binning and block normalization. The HOG tries to capture information about gradients. A face image will be divided into $n \times n$ cells and the histogram of gradient directions are compiled for each pixel in the cells. Figure 2.10 views the histogram of oriented gradients for the cropped face image.

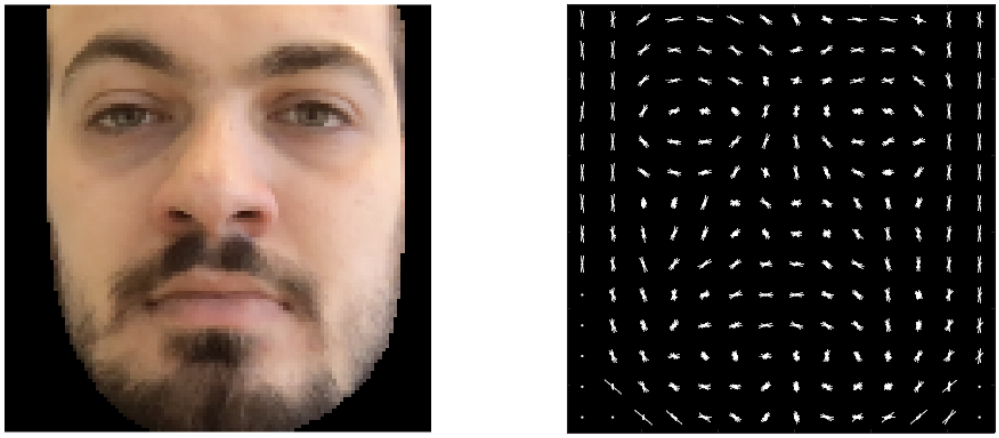


Figure 2.10: Left aligned image and right Histogram of Oriented Gradients.

Scale Invariant Feature Transform
Scale Invariant Feature Transform is a feature detector and descriptor proposed by Lowe [64]. The idea is that the image is transformed into local feature coordinates. These coordinates are invariant to image parameters as translation, rotation, scale. The SIFT searches for stable features across multiple scales. Histograms of oriented gradient directions will be applied to determined key points. These key points correspond to edges, corners and other informative changes in the image. Finally a descriptor will be computed for the local image regions.

Feature selection and reduction
Using dimensionality reduction techniques the number of random variables could be reduced without losing much information. The Principal Component Analysis is one of the approaches, which extracts the important features as a component [50]. PCA reduces multidimensional data to lower dimensions while retaining most of the information. It covers standard deviation, covariance, and eigenvectors. The Linear Discriminant Analysis finds a linear combination of features that may separate two or more classes of objects [71]. The objective is to reduce the dimensions by removing the redundant and dependent features by transforming the features from higher dimensional space to a space with lower dimensions. Besides feature reduction techniques, there also exist feature selection approaches. These approaches include and exclude attributes present in the data without changing them. Minimal-redundancy-maximal-relevance is a feature selection method proposed by Peng et al. [78]. This approach selects the features with the highest relevance with the target class and is also minimally redundant.

2.3.3 Classification

Age is a number of elapsed years since birth to a point in life. The estimation is a process of estimating someone's actual age. Generally, Facial age estimation techniques includes two components: feature extraction (section 2.3.2) and classification (section 2.3.3). The classification problem classified each face image into age classes, where each age is considered as a class label. In the regression problem, each age is considered a regression value. Both problems can be combined as a hybrid approach to finding the relationship between extracted feature vectors and age labels. There exist several algorithms to solve the age estimation problem. One of the most popular algorithms for solving age estimation problems are Support Vector Machine (SVM) [17], k-Nearest Neighbor (KNN) [4], and Convolutional Neural Network (CNN) [61].

A SVM is an algorithm which that be used for classification and regression. It performs classification by finding the hyperplane that maximizes the margin between the two classes. The advantages of using the SVM are that it performs well on datasets with numerous attributes and uses a subset of training points in support vectors, which is memory efficient. A few limitations exist such as it does not perform well on a large dataset, while the required training time is higher than training a small dataset. One of the biggest challenges for the SVM classifier is determining the kernel and regularization parameters. A wrong choice in kernel moves to the problem of overfitting.

KNN is another algorithm for solving classification problems in age estimation. Similarly to SVM, it can be used as a classifier and regressor but mostly used for classification. It is This is a technique for classifying objects based on closest training examples in the feature space. It is a simple algorithm that stores all available cases and classifies new cases by a majority vote of its k neighbors. This algorithm separates unlabeled data points into well-defined groups. This algorithm comprises storing feature vectors and labels during the training. The limitations of the KNN algorithm can be very expensive in memory usage. One of the advantages of KNN is a high accuracy, but this is not competitive compared to other supervised learning models.

Recently, CNN is introduced to solve age estimation problems as well. It is generally used to analyze image data and can learn spatial hierarchies of features through backpropagation using layers as convolution layers, pooling layers, and fully connected layers. An advantage of using CNN is that the

feature extractor can be trained to a new task. So, the attributes from a trained CNN can be feed to another task. This saves computation time and also avoids training the network from scratch. Another advantage is that CNN performs well on a large dataset compared to SVM. However, using a large dataset is computationally expensive when training the network. One of the common challenges of CNN is overfitting, where the network does not generalize well from the training data. There are also other approaches used for estimating ages from facial images, such as Multi-layer Perceptron (MLP) and Artificial Neural Network (ANN) [55, 57].

2.4 Analysis of facial expressions

The definition of facial expression is described as the change in the face due to the activity of the facial muscles. They play an important role in revealing the psychological state of a person [95]. Paul Ekman conducted that a human face can express 6 basic emotions: happiness, fear, anger, disgust, surprise, and sadness.

As mentioned before, facial expressions can be measured with FACS. The FACS describes 44 unique action units and each represents a certain component of facial muscles movement. Table 2.11 describes the several AUs and in table 2.12 each emotion with the set of AUs are described. The expressions are encoded with FACS, that is using some rules to detect AU's in the facial image.

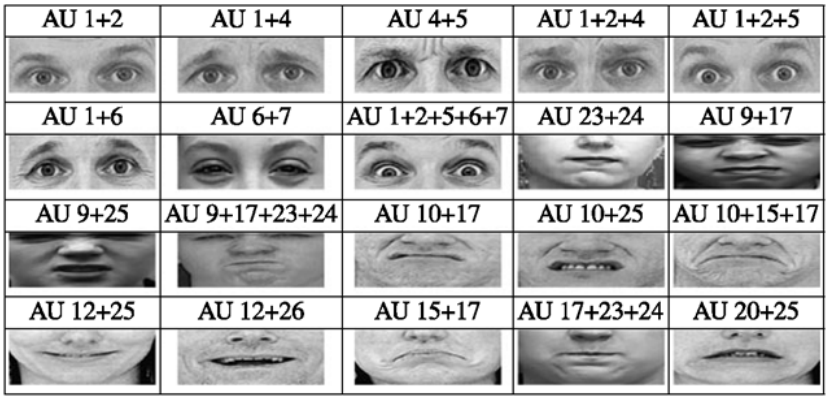


Figure 2.11: Example of Action Units.

Analyzing facial expression begins with a face acquisition phase, where

face processing is applied. The face and their landmarks will be detected in the image. The detection can be used on a single frame and or set of frames. After the face is located, the geometric and appearance features can be extracted that are caused by facial expression. The last phase is to classify facial expression.

To find specific frames from video data the corresponding onset, apex, and offset phases can be calculated. The Onset is the initial phase of facial expression, and it defines the duration from neutral to expressive state. The apex phase is the phase between the onset and the offset, which is a stable peak period of facial expression. Likewise, the offset is the final phase from expressive to neutral state [49].

Basic expressions	Involved Action Units
Surprise	AU 1, 2, 5,15,16, 20, 26
Fear	AU 1, 2, 4, 5,15,20, 26
Disgust	AU 2, 4, 9, 15, 17
Anger	AU 2, 4, 7, 9,10, 20, 26
Happiness	AU 1, 6,12,14
Sadness	AU 1, 4,15, 23

Figure 2.12: Action Units for six basic emotions.

2.5 Face analysis for pathologies

Dementia is a general name for diseases where brain activities are not functioning as expected. The most common form of dementia is Alzheimer's disease [56]. This disease involves the parts of the brain that controls memory, thoughts, and language. Age is the greatest risk factor for Alzheimer's disease. This mostly occurs at persons after the age of 65 years, however, there are young patients with this disease. Although, it is not clear how patients get this disease, so any person may get Alzheimer's disease. However, Alzheimer's disease can be caused by a genetic mutation passed from one generation to the next. This is known as familial Alzheimer's disease, which occurs at a young age.

Alzheimer's disease can be excluded into several stages. At the early stage, it starts with small brain impairments that affect learning, memory, and thinking. During the mid-stage, the person starts with struggling with communication problems that will impact their personal life. In the late stage, most of the brain has been damaged and the person will lose several

abilities such as memorizing and communicating.

According to Seidl et al., Alzheimer's disease influences facial expression [91]. Patients with the disease have issues in expressing and recognizing facial emotions [19, 89]. Losing this ability leads to the loss of appropriate behavior in social settings [9]. Smith studied facial expressions of subjects with Alzheimer's disease [92]. The Facial Action Coding System (FACS) has been used tor analyzing the expressions from videos. A higher amount of negative facial expressions were observed from the subjects when they viewed sad vignettes. According to Smith, negative facial expression increases by mild Alzheimer's disease.

2.6 Related work

Automatic age estimation from facial images has been studied for around 20 years. Researches encountered several challenges during these periods and tried to automate age estimation with computers. One of the challenging problems is in the field of face analysis. Especially, illumination, pose, resolution, and facial expression variations made age analysis difficult [29]. The automatic age estimations with computers show promising results compared to humans in the case of when the image conditions are good and the algorithms are trained with large datasets [38]. The best way to start with discovering image based age estimation is to survey papers about how a human face is progressing during the aging. Several papers discussed the understanding of the aging process of a human face [29, 84, 6]. They tried to analyze answers to questions like how humans perceive aging and how does aging impact the facial estimation performance.

An accurate age estimation task requires large annotated datasets with facial images. It is difficult task to collect datasets that contain facial images including a great differences in age range. The most common datasets used in age estimation are described below:

- The PAL contains 575 images within four age groups [72]. The age ranges between 18 and 70+ years old. The groups are as follow: 18-29 containing 218 images; 30-49 containing 76 images; 50-69 containing 123 images; 70+ containing 158 images.
- The FG-NET includes 1,002 images from age range 0-69 years old [28]. The database has annotations for different human races and contains

diversity of head poses and some facial expression as well as some illumination on the images.

