الجمهورية الجزائرية الديمقراطية الشعبية

Democratic and Popular Republic of Algeria

وزارة التعليم العالي و البحث العلمي

Ministry of Higher Education and Scientific Research

Mohamed Khider university – Biskra
Faculty of Science and Technology
Department of Electrical Engineering
Réf : .....

جامعة محمد خيضر بسكرة
كلية العلوم و التكنولوجيا
قسم الهندسة الكهربائية
المرجع: .....

Thesis presented for obtaining the LMD Doctorate degree

Electronics

Facial Age Estimation

Defended by:

GUEHAIRIA Oussama

In front of the jury composed of:

KAZAR Okba	Professor, Biskra university	president
OUAFI Abdelkrim	Professor, Batna university	Examiner
AJGOU Riad	Dr, Oued university	Examiner
OUAMANE Abdelmalik	Dr, Biskra university	Supervisor
DORNAIKA Fadi	Professor, San Sebastian university Spain	Co-supervisor

Acknowledgements

First and foremost, my deep thanks and gratitude to Almighty ALLAH, for His graces and blessings and for giving me strength and ability throughout the years of the PhD to successfully complete my thesis.

I would like to thank very sincerely my thesis supervisor Dr. ***Abdelmalik OUAMANE*** Associate Professor at Biskra University, for welcoming me in his research team. I would like to express my deep gratitude for the trust that he has placed in me and for his support and advices.

I address my deep appreciation to my co-supervisor Dr. Fadi DORNAIKA Professor at IKERBASQUE Research Foundation,Spain and Pr. Abdelmalik Taleb-Ahmed Professor at Polytechnic University of North France, Valenciennes for their relevant advices, explanations, insightful comments, and their support throughout the years of this thesis.

I should like to thank all the ***members of the Jury*** for your invaluable work. thank you for your effort and for giving me the opportunity to present our work front of you.

At the end of this long-term work, it is with great pleasure that I would like to thank my family: ***my parents***, my sisters ***Asma*** and ***Imen*** , ***my Uncles***, my brothers ***Imed OTHMANI*** for his motivation and support, ***Zakaria Siad.*** my friends ***khaireddin and yaakoub***

My mother **ROUKAIA** thank you for encouraging me and for being constant support at every step in my life. In addition, I thank my father **Nour eddin,** who helped me to go successfully through all these years.

I would like also to thank my fiancee ***Meriem BELALOUI*** for her encouragements and support.

To my family

To all my friends.

Publications & Communications associated with the thesis

I. International publication

1. **O Guehairia**, A Ouamane, F Dornaika, and A Taleb-Ahmed. "**Feature Fusion Via Deep random forest for facial age estimation**." Neural Networks, 130:238–252, 2020.

II. International Communication

2. **O. Guehairia**, A. Ouamane, F. Dornaika, & A. Taleb-Ahmed (2020). **Deep Random Forest for Facial Age Estimation Based on Face Images**. In 2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP) (pp. 305-309). El-Oued University, El-Oued ,Algeria.

Abstract

In the last few years, human age estimation from face images attracted the attention of many researchers in computer vision and machine learning fields. This is due to its numerous applications.

In this thesis, we propose a two new architectures for age estimation based on facial images.The first one is mainly based on a cascade of classification trees ensembles, which are known recently as a Deep Random Forest (**DRF**). This first proposed architecture is composed of two types of DRF. The first type extends and enhances the feature representation of a given facial descriptor. The second type operates on the fused form of all enhanced representations in order to provide a prediction for the age while taking into account the fuzziness property of the human age. While the proposed methodology is able to work with all kinds of image features, the face descriptors adopted in this work used off-the-shelf deep features allowing to retain both the rich deep features and the powerful enhancement and decision provided by the proposed architecture. Experiments conducted on six public databases prove the superiority of the proposed architecture over other state-of-the-art methods.

The seconde method extends and improves the previous scheme for fusing multiple deep face features for age estimation. This scheme was based on Deep Random Forests (**DRF**). We propose a new pipeline that integrates tensor based subspace learning before applying the DRFs. Deep face features of a training set are represented as a 3D tensor. Multi-linear Whitened Principal Component (MWPCA) and Tensor Exponential Discriminant (TEDA) are used to extract the most discriminant information. The features of the tensor subspace are then fed to DRFs in order to predict the age. Experiments conducted on five public face databases show that the method can compete with many state-of-the art methods.