- The MORPH dataset consists of 2 albums [86]. The first album contains 1724 images and second album has 55,134 images. The ages of the individuals range from 16 to 77. The dataset is labeled with race, gender, birth date, and acquisition date.
- The IMDB-WIKI includes 523,051 and is one of the largest datasets containing facial images [88]. The data is collected from IMDB and Wikipedia. The ages ranges from 0 to 100 years.
- The Adience contains 26,580 images divided into 8 age groups from an age range from 0 to 60 years old [24, 61].
- The WebFace has 77,021 facial images where age ranges from 1 to 80 years old [73]. The facial images are collected from Flicker and Google Image..

The facial age estimation task can be separated into three components: pre-processing, feature extraction, and age estimation. The aim of the preprocessing is to align and normalize the features in the face. Any preprocessing starts with a face detection task, which is detecting the face from the facial image. Recent approaches to face detection algorithms use deep neural networks to detect face and facial landmarks. A recent face detection framework was proposed by Zhang et al. [108]. A multi-task cascaded CNN's based framework is developed for alignment and joint face detection. This approach outperforms other state-of-the-art face detection methods. Face detection is separated in different stages: fast proposal network, refinement network, and output network. The network produces a bounding box and predicts face and landmark locations. Qin et al. used similar joint training cascaded CNN's for face detection [83]. This method achieves better accuracy compared to the proposed approach of Zhang et al. [108]. Recently, Yang et al. proposed a new face detection model through Deep Facial Part Responses [105]. Generic object proposal approaches are used to detect faces from the model. Those approaches are often used to provide high-quality and category-independent bounding boxes. Ranjan et al. proposed a deep learning classification algorithm for face detection, locating facial landmarks, and age estimation called HyperFace [85]. They used two different deep learning networks: AlexNet and ResNet-101 for constructing the framework. The HyperFace is based on a multi-tasking network to combine the intermediate layers of the network and fuse the features. Wu et al. tweaked an existing

vanilla convolutional neural network with fine-tuning and alignment-sensitive data augmentation to obtain more accurate landmark detection from images [103].

Another important task of pre-processing is face alignment. The face alignment is applied after the face detection. This technique identifies the geometric structure of the human face. It aims to align the face based on the translation, scale, and rotation. The most face alignment techniques rely on facial landmarks. Various approaches have been proposed to detect facial landmarks. Burgos-Artizzu et al. proposed a facial landmark estimation to detect landmarks from faces to deal with the occlusions due to differences in the pose, expression, and other [11]. Their approach, Robust Cascaded Pose Regression (RCPR), uses robust shape-indexed features to detect occlusions and estimate the facial landmark positions. Zhang et al. proposed a CNN for facial landmark detection [111]. They optimized the facial landmark detection to age estimation task, but also on other tasks as facial expression recognition. The model uses a deep neural network to share representation between the tasks to learn the facial landmark points.

The feature extraction component extracts features from the facial image. Those features are important in the age estimation process, while they affect the performance of the estimation results. As there are many algorithms for feature extraction, we will describe several popular ones. Image-based age estimation approaches view the face image as a textured pattern. Several papers [96, 69, 6, 20] compared various feature extraction techniques. Features like LBP [32], HOG [3], local ternary patterns (LTP) [100], local phase quantization (LPQ) [76], and BIF [35] are used. Mandal et al. use optical0based features, which can perform similar to face component based features [66]. But, their tracking is initialized by manually annotated facial landmarks. Pontes et al. proposed a flexible hierarchical approach for facial age estimation based on multiple features [79]. There are two different features extracted: global features and local features. LBP, Gabor wavelets [67] and LPQ descriptors are used to extract Wrinkles and skin spots. These approaches are used after detecting skin areas.

The last component of the facial age estimation task is classification. The goal of the classification algorithm to estimate the age of a given facial image. Age can be seen as classification, regression, or hybrid problem. Age is as a classification problem when the age is estimated into an age group. For the regression problem, age is estimated as a single label. A hybrid problem classifies the age first into an age group and then estimates the age from

that age group. Algorithms such as Support Vector Machine [35], Nearest Neighbor [57], Quadratic function [58], Neural Networks [55], Support Vector Regression [34] have been used in the past as classifiers. Recently deep learning methods are mostly used for age estimation tasks.

The core idea of this research is introduced in [88]. However, a different approach will be used for our age estimation framework. Rothe et al. proposed a deep learning method for age estimation without using facial landmarks. They use the VGG-16 architecture of CNN that is pre-trained on the ImageNet dataset for image classification. The model is fine-tuned with the IMDB-WIKI dataset, which they have introduced.

Chapter 3

Automatic image-based age estimation

This chapter describes the methodology of the implemented approach to estimate the age between Alzheimer's Disease patients and healthy subjects by residual networks. Firstly, give an introduction to convolutional neural networks. Then the datasets used for the experiments will be given. Finally, the implemented age estimation framework will be discussed.

3.1 Convolutional Neural Network Preliminaries

Convolutional Neural Network (CNN) is a deep learning approach inspired by the biology of the animal visual cortex [59]. CNN is commonly used for applications as image processing. The architecture of a CNN is constructed by several stacks of several layers that make a class prediction for a given input image. Those layers are discussed in the next section. A CNN architecture can differ in number and type of layers. There are no criteria for the amount and type of layers used in a CNN model. Although, recently CNN in image processing is divided into two parts, feature extraction, and classification. An example architecture of CNN is shown in Figure 3.1.

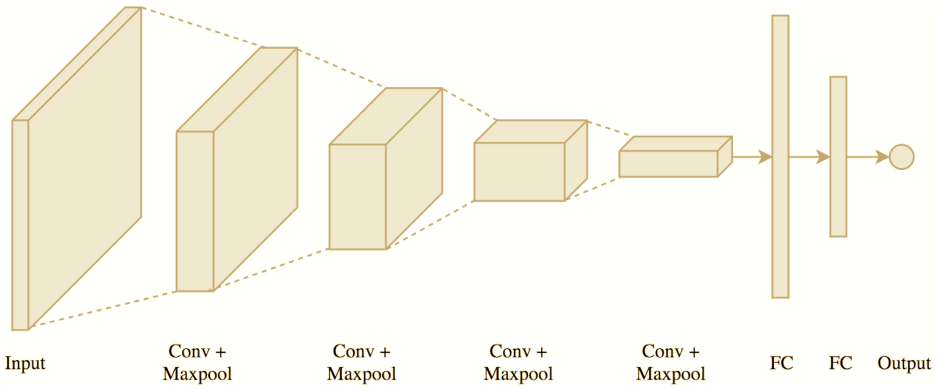


Figure 3.1: An image is passed through the convolutional, pooling, and fully connected layers, that gives an output of probability of the the class that the image belongs.

A CNN consists of several layers such as convolutional, pooling, and fully connected layers. Basically, the input layer of a CNN is an image, where convolutional is performed on the next layer, known as the convolutional layer. An image is represented as a matrix in the form of 3D color channels (red, green, blue). For example, an image with a size of 224×224 contains 50 176 pixels. In CNN, the input shapes of the network are formatted as $H \times W \times D$, where D is the depth of the color channel.

A convolutional layer is followed after the input layer, where a convolutional operation is performed on the input image using a filter, also known as the kernel. The convolutional operation is the sum of a dot product between the input image and the filter. The output of the operation is a feature map which is the input for all other layers. In later stages, the input for the other layers are feature maps. The convolutional operation is defined as:

$$FeatureMap = I_{image} \otimes F_{filter} = \sum_{n} \sum_{m} I[n, m]F[x - n, y - m], \qquad (3.1)$$

where the x,y are coordinates in the matrix and n,m the size of the matrix. The filter has a smaller size compared to the image input, where it slides as window over the input image. Figure 3.2 gives a visual example of the convolutional operation.

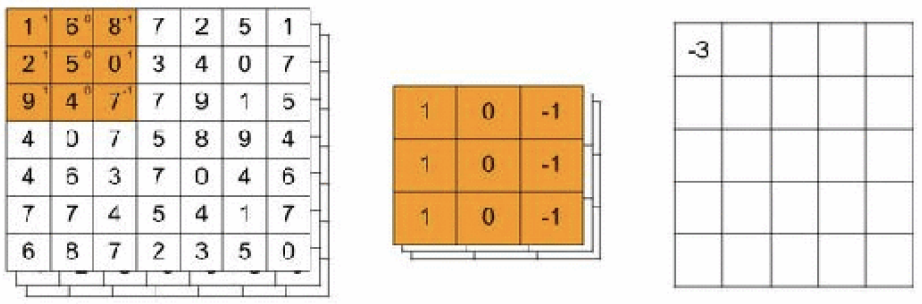


Figure 3.2: A filter of size 3×3 slides over the image (7×7), that results into feature map 5×5 size.

The convolutional layer includes some parameters (padding, stride, and kernel size) that have an impact on the size of the feature map. Padding is used to keep the size of the output the same as the input image. The size of the data will be reduced when the padding is not defined. The stride determines the shifting size of the filter on the input.

After each convolution layer, an activation function is applied to solve the non-linear problem. A non-linear activation function is used to make more complex decisions. The most common activation function used in CNN is a Rectified Linear Unit (ReLU), which is favored for its faster training speed. The function returns 0 if it receives any negative input, but for any positive value x, it returns that value back. The ReLU is defined as:

$$f(x) = max(0, x), \tag{3.2}$$

layers followed by the ReLU will only used when f(x) does not equal to 0.

A pooling layer reduces the spatial size of an image, which is downsampling the feature maps by reducing the height and width dimensions of it. This layer is used after a convolution layer. The most used pooling layers are max and average pooling. The average pooling takes a filter of $N \times N$ size and stride of N length to output the average of the neighbor units of the feature map. The max pooling takes a filter of $N \times N$ size and stride of N length to output the maximum value in every region of the feature map.

$$MaxPooling = max_{x,y=1}^{h,w}, X_{ij}, \tag{3.3}$$

$$AveragePooling = \frac{1}{len(hw)} \sum_{x,y=1}^{h,w}, X_{ij}, \tag{3.4}$$

Using a pooling layer can prevent the CNN from overfitting and improve the computational efficieny. However, it can lead to underfitting when use of large amount of pooling layers.

Fully Connected Layer is initialized after the last pooling layer that receives the output. As the Fully Connected Layer expects a 1D arrays as input, the pooled feature map is converted into a long continuous linear vector. This process is called Flattening. The Fully Connected layers perform classification based on the features extracted by the previous layers. This layer contains softmax activation function is defined as,

$$p(y_i) = \frac{e^{x_i}}{\sum_j^k e^{x_j}}, \tag{3.5}$$

returns a probability between 0-1 for each label. A softmax activation function is a form of logistic regression that normalizes an input value into a vector of values.

A loss function is used to evaluate the CNN model. The loss function estimates the error of the model. A cross entropy is used for a multi class network.

$$Cross - Entropy = -\sum_i^C t_i log(f(s)_i)), \tag{3.6}$$

Another layer that is often used in a CNN model is a dropout layer. The Dropout layer is useful to avoid the overfitting of the data. The basic idea is that random amount of neurons will be dropped during the training.

A batch normalization layer is a method that lets the CNN train in an stable way [46]. The output of each layer will be normalized by shifting the inputs to zero-mean and unit variance. The next sections will discuss the intuition behind the residual networks and transfer learning.

3.1.1 Residual Networks

ResNet is a residual network developed by Kaiming He [39]. This network uses residual blocks to solve one of the problems in CNN, training more complex and deeper networks. Basically, the idea is that the activation of the previous layer will be the same as residual connections and learn new representation with the new layer. For example, the network will try to

learn the residual $F(x) = H(x) - x$ by $F(x) + x$ instead of learning the $H(x)$ directly as shown in figure 3.3. Simply, the residual block adds a skip connection between the convolutional layers. Compared to the VGG-16, dimensionality reduction in the residual blocks leads to less number of parameters. This means that the complexity of the model will be lower.

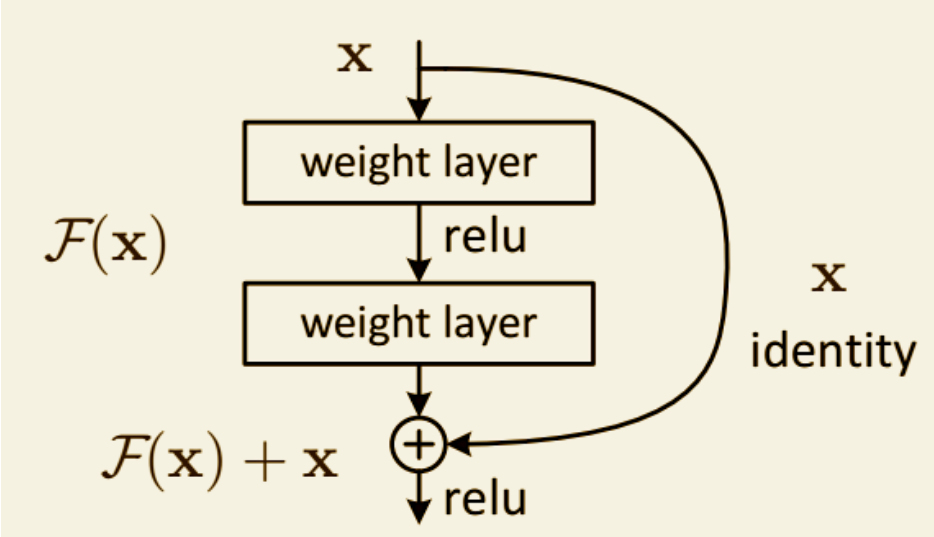


Figure 3.3: A residual block.

The ResNet-50 architecture starts with a $7x7$ convolutional layer and a max-pooling layer of $3x3$, both layers have stride 2. Every block of ResNet-50 is composed with 3 convolutional layers $[1x1, 3x3, 1x1]$. After the max-pooling layer, there are the following residual blocks: $3x$ with depth dimension of 256 ,4 with depth dimension of 512, 6 with depth dimension of 1024, and 3 with depth dimension of 2048. The architecture ends with a $7x7$ average pooling and a softmax classifier.

3.1.2 Transfer Learning

Transfer learning is an approach for CNN, where the knowledge of a pre-trained model on a larger database is transferred to a smaller database. The layers of the pre-trained model are frozen, with only modifying the last few layers to make predictions. Transfer learning is a very popular approach nowadays, while it saves computation time to train larger deep networks.

The pipeline for transfer learning is loading a pre-trained model that is trained on different datasets, freeze some convolutional layers, remove the

output layer, and apply it to the specific problem. For age estimation, a model needs to be fine-tuned on facial images to obtain accurate predictions instead of using a model that is trained on multiple problems as vehicle detection.

3.2 Framework

This section discusses the implementation of the automatic image-based age estimation framework. An overview is given in figure 3.4. The architecture of our framework is very similar to the general image-based age estimation approaches [88, 61]. The framework starts with a general pre-processing step for facial images. The facial images are prepared by face detection, cropping, alignment and resizing, explained in section 3.4. Then, a residual neural network has been build for feature extraction and classification, explained in section 3.5.

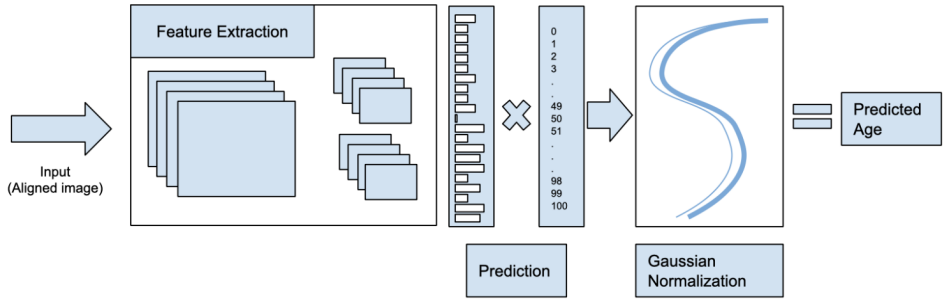


Figure 3.4: An overview of the residual age estimation architecture.

This framework is used for estimating the age of both AD patients and healthy subjects. The objective is to implement a CNN for age estimation to examine the performance of estimation on databases with faces above 50+ age. The main objective is to have an accurate CNN model for age estimation to examine the performance of estimation on databases with faces above 50+ age. Naturally, the training is not restricted to elderly faces, but we are more interested in the performance on faces of 50+ ages. The approach is based on the implementation of Deep EXpectation (DEX) of apparent age [88]. The DEX architecture predicts the apparent age by first detecting the face on the image and then by applying a deep convolutional neural network

with the VGG-16 structure for the prediction. Instead of using VGG-16 architecture, we use a ResNet-50 architecture and fine-tune the pre-trained ImageNet model. The age estimation is treated as a classification problem, where we have 101 classes, regarding the ages from 0 to 100. Although, a regression approach is better for continuous age estimation. We used a post-processing stage to accumulate results from the 100-dimensional classification output, and there perform the switch from discrete to continuous by using a softmax classifier. Each softmax output will be multiplied with their age class and then summed for the prediction. This will give us several predictions for each subject's age, given a neutral image selected automatically from the video as input. Given the predictions, we fit a curve to the age distribution to obtain a single prediction as to the maximum of the curve. The next section discusses the databases used during the experiments.

3.3 Data

The facial age estimation consists of various collected databases with different conditions. The databases are based on their real age or apparent age, and both as well. As many databases are accessible online, they differ in image conditions as other aspects. To train the model with accurate results to our testing database, we selected the databases based on several conditions. One of the most important aspects is that there is enough data for training the model for accurate results to make an accurate estimation in the elderly.

Selection criteria of databases:

- database image size, it is important to have large amount of training data for the performance.
- Large amount of elderly images: the most images contain less faces of elderly. As our test database contains images from persons above 50 years old. It is important to have also a minimum amount of elderly images in the training set.
- Image conditions: from the selected databases inappropriate images are removed based on the conditions at section 2.3.1.

The following subsections will describe the databases that are used for training purposes. We also used databases for testing purposes to achieve our goal.

3.3.1 ImageNet

ImageNet is a large database that contains images for many categories such as human faces, animals, objects, and more [90]. This database is mainly used for object detection purposes. Although, we used this database from a pre-trained CNN model for more accurate age classification. The database contains around 14 million images that have been hand-annotated to indicate the image regions. ImageNet is used in Large Scale Visual Recognition Challenge (ILSVRC), where software programs compete to correctly classify and detect objects and scenes.

3.3.2 UTKFace

The UTKFace data consists of images of human facial. This database has more than 20.000 facial images from an age range from 0 to 116 years old [109]. The images are annotated with age, gender, and ethnicity. Additionally to full facial images, the database also contains aligned and cropped face images. The purpose of using this database is to improve the accuracy of the age estimation model. Sample images of the UTKFace are given in figure 3.5.

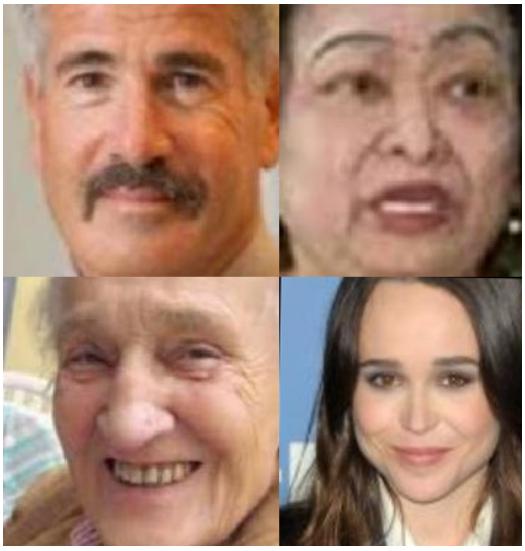


Figure 3.5: Samples from UTKFace database.

The UTKFace database is annotated as:

- Age: An integer that ranges from 0 to 116 indicating the age of the person in the image.

- Gender: An integer that is a 0 (male) or 1 (female), indicating the gender of the person.
- Race: An integer that ranges from 0 to 4, indicating the race of the person. expressed as White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern).
- Datetime: provides the date and time that an image was collected in the following format yyyymmddHHMMSSFFF.

As our goal is to estimate age, we are only interested in the age labels. The images are divided into folders corresponding to their age. Those folders are used as input for the data-generator of Keras, explained in section 3.4.

An example UTKFace folder structure:

- 0: Contains images with age 0.
- 1: Contains images with age 1.
- ...: ...
- 99: Contains images with age 99.

3.3.3 Appa-real

APPA-REAL is a database that is provided by ChaLearn LAP [23]. This database contains around 7,591 facial images and is annotated with the real and apparent ages. The facial images are already aligned and cropped with a margin of 40% that is obtained from a face detector [68]. There are many inappropriate cropped images as a result of a failure in face detection.

The annotations of the real age and apparent age are provided by CSV files. Those facial images are already separated into train, test, and validation sets. Each file contains the filename, real age, apparent age, worker age, worker gender. The last two labels are indications of the annotators, which annotated the apparent age of the person in the image. Sample images of the APPA-REAL data are given in figure 3.6.

Figure 3.6: Samples from APPA-REAL database.

The purpose of using the UTKFace and APPA-REAL databases are for fine-tuning the pre-trained ImageNet model.

3.3.4 FG-NET

The FG-NET database contains 1,002 images from age range 0-69 [28]. Since this database is small, we only use it for evaluating the age group 60-69 for estimating older ages.

3.3.5 Alzheimer's Disease database

The Alzheimer's Disease database is collected by CAPA Hospital in Istanbul, Turkey. It includes 104 videos from patients and healthy subjects. Each video has a different length and contains variations due to pose, illumination, and scale. The database is composed in RGB videos and is recorded in 1920 x 1080 pixels at a rate of 25 bits per second. It contains 104 videos of smile expressions from different subjects. The Alzheimer's database is used for testing our main hypothesis.

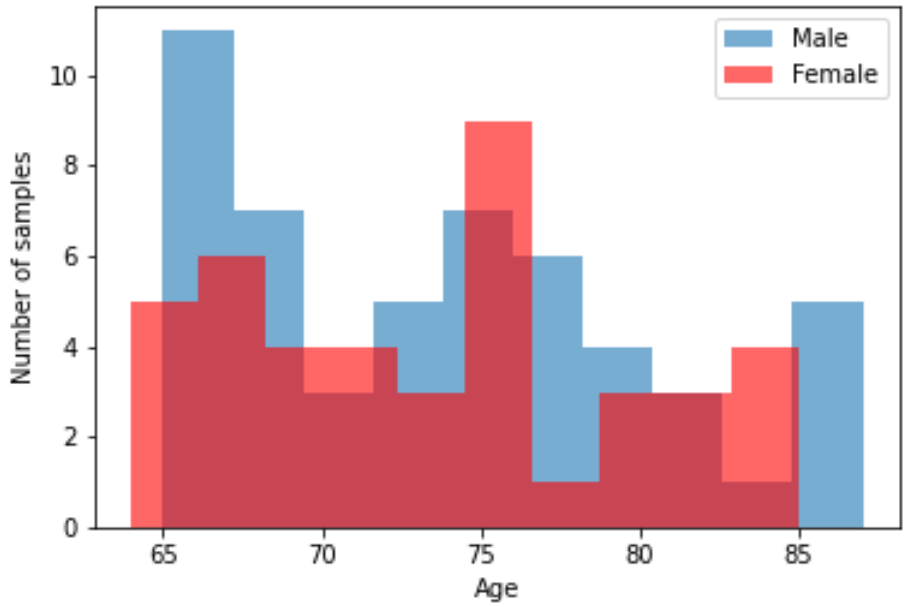


Figure 3.7: Number of samples of the AD database per age grouped by gender.

3.4 Data preparation

The databases contain facial images with different imaging conditions such as face poses, illumination, and size. We apply face detection, alignment, and cropping to obtain only the face from the image.

The first step is to detect face from the image. As only facial information is important, we discarded the background information. The Haar Cascade based face detector (Viola-Jones) is used to detect faces from the image (see section 2.3.1.1) [101]. A more elaborated face detection approach is not needed, while our databases are frontal and clearly illuminated, with good resolution. We also experimented with a more recent, neural network based face detector, called Single-Shot-Multibox Detector (SSMD) [63]. The SSMD face detector is using a CNN that uses multi-scale feature maps to detect faces. The main difference between the Viola- Jones face detector and SSMD is that the SSMD includes more of the forehead of the face, which can be important in age estimation for estimating the age based on the forehead wrinkles.

After detecting the face, figure 3.9, a rectangular area is cropped from the image. For face alignment, we used a facial landmark approach to align the face [51]. An annotated training set of images is used for training an ensemble of regression trees in this approach. As a result, the locations of 68 facial landmarks are estimated from their pixel intensities. The dlib library [1] has a pre-trained facial landmark detector to detect these 68 landmark locations. Given the facial landmarks, the faces can be aligned such that the faces are rotated where the eyes lie horizontally at the same y-axis, as in Figure 3.8. The angle is determined by Angle determined by D_x and D_y (differences between left and right eye center). As the images are aligned, figure 3.10, the faces are cropped to a size of 224 224 pixels. A few images were missed by the face detectors due to the image quality. As those images could decrease the performance for age estimation on the elderly, we have removed them from the training set.

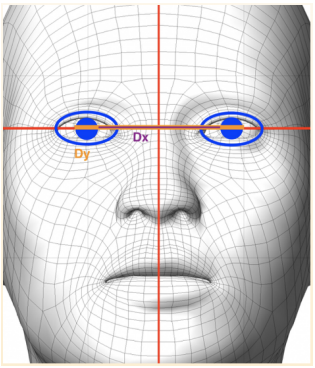


Figure 3.8: Eye alignment.

Figure 3.9: Example image.

Figure 3.10: Preprocessed image.

3.5 Network

The feature extraction and classification of the automatic age estimation is based on a residual network [39]. The apparent age is estimated with a residual convolutional neural network. A ResNet-50 architecture, which is a 50 layer Residual Network and includes a pre-trained ImageNet model [90]. The pre-trained network is first fine-tuned with the APPA-REAL database and then with the UTKFace database. The input of the residual network uses images with 224224 pixels. As the pre-trained ImageNet model con-

[1] http://dlib.net/

tains an output layer of 1000 classes. We removed this output layer and replaced it with 100 classes, to represent the age range from 0-100, and retrain with facial age data. We compute the softmax expected value on the output probabilities of those classes [88]. During fine-tuning, the layers of the pre-trained model are frozen and the last fully connected layer is removed from the model. Instead, several new layers are added to create a new classification or regression model. The features of the pre-trained model are thus transferred to the new model.

The CAPA Alzheimer's Disease database is used for testing the models for estimation accuracy. As this database consists of video data, we automatically select a neutral frame from the video of the subject. Then extract the frames with an open-source facial behavior analysis toolkit, namely OpenFace [7]. This toolkit detects the locations of lip corners. The lip corners are used for calculating the temporal phases of facial expressions as in [21]. The smile amplitude is calculated as the mean amplitude of right and left lip corners, normalized by the length of the lip. The Onset is the initial phase of facial expression, and it defines the duration from neutral to expressive state. The apex phase is the phase between the onset and the offset, which is a stable peak period of facial expression. Likewise, the offset is the final phase from expressive to neutral state [48]. We automatically selected one neutral frame from the onset phase.

Chapter 4

Experiment

This chapter describes the experiments for the implemented automatic age estimation. As the performance of the classifications is mostly determined by pre-processing and training parameters. We conducted several experiments on how those factors could affect the performance of the age estimation. Firstly, we will provide the environmental settings and evaluated metrics for the age estimation task. Then, we demonstrate the training parameters and experimental results of the age estimation tasks.

4.1 Experimental Setups

This section describes the experimental settings that the framework is evaluated on. Firstly, we discuss the metrics where age estimation is evaluated. Then, we give an overview of the technical settings used in this approach.

4.1.1 Evaluation Metrics

The common performance measures for age estimation are the Mean Absolute Error (MAE). The MAE is the average of absolute error between the real age and the estimated age.

$$MAE = \frac{1}{n}\sum_{n=1}^{n} \|x_i - x\| \tag{4.1}$$

Also, the Mean Error (ME) is measured to estimate the mean error between the real age and the estimated age.

$$ME = \frac{1}{n}\sum_{n=1}^{n} x_i - x \tag{4.2}$$

4.1.2 Environment

The following software and hardware specifications are used for realizing the implementation. The code is implemented on an Apple Macbook Pro with a 2,6 GHz Intel Core i7-8850H, 16 GB 2400 MHz DDR4, and Intel UHD Graphics 630 1536 MB. The operating system was macOS Mojave 10.14.3. We used the Keras API [31], which is a high-level neural networks API, programmed in Python. A ResNet50 model is a 50 layer Residual Network, which includes a pre-trained ImageNet model available from Keras API.

The age estimation framework is realized and implemented with Python. As the hardware specifications affect the training of a deep learning model. We used a free cloud service of Google to train the network. Google Colaboratory free to use and provides free GPU. With this service, we can implement deep learning applications using libraries as Keras, TensorFlow and OpenCV. MacBook Pro Retina is used for additional experiments such as estimating age from facial images and plotting heatmaps, which is a graphical representation of data depicted with color. The heatmap gives an idea of where the model pays attention when it comes to a prediction. The specifications are described below.

Model	MacBook Pro (Retina, 15-inch, 2018)
CPU	2,6 GHz 6-Core Intel Core i7
Memory	16 GB 2400 MHz DDR4

Table 4.1: Specifications of the MacBook Pro used for the experiments.

The specifications of Google Colaboratory:

- GPU: 1xTesla K80, compute 3.7, having 2496 CUDA cores, 12GB GDDR5 VRAM
- CPU: 1xsingle core hyper threaded Xeon Processors @2.3Ghz i.e(1 core, 2 threads)
- RAM: 12.6 GB Available
- Disk: 320 GB Available

We used Matlab r2019a and SPSS Statistics to analyze the MAE of Alzheimer's dataset. Also, we used Matlab r2019a for the post-processing

stage, where we calculate the Gaussian distribution of 100-dimensional classification output from the single image. Then we use curve-fitting to model the distribution and select the maximum of the curve as the prediction.

4.2 Automatic Age Estimation Model

We conducted several experiments to obtain an accurate age estimation model and improve the age estimation performance for the elderly. A deep learning model has several configurable parameters that determine the model structure and how the model is trained. Those parameters are set before the training of the model. During the experiments, we trained the models with various parameters. The training results are described in section 4.2.1.

The following parameters are needed for the model:

- The Learning rate defines how quickly the parameters are updated by the model.
- Epoch is the number of times that the entire training data is used once to update the weights in the model.
- Optimizer is used for updating the weights and minimizing the loss function. The loss function tells the optimizer if it moves to the right or wrong direction.
- Momentum, the movement of the "optimizer's" direction depends on the size and direction of the movements in the previous update.
- Dropout is a regularization technique to reduce overfitting.
- L2 regularization is a weight decay algorithm that stables the model.
- Batch size indicates the number of sub samples that the model gets updated.

4.2.1 Training results

We trained several models to find the optimal age estimation model for estimating the elderly. First, we experimented with two different optimizers, Stochastic gradient descent (SGD) [10] and Adam [53]. Then, for the best optimizer, we fine-tuned the model based on the APPA-REAL and UTKFace

datasets. Finally, we found our optimal model by fine-tuning the network twice with two different datasets instead of one time fine-tuning of a combined dataset. Estimation results between those models can be found in figure 4.6. The predictions were made on some samples of the IMDB-WIKI dataset.

Experiment with different optimizers

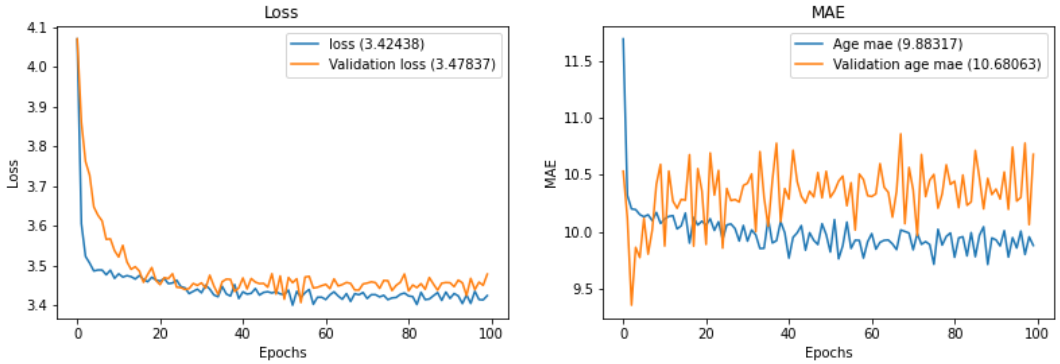
We conducted experiments based on two different optimization techniques. Figure 4.1 shows the loss and MAE for the model trained on the SGD optimizer. The loss and MAE for the model trained with Adam optimizer are shown in figure 4.2. The models are trained on combined APPA-REAL and UTKFace datasets and we only took images where the age is greater than 50 years. The results show that MAE and the validation MAE for the Adam model are around 9.66 and 9.69, respectively. While the MAE and the validation MAE for the SGD model is around 9.88 and 10.68. As the model is trained on images above 50+ age, the model predicts the age for younger persons above the 50 years old. This shows that the amount of data and annotations of it are important and very effective on the performance of the estimations. From the results in figure 4.6, we can conduct that the estimations with Adam optimizer are more accurate compared to the SGD optimizer. The parameters used at the training of the optimizer experiment are described in tables 4.2 and 4.3.

Learning rate
0.001
Optimizer
SGD
Epochs
100
Momentum
0.9
Training set split
90%
Validation set split
10%
Batch size
64

Table 4.2: Parameters of the SGD model.

Learning rate
0.001
Optimizer
ADAM
Epochs
100
Training set split
90%
Validation set split
10%
Batch size
64

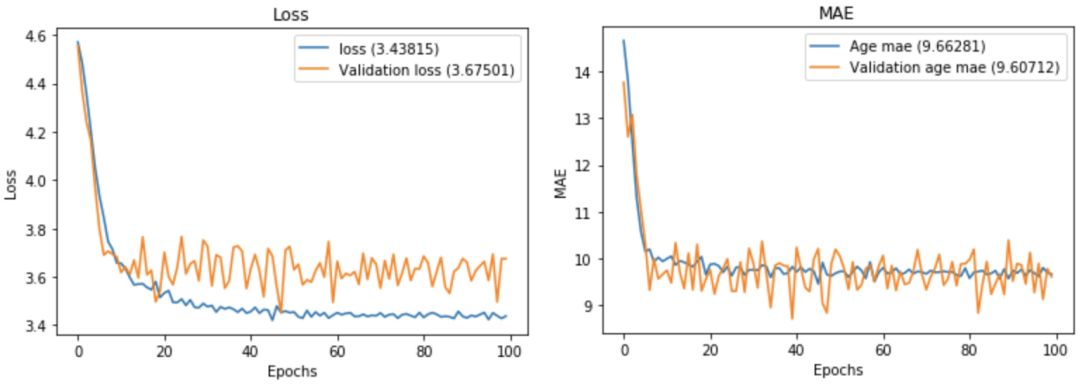
Table 4.3: Parameters of the ADAM model.



(a) The loss of the model.

(b) The MAE of the model.

Figure 4.1: The loss and MAE of fine-tuned model with SGD.



(a) The loss of the model. (b) The MAE of the model.

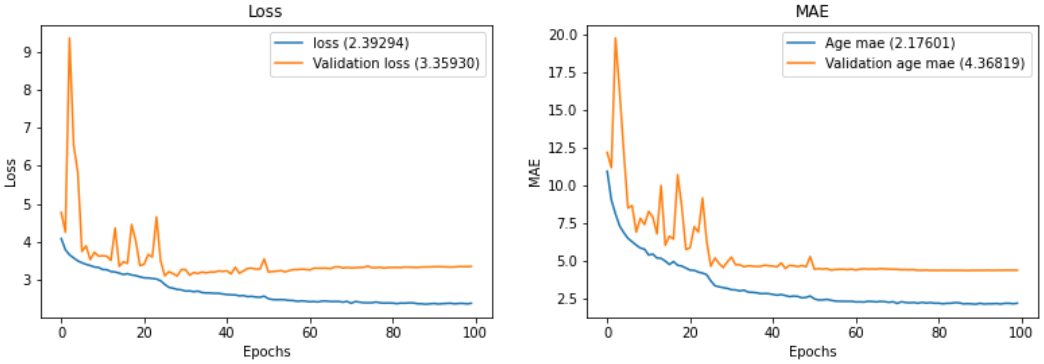
Figure 4.2: The loss and MAE of fine-tuned model with Adam.

Experiment with apparent and real age

After selecting the optimizer, we conducted experiments on the APPA-REAL dataset to see the differences in using the real age or the apparent ages. Those models were trained with Adam optimizer. The APPA-REAL and UTKFace datasets are combined and trained with an age range from 0 to 100. The validation loss and mean absolute error are 3.344 and 4.350, respectively. As shown in Figure 4.3, the model is underfitting and there will be greater errors for predictions. However, as shown in Figure 4.4, the validation loss and mean absolute error are 3.748 and 6.574, respectively. This model is also underfitting. The MAE of the model trained on real ages is greater compared to the apparent model.

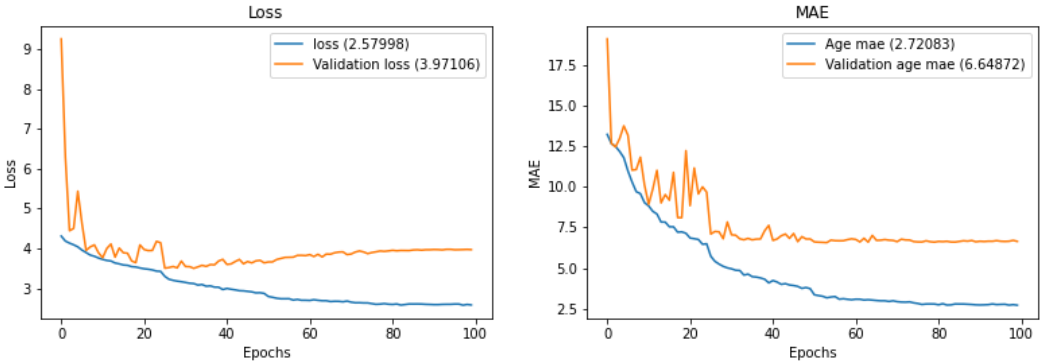
Learning rate	0.001
Optimizer	ADAM
Epochs	100
Training set	90%
Validation set	10%
Batch size	64

Table 4.4: Parameters of apparent and real age models.



(a) The loss of the model. (b) The MAE of the model.

Figure 4.3: The loss and MAE of fine-tuned ResNet-50 on APPA-REAL Apparent Age and UTKFace.



(a) The loss of the model. (b) The MAE of the model.

Figure 4.4: The loss and MAE of fine-tuned ResNet-50 on APPA-REAL RealAge and UTKFace.

The Optimal Model

The model is trained with the ADAM optimizer with a learning rate $\alpha = 0.001$ and a batch size of 64 [53]. Then, we fine-tuned the pre-trained on the APPA-REAL database for 100 epochs. Followed by a fine-tune on the UTKFace database that is trained for a further 35 epochs, shorter as we applied the early stopping callback [80].

An L2 regularization with $\lambda = 0.0005$ weight on the loss function is applied for the UTKFace fine-tuned model. We also applied a dropout (0.3) for both models and fine-tuned both models with two losses: the cross-entropy

loss and the Euclidean loss. Freezing the weights of a pre-trained model prevents the update of a specific weight during backpropagation. We froze the first 26 layers. The databases are each split into two folds, as 90% training set and 10% validation set, respectively. As we want to predict the age of Alzheimer's Disease and healthy subjects, we have tested the model on the CAPA Alzheimer's Disease database, which was not used during training and parameter selection.

Comparing both figures with the final model, in Figure 4.5, fine-tuning the model twice with two different databases leads to more efficient predictions. In this figure, we show the training curve of the second fine-tuning of the residual network on the UTKFace database. The learning rate was set to 0.001 with an L2 regularization of $\lambda = 0.0005$ and a dropout is set to 0.3. The validation loss is higher than of the training loss, which suggests that the model is overfitting a little bit compared to the other models. Using L2 regularization and fine-tuning twice a model improves estimations for old people.

As shown from the results in figure 4.6, we can conduct that the estimations with the optimal model are average. However, the face detector had an impact on the results of those estimations. Especially for persons with a larger forehead, the face detector only detected the face below the eyebrows.

Learning rate	$\alpha = 0.001$
Optimizer	ADAM
Epochs (fine-tuning APPA-REAL)	100
Epochs (fine-tuning UTKFace)	35
Dropout	0.3
L2 regularization	$\lambda = 0.0005$
Callbacks	Early stopping
Training set	90%
Validation set	10%
Batch size	64

Table 4.5: Parameters of the optimal model.

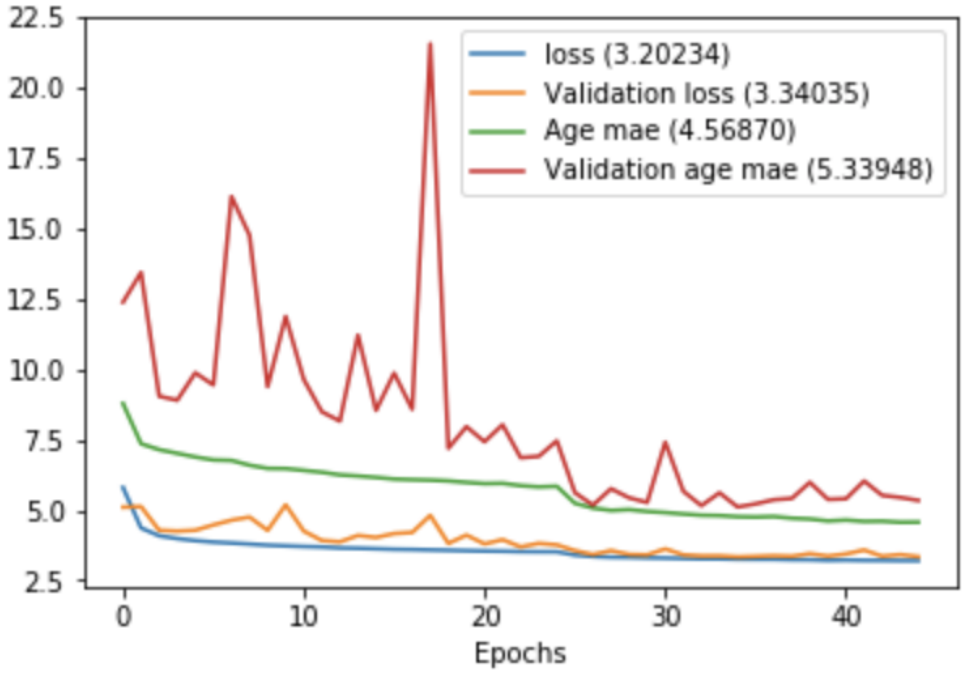
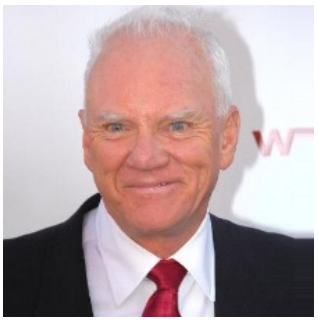


Figure 4.5: The loss and MAE of the optimal model.

Real Age: 28
SGD model: 63
ADAM model: 57
RA model: 30
AP model: 29
Ours: 36

Real Age: 22
SGD model: 58
ADAM model: 57
RA model: 25
AP model: 26
Ours: 26

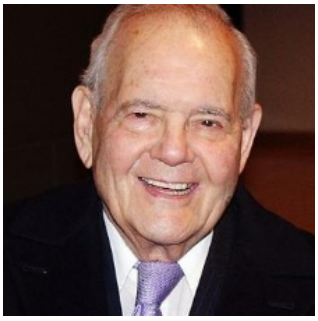


Real Age: 64
SGD model: 59
ADAM model: 65
RA model: 51
AP model: 53
Ours: 53.

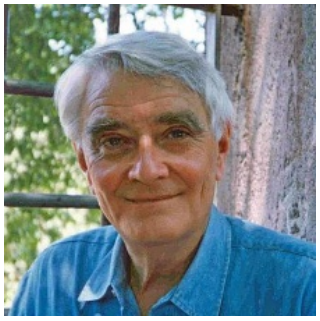
Real Age: 38
SGD model: 57
ADAM model: 57
RA model: 36
AP model: 34
Ours: 33

Real Age: 57
SGD model: 59
ADAM model: 55
RA model: 52
AP model: 40
Ours: 49

Real Age: 79
SGD model: 66
ADAM model: 64
RA model: 54
AP model: 52
Ours: 58



Real Age: 83
SGD model: 63
ADAM model: 67
RA model: 68
AP model: 63
Ours: 71



Real Age: 66
SGD model: 67
ADAM model: 60
RA model: 59
AP model: 51
Ours: 62

Real Age: 35
SGD model: 57
ADAM model: 56
RA model: 35
AP model: 33
Ours: 32

Figure 4.6: Age estimation samples on the IMDB-WIKI dataset predicted on the trained models. SGD = model trained on SGD optimizer; ADAM = model trained on Adam optimizer; RA = fine-tuned model on real age; AP = fine-tuned model on apparent age; Ours = Our optimal model.

4.2.2 Age estimation results of FG-NET

Most classification methods use the FG-NET database for classifying age into age groups. To illustrate the performance of our approach in a comparable way, we report here results with FG-NET as well. We conducted experiments for previous studies that used the FG-NET database for classifying age. As shown in the Table 4.6, we perform better than most of the age estimation methods on ages above 60+. The AAM and LBP methods are evaluated on three samples, while the other five samples are used for training. A fair comparison between those methods cannot be made. We achieve MAE of 8.14 for ages above 60+, but clearly, FG-NET is not adequate for this task, as there are only eight samples with ages over 60. The fine-tuning on the elderly requires a two-level classification approach, where different classifiers are used for the younger and older subjects. Subsequently, we do not report additional results on the younger subjects.

Method	**# of 60-69 Samples**	**MAE**
IIS-LLD (Single) [30]	8 images	32.13
IIS-LLD (Triangle)[30]	8 images	26.25
Enhanced BIF [25]	8 images	26.25
IIS-LLD (Gaussian)[30]	8 images	24.00
C-lsLPP [12])	8 images	23.37
C-lsMFA [12]	8 images	22.25
PCA + LPP + SS [13]	8 images	17.33
OURS	8 images	8.14
SR-AAM [65]	Not mentioned	5.32
AAM [22]	3 images	4.62
LBP [22]	3 images	**4.53**

Table 4.6: Comparison of age estimation methods on the FG-NET database with ages greater than 60 years old.

4.2.3 Age estimation results between healthy subjects and AD patients

The following results are estimations between healthy subjects and AD patients. Firstly, we analyze the effect of Alzheimer's Disease. Then, we examine the effect of gender between healthy subjects and AD patients.

4.2.3.1 Effect of Alzheimer's Disease

The MAEs and mean errors (ME) for Alzheimer's Disease patients and healthy subjects are given in Table 4.7. The database has 50 AD patients and 44 healthy subjects. As shown in the table, the MAEs of AD patients is higher compared to healthy subjects. Automatic algorithms almost always underestimate the age of elderly subjects, and our approach is no exception to this general trend (as shown by negative ME results). The results show that on average, AD patients are estimated to be 5.22 years younger than healthy subjects in the same age range. Using t-test, the difference is found to be statistically significant ($p = 0.0071$). The effect of the gender on the Alzheimer's Disease has been assessed to understand that the AD patients seem younger.

Groups	**# of samples**	**MAE**	**ME**
AD Patients	50	10.9	-9.7
Healthy subjects	44	8.16	-4.48

Table 4.7: Comparison for the age estimation on the CAPA Alzheimer's Disease database.

4.2.3.2 Effect of gender

The effect of gender on the results has been assessed. The MAEs and MEs for male and female subjects are given in Table 4.8. The MAEs for female and male estimates differ slightly. The automatic method estimates the age of males younger than that of females both for AD patients (3.23 years younger) and healthy subjects (5.38 years younger). However, these differences are not significant ($p \geq 0.05$). Also, notice that the dataset is not large enough to derive any conclusions from these results. It is interesting to note that the difference in average error between healthy and AD male subjects is quite large (3.67). If this is a systematic bias that would manifest itself in larger sample sizes, it may suggest that age indicator is more reliable for diagnostic purposes in males, compared to females.

Groups	**Gender**	**# of samples**	**MAE**	**ME**
AD Patients	Male	27	11.63	−11.19
	Female	23	10.04	−7.96
Healthy subjects	Male	24	7.96	−6.80
	Female	20	8.40	−1.42

Table 4.8: Comparison of the genders for the age estimation on the CAPA Alzheimer's Disease database.

We investigated the facial regions and their contributions to the age estimation task. This, of course, is a computer-based analysis, and it may be the case that humans pay attention to different cues. Figure 4.7 shows a heatmap depicting the relative importance of facial areas for 4 subjects the residual image-based age estimation model. We obtained the results by systematically blurring patches on the faces and observing the differences in estimated ages. A larger difference indicates to a more important area. Those depictions show that the model mostly relies on eye corners, top of the mouth, and the cheeks. Also, the vertical wrinkles on the cheeks on either side of the mouth are the most important marks of age, and blurring them decreases the apparent age of the subject.

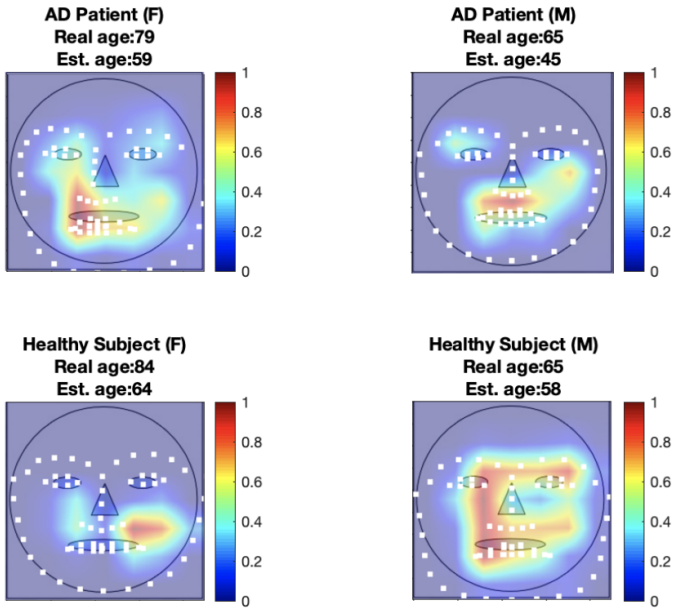


Figure 4.7: Heatmap for four subjects from the AD database. The facial landmarks indicate the facial shape and location of the real subjects.

Figure 4.8 provides some examples of the age estimations. The estimations are well predicted on older faces compared to young ones. The error for the image at right is +6 years, while the image in the middle is predicted with an error of -3. As the age estimation works properly for old faces, it has some constraints for correct estimation at younger faces. Because the model is trained moreover on older faces.

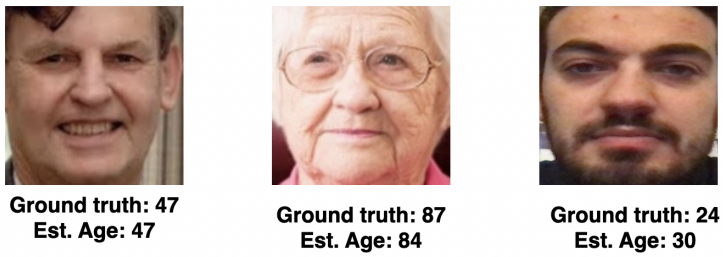


Figure 4.8: Age estimations for three subjects.

Chapter 5

Discussion

Estimating the age of AD patients and healthy subjects show that the proposed age estimation approach indeed estimates AD patients to be younger than their real age. The effect of gender is particularly stronger at male subjects. Although, the male healthy subjects are estimated younger as well. This finding is a small step towards supporting our hypothesis and encourages us to investigate this issue further. In general, the female subjects contain more wrinkles on their faces compared to the male subjects.

The automatic age estimation uses a deep residual CNN architecture. We decided to fine-tune the model several times to obtain most accurate model for old faces. The model seeks to improve the performance on elderly subjects by fine-tuning the network with two databases, APPA-REAL and UTKFace.

We utilize OpenFace [7] for detecting the locations of lip corners. With those locations we calculated the mean amplitude of right and leftlip corners [21]. The mean amplitude is used to find the onset, apex and offset phase. Then we automatically selected one neutral from the onset phase. The pretrained network is accurate to detect the corner locations with minimum error. Although, OpenFace had sometimes issues with face alignment for some faces. So, we used the face alignment techniques described in Chapter 3.4.

A representative face with a heat map is shown in Figure 5.1. The heat map depicts the importance of the facial area for our residual image based age estimation model. It is obtained by systematically blurring patches on the faces to observe the differences in estimated ages. A larger difference illustrates the more important area. We can see that the model mostly relies on eye corners, top of the mouth, and the cheeks. So, blurring the regions

decreases the apparent age of the subject.

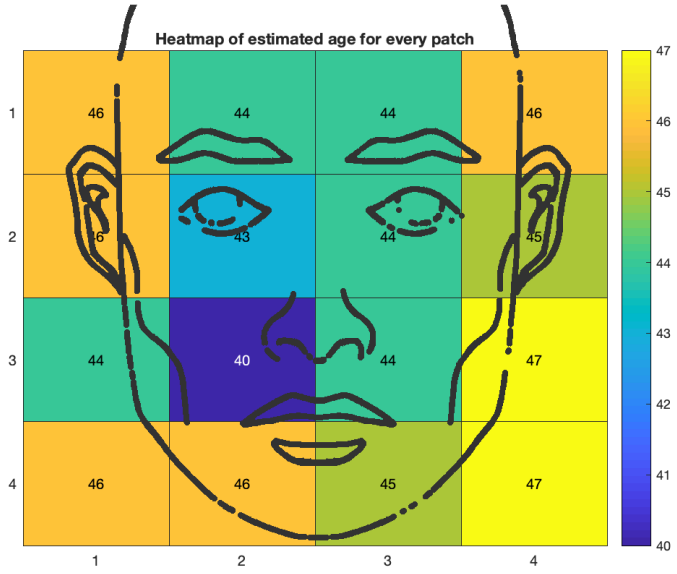


Figure 5.1: Important regions of the face for age estimation. The colors and values indicate the estimated age.

As there are a lot of various facial databases based on specific features. We decided to take images with frontal pose and specific qualities. Most databases had less elderly images which influenced our case of study for estimations. Our goal was to analyze the estimations of Alzheimer's patients, where Alzheimer's Disease occurs mostly above the age of 50 years old. By finding databases that contain similar image conditions with a lot of elderly images was one of the limitations. Most databases contain facial images of persons between 15-60 years old. Even a combination of several databases leads to poor estimations of age estimation. Creating an own collection of elderly databases could have improved this study. Due to the limited databases, we could not make more comparisons of age estimation for Elderly. Most approaches focus on young people, while we have aimed to estimate the age accurately for the elderly. We were not able to make comparison to other approaches that made estimations for those subjects.

Chapter 6

Conclusion and Future Work

This chapter reviews the results of this Master Thesis and presents the final analysis of the experiments. The ideas for future work on this Master Thesis are introduced in section 6.2.

6.1 Conclusion

In this Master Thesis, we introduced an automatic age estimation to analyze the age of Alzheimer's patients and healthy subjects. The approach is based on residual networks to estimate the age of Alzheimer's Disease and healthy subjects. As the Alzheimer's Disease occurs moslty at elderly, it is a challenging task to find a suitable databases that contains more older persons than currently exists. We have dealt with this problem by fine-tuning the model multiple times. However, due to this we were not able to make comparison to other approaches to compare the results on Elderly. Most of the approaches focuses on young.

The first experiments in this work show that age estimation approaches perform better MAE when estimating the apparent age rather than real age. The reason for this could be that the real age can not be always related to the appearance. Also, the experiments show that the parameters can highly affect the results of the estimations. Expanding the training database with more images of various qualities would affect the performance of the estimations for inappropriate images.

The experiment with estimating age of AD patients and healthy subjects show that the proposed age estimation approach indeed estimates AD patients to be younger than their real age, and in particular, this effect is

stronger in male subjects. This finding is a small step towards supporting our hypothesis, and encourages us to investigate this issue further. The residual age estimation is a standard ResNet architecture, but seeks to improve performance on elderly subjects by fine-tuning the network with two databases.

We were able to compare the residual age estimation to other approaches for classifying age above 60+. The model was able to improve other approaches within 8 images. The results were improved from 17.33 to 8.14 MAE.

A contribution of this thesis is a published paper to IPTA Conference "Do Alzheimer's Patients Appear Younger than Their Age? A Study with Automatic Facial Age Estimation"s. The published paper is part of this thesis to share the results of our facial age analysis of Alzheimer's patient. This paper contributes to show that the future investigations could be made between age estimation with diseases in broad context.

6.2 Future Work

The accuracy of age estimation can be improved by using attention mechanism. The attention mechanism can focus on specific regions, which can differentiate young and old faces during the training.

Using Keras is not really recommended due to the limited changes that can be applied to the intermediate layers. A trained network is difficult to modify, where the network could get broken. However, libraries such as PyTorch and Lua more advanced and ca properly deal with those modification problems to the intermediate layers.

To analyze the facial dynamic it is not recommended to use the Keras Library, while it is limited in such cases. However, libraries such as PyTorch and Lua are properly more advanced for implementing the facial dynamics as it can deal with this kind of problems.

We showed that age estimation is highly affected by pre-processing. Using various local features can represent texture information such as wrinkles. Those kind of features could improve the age estimation accuracy as well. Also, using dynamic features and motion features combined with facial static features can result into more accurate model. The dynamic features contain

temporary information of the facial appearance. The motion features contains complementary feature information from the frame differences.

The MAE will be significantly improved for elderly when the facial images in the dataset increases. Although, the number of images for elderly training the classifier are still insufficient. Therefore, collecting a own database should be sufficient.

Bibliography

[1] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12):2037–2041, 2006.

[2] A Midori Albert, Karl Ricanek Jr, and Eric Patterson. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic science international*, 172(1):1–9, 2007.

[3] Fares Alnajar, Caifeng Shan, Theo Gevers, and Jan-Mark Geusebroek. Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions. *Image and Vision Computing*, 30(12):946–953, 2012.

[4] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[5] Kjeld Andersen, Lenore J Launer, Michael E Dewey, L Letenneur, A Ott, JRM Copeland, J-F Dartigues, P Kragh-Sorensen, M Baldereschi, C Brayne, et al. Gender differences in the incidence of ad and vascular dementia: The eurodem studies. *Neurology*, 53(9):1992–1992, 1999.

[6] Raphael Angulu, Jules R Tapamo, and Aderemi O Adewumi. Age estimation via face images: a survey. *EURASIP Journal on Image and Video Processing*, 2018(1):42, 2018.

[7] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 59–66. IEEE, 2018.

[8] M Bartlett and Jacob Whitehill. Automated facial expression measurement: Recent applications to basic research in human behavior, learning, and education. *Handbook of Face Perception. Oxford University Press, USA*, 2010.

[9] Benoit Bediou, Ilham Ryff, Bernadette Mercier, Maud Milliery, Marie-Anne Henaff, Thierry D'Amato, Marc Bonnefoy, Alain Vighetto, and Pierre Krolak-Salmon. Impaired social cognition in mild alzheimer disease. *Journal of geriatric psychiatry and neurology*, 22(2):130–140, 2009.

[10] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[11] Xavier P Burgos-Artizzu, Pietro Perona, and Piotr Dollár. Robust face landmark estimation under occlusion. In *Proceedings of the IEEE international conference on computer vision*, pages 1513–1520, 2013.

[12] Wei-Lun Chao, Jun-Zuo Liu, and Jian-Jiun Ding. Facial age estimation based on label-sensitive learning and age-oriented regression. *Pattern Recognition*, 46(3):628–641, 2013.

[13] Cuixian Chen, Yaw Chang, Karl Ricanek, and Yishi Wang. Face age estimation using model selection. In *Proc. IEEE CVPR Workshops*, pages 93–99. IEEE, 2010.

[14] Jen-Hau Chen, Kun-Pei Lin, and Yen-Ching Chen. Risk factors for dementia. *Journal of the Formosan Medical Association*, 108(10):754–764, 2009.

[15] Shixing Chen, Caojin Zhang, Ming Dong, Jialiang Le, and Mike Rao. Using ranking-cnn for age estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5183–5192, 2017.

[16] Sydney R Coleman and Rajiv Grover. The anatomy of the aging face: volume loss and changes in 3-dimensional topography. *Aesthetic surgery journal*, 26(1_Supplement):S4–S9, 2006.

[17] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[18] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[19] Antitza Dantcheva, Piotr Bilinski, Jean Claude Broutart, Philippe Robert, and Francois Bremond. Emotion facial recognition by the means of automatic video analysis. *Gerontechnology Journal*, 2016.

[20] Twisha Dhimar and Kinjal Mistree. Feature extraction for facial age estimation: A survey. In *Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on*, pages 2243–2248. IEEE, 2016.

[21] Hamdi Dibeklioğlu, Theo Gevers, Albert Ali Salah, and Roberto Valenti. A smile can reveal your age: Enabling facial dynamics in age estimation. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 209–218. ACM, 2012.

[22] Chi Nhan Duong, Kha Gia Quach, Khoa Luu, Hoai Bac Le, and K Ricanek. Fine tuning age estimation with global and local facial features. In *IEEE International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic*, pages 2032–2035, 2011.

[23] S Escalera X Baro I Guyon R Rothe. E Agustsson, R Timofte. Apparent and real age estimation in still images with deep residual regressors on appa-real database. In *12th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2017*. IEEE, 2017.

[24] Eran Eidinger, Roee Enbar, and Tal Hassner. Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security*, 9(12):2170–2179, 2014.

[25] Mohamed Y El Dib and Motaz El-Saban. Human age estimation using enhanced bio-inspired features (ebif). In *2010 IEEE International Conference on Image Processing*, pages 1589–1592. IEEE, 2010.

[26] Mahyar Etminan, Sudeep Gill, and Ali Samii. Effect of non-steroidal anti-inflammatory drugs on risk of alzheimer's disease: systematic review and meta-analysis of observational studies. *Bmj*, 327(7407):128, 2003.

[27] Oren Friedman. Changes associated with the aging face. *Facial Plastic Surgery Clinics*, 13(3):371–380, 2005.

[28] Yanwei Fu, Timothy M. Hospedales, Tao Xiang, Yuan Yao, and Shaogang Gong. Interestingness prediction by robust learning to rank. In *ECCV*, 2014.

[29] Yun Fu, Guodong Guo, and Thomas S Huang. Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 32(11):1955–1976, 2010.

[30] Xin Geng, Chao Yin, and Zhi-Hua Zhou. Facial age estimation by learning from label distributions. *IEEE transactions on pattern analysis and machine intelligence*, 35(10):2401–2412, 2013.

[31] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019.

[32] Asuman Gunay and Vasif V Nabiyev. Automatic age classification with lbp. In *Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on*, pages 1–4. IEEE, 2008.

[33] Asuman Günay and Vasif V Nabiyev. Age estimation based on hybrid features of facial images. In *Information Sciences and Systems 2015*, pages 295–304. Springer, 2016.

[34] Guodong Guo, Yun Fu, Thomas S Huang, and Charles R Dyer. Locally adjusted robust regression for human age estimation. In *Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on*, pages 1–6. IEEE, 2008.

[35] Guodong Guo, Guowang Mu, Yun Fu, and Thomas S Huang. Human age estimation using bio-inspired features. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 112–119. IEEE, 2009.

[36] Zhenhua Guo, Lei Zhang, and David Zhang. A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*, 19(6):1657–1663, 2010.

[37] Abdenour Hadid, Matti Pietikäinen, and Stan Z Li. Learning personal specific facial dynamics for face recognition from videos. In *International Workshop on Analysis and Modeling of Faces and Gestures*, pages 1–15. Springer, 2007.

[38] Hu Han, Charles Otto, Xiaoming Liu, and Anil K Jain. Demographic estimation from face images: Human vs. machine performance. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):1148–1161, 2015.

[39] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[40] Marko Heikkilä, Matti Pietikäinen, and Cordelia Schmid. Description of interest regions with local binary patterns. *Pattern recognition*, 42(3):425–436, 2009.

[41] John R Hodges, David P Salmon, and Nelson Butters. Recognition and naming of famous faces in alzheimer's disease: A cognitive analysis. *Neuropsychologia*, 31(8):775–788, 1993.

[42] Di Huang, Yunhong Wang, and Yiding Wang. A robust method for near infrared face recognition based on extended local binary pattern. In *International Symposium on Visual Computing*, pages 437–446. Springer, 2007.

[43] Yonggang Huang, Yunhong Wang, and Tieniu Tan. Combining statistics of geometrical and correlative features for 3d face recognition. In *BMVC*, pages 879–888. Citeseer, 2006.

[44] V Ilankovan. Anatomy of ageing face. *British Journal of Oral and Maxillofacial Surgery*, 52(3):195–202, 2014.

[45] Bas A In'T Veld, Annemieke Ruitenberg, Albert Hofman, Lenore J Launer, Cornelia M van Duijn, Theo Stijnen, Monique MB Breteler, and Bruno HC Stricker. Nonsteroidal antiinflammatory drugs and the risk of alzheimer's disease. *New England Journal of Medicine*, 345(21):1515–1521, 2001.

[46] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[47] Taskeed Jabid, Md Hasanul Kabir, and Oksam Chae. Facial expression recognition using local directional pattern (ldp). In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1605–1608. IEEE, 2010.

[48] Bihan Jiang, Michel Valstar, Brais Martinez, and Maja Pantic. A dynamic appearance descriptor approach to facial actions temporal modeling. *IEEE transactions on cybernetics*, 44(2):161–174, 2013.

[49] Bihan Jiang, Michel F Valstar, Brais Martinez, and Maja Pantic. A dynamic appearance descriptor approach to facial actions temporal modeling. *IEEE Trans. Cybernetics*, 44(2):161–174, 2014.

[50] Ian Jolliffe. Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer, 2011.

[51] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874, 2014.

[52] Terrence C Keaney. Aging in the male face: intrinsic and extrinsic factors. *Dermatologic Surgery*, 42(7):797–803, 2016.

[53] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[54] Russell A Kirsch. Computer determination of the constituent structure of biological images. *Computers and biomedical research*, 4(3):315–328, 1971.

[55] Sarah N Kohail. Using artificial neural network for human age estimation based on facial images. In *Innovations in Information Technology (IIT), 2012 International Conference on*, pages 215–219. IEEE, 2012.

[56] Igor O Korolev. Alzheimer's disease: a clinical and basic science review. *Medical Student Research Journal*, 4:24–33, 2014.

[57] Andreas Lanitis, Chrisina Draganova, and Chris Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):621–628, 2004.

[58] Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455, 2002.

[59] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[60] Seong-Whan Lee. Off-line recognition of totally unconstrained handwritten numerals using multilayer cluster neural network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6):648–652, 1996.

[61] Gil Levi and Tal Hassner. Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 34–42, 2015.

[62] Chin-Teng Lin, Dong-Lin Li, Jian-Hao Lai, Ming-Feng Han, and Jyh-Yeong Chang. Automatic age estimation system for face images. *International Journal of Advanced Robotic Systems*, 9(5):216, November 2012.

[63] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[64] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[65] Khoa Luu, Tien Dai Bui, Ching Y Suen, and Karl Ricanek. Spectral regression based age determination. In *Proc. IEEE CVPR Workshops*, pages 103–107. IEEE, 2010.

[66] Bappaditya Mandal and Nizar Ouarti. Spontaneous vs. posed smiles-can we tell the difference? *arXiv preprint arXiv:1605.07026*, 2016.

[67] Bangalore S Manjunath and Wei-Ying Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8):837–842, 1996.

[68] Markus Mathias, Rodrigo Benenson, Marco Pedersoli, and Luc Van Gool. Face detection without bells and whistles. In *European conference on computer vision*, pages 720–735. Springer, 2014.

[69] Seyyid Ahmed Medjahed. A comparative study of feature extraction methods in images classification. *International Journal of Image, Graphics and Signal Processing*, 7(3):16, 2015.

[70] Bryan Mendelson and Chin-Ho Wong. Changes in the facial skeleton with aging: implications and clinical applications in facial rejuvenation. *Aesthetic plastic surgery*, 36(4):753–760, 2012.

[71] Sebastian Mika, Gunnar Ratsch, Jason Weston, Bernhard Scholkopf, and Klaus-Robert Mullers. Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop.*, pages 41–48. Ieee, 1999.

[72] Meredith Minear and Denise C Park. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4):630–633, 2004.

[73] Bingbing Ni, Zheng Song, and Shuicheng Yan. Web image and video mining towards universal and robust age estimator. *IEEE Transactions on Multimedia*, 13(6):1217–1229, 2011.

[74] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.

[75] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[76] Ville Ojansivu and Janne Heikkilä. Blur insensitive texture classification using local phase quantization. In *International conference on image and signal processing*, pages 236–243. Springer, 2008.

[77] Sampo V Paunonen, Ken Ewan, Jillian Earthy, Sarah Lefave, and Heather Goldberg. Facial features as personality cues. *Journal of personality*, 67(3):555–583, 1999.

[78] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238, 2005.

[79] Jhony K Pontes, Alceu S Britto Jr, Clinton Fookes, and Alessandro L Koerich. A flexible hierarchical approach for facial age estimation based on multiple features. *Pattern Recognition*, 54:34–51, 2016.

[80] Lutz Prechelt. Early stopping-but when? In *Neural Networks: Tricks of the trade*, pages 55–69. Springer, 1998.

[81] Judith MS Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1):15–19, 1970.

[82] N Puizina-Ivic. Skin aging. *Acta Dermatovenerologica Alpina Panonica Et Adriatica*, 17(2):47, 2008.

[83] Hongwei Qin, Junjie Yan, Xiu Li, and Xiaolin Hu. Joint training of cascaded cnn for face detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3456–3465, 2016.

[84] Narayanan Ramanathan, Rama Chellappa, Soma Biswas, et al. Age progression in human faces: A survey. *Journal of Visual Languages and Computing*, 15:3349–3361, 2009.

[85] Rajeev Ranjan, Vishal M Patel, and Rama Chellappa. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[86] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 341–345. IEEE, 2006.

[87] Michael J Richard, Carrie Morris, Byron F Deen, Linda Gray, and Julie A Woodward. Analysis of the anatomic changes of the aging facial skeleton using computer-assisted tomography. *Ophthalmic Plastic & Reconstructive Surgery*, 25(5):382–386, 2009.

[88] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2-4):144–157, 2018.

[89] M Roudier, P Marcie, A-S Grancher, C Tzortzis, S Starkstein, and F Boller. Discrimination of facial identity and of emotions in alzheimer's disease. *Journal of the Neurological Sciences*, 154(2):151–158, 1998.

[90] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[91] Ulrich Seidl, Ulrike Lueken, Philipp A Thomann, Andreas Kruse, and Johannes Schröder. Facial expression in alzheimer's disease: impact of cognitive deficits and neuropsychiatric symptoms. *American Journal of Alzheimer's Disease & Other Dementias®*, 27(2):100–106, 2012.

[92] MC Smith. Facial expression in mild dementia of the alzheimer type. *Behavioural Neurology*, 1995.

[93] Irwin Sobel and Gary Feldman. A 3x3 isotropic gradient operator for image processing. *a talk at the Stanford Artificial Project in*, pages 271–272, 1968.

[94] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2014.

[95] I-O Stathopoulou and George A Tsihrintzis. An improved neural-network-based face detection and facial expression classification system. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 1, pages 666–671. IEEE, 2004.

[96] M Merlin Steffi and J John Raybin Jose. Comparative analysis of facial recognition involving feature extraction techniques. 2018.

[97] Jinli Suo, Tianfu Wu, Songchun Zhu, Shiguang Shan, Xilin Chen, and Wen Gao. Design sparse features for age estimation using hierarchical face model. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008.

[98] Kestutis Sveikata, Irena Balciuniene, and Janina Tutkuviene. Factors influencing face aging. literature review. *Stomatologija*, 13(4):113–6, 2011.

[99] Michael A Taister, Sandra D Holliday, and HIM Borrmman. Comments on facial aging in law enforcement investigation. *Forensic science communications*, 2(2), 2000.

[100] Sibt ul Hussain and Bill Triggs. Visual recognition using local quantized patterns. In *Computer Vision–ECCV 2012*, pages 716–729. Springer, 2012.

[101] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001.

[102] Karin Wolffhechel, Jens Fagertun, Ulrik Plesner Jacobsen, Wiktor Majewski, Astrid Sofie Hemmingsen, Catrine Lohmann Larsen, Sofie Katrine Lorentzen, and Hanne Jarmer. Interpretation of appearance: The effect of facial features on first impressions and personality. *PloS one*, 9(9):e107721, 2014.

[103] Yue Wu, Tal Hassner, KangGeon Kim, Gerard Medioni, and Prem Natarajan. Facial landmark detection with tweaked convolutional neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 2017.

[104] Allan E Wulc, Pooja Sharma, and Craig N Czyz. The anatomic basis of midfacial aging. In *Midfacial Rejuvenation*, pages 15–28. Springer, 2012.

[105] Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Faceness-net: Face detection through deep facial part responses. *IEEE transactions on pattern analysis and machine intelligence*, 40(8):1845–1859, 2018.

[106] Agustín G Yip, Robert C Green, Matthew Huyck, L Adrienne Cupples, and Lindsay A Farrer. Nonsteroidal anti-inflammatory drug use and alzheimer's disease risk: the mirage study. *BMC geriatrics*, 5(1):2, 2005.

[107] Abdullah Emir Zeylan, Albert Ali Salah, Hamdi Dibeklioğlu, Zeynep Tüfekçioğlu, Başar Bilgiç, and Murat Emre. Do alzheimer's patients appear younger than their age a study with automatic facial age estimation. In *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2019.

[108] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[109] Song Yang Zhang, Zhifei and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[110] Wenchao Zhang, Shiguang Shan, Wen Gao, Xilin Chen, and Hongming Zhang. Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 786–791. IEEE, 2005.

[111] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Facial landmark detection by deep multi-task learning. In *European conference on computer vision*, pages 94–108. Springer, 2014.

[112] Guoying Zhao and Matti Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):915–928, 2007.

[113] Marc S Zimbler, Mimi S Kokoska, and J Regan Thomas. Anatomy and pathophysiology of facial aging. *Facial plastic surgery clinics of North America*, 9(2):179–87, 2001.