Keywords:

Age estimation, deep features, decision trees, deep learning,random forest, deep random forest,feature (DRF) fusion, Multi-linear Whitened Principal Component, Tensor Exponential Discriminant Analysis, Tensor Based Subspace.

Résumé

Au cours des dernières années, l'estimation de l'âge humain à partir d'images de visage à attiré l'attention de nombreux chercheurs dans les domaines de la vision par ordinateur et de l'apprentissage automatique. Cela est dû à ses nombreuses applications.

Dans cette thèse, nous proposons deux nouvelles architectures pour l'estimation de l'âge basées sur des images faciales. La première est principalement basée sur une cascade d'ensembles d'arbres de classification (Random Forest), qui sont connus récemment comme une forêt au hasard profond (DRF). Cette première architecture proposée est composée de deux types de DRF. Le premier type étendre et améliore la représentation des caractéristiques d'un descripteur facial donné. Le deuxième type fonctionne sur la forme fusionnée de toutes les représentations déjà améliorées afin de fournir une prédiction de l'âge tout en tenant compte de la propriété ambiguïté de l'âge humain. Bien que la méthodologie proposée soit en mesure de fonctionner avec toutes sortes de caractéristiques d'image, les descripteurs de visage adoptés dans ce travail ont utilisé des caractéristiques profondes disponibles permettant de conserver à la fois les riches fonctionnalités profondes et la puissante amélioration et la décision fournies par l'architecture proposée. Les expériences menées sur six bases de données publiques prouvent la supériorité de l'architecture proposée par rapport à d'autres méthodes de pointe.

La deuxième méthode s'étend et améliore le schéma précédent, par fusionner plusieurs caractéristiques profondes du visage pour l'estimation de l'âge. Ce schéma était basé sur les forêts aléatoires profondes (DRFs). Nous proposons un nouveau pipeline qui intègre l'apprentissage subspatial basé sur des tenseurs avant d'appliquer les DF. Les caractéristiques profondes d'un ensemble de formation sont représentées comme un tenseur 3D. Multi-linéaire Whitened Principal Component (MWPCA) et Tensor Exponential Discriminant (TEDA) sont utilisés pour extraire les informations les plus discriminantes. Les caractéristiques du sous-espace tenseur sont ensuite alimentées aux DRFs afin

de prédire l'âge. Des expériences menées sur cinq bases de données publiques montrent que la méthode peut rivaliser avec de nombreuses méthodes dans la littérature scientifique.

Mots-clés :

Estimation d'âge, caractéristiques profondes, arbres de décision, Apprentissage profond, forêt aléatoire, forêt aléatoire profonde, fusion de caractéristiques, composant principal blanchi multi-linéaire, les tenseurs, MWPCA, TEDA, sousespace basé sur un tenseur.

ملخص

في السنوات القليلة الماضية، جذبت عملية تقدير العمر البشري من صور الوجه انتباه العديد من الباحثين في مجالات رؤية الكمبيوتر والتعلم الآلي. ويرجع ذلك إلى العديد من التطبيقات .

في هذه الأطروحة، نقترح اثنين من النماذج الجديدة لتقدير العمر على أساس صور الوجه. ويستند الأول أساسا على سلسلة من مجموعات الأشجار التصنيف، والتي تعرف مؤخرا باسم غابة عشوائية عميقة. يتكون هذا الهيكل المقترح الأول من نوعين من النوع الأول يمتد ويعزز تمثيل ميزة واصف الوجه معين .DRF .النوع الثاني يعمل على شكل تنصهر من جميع التمثيلات المحسنة من أجل توفير تنبؤ للعمر مع الأخذ في الاعتبار خاصية الزغب في عصر الإنسان. في حين أن المنهجية المقترحة قادرة على العمل مع جميع أنواع ميزات الصورة، فإن واصفات الوجه المعتمدة في هذا العمل تستخدم ميزات عميقة جاهزة تسمح بالاحتفاظ بكل من الميزات العميقة الغنية وتعزيز قوي وقرار المقدم من البنية المقترحة. التجارب التي أجريت على ست قواعد بيانات عامة تثبت تفوق النماذج المقترحة على غيرها من الأساليب الحديثة .

طريقة الثانية تمتد ويحسن المخطط السابق لصهر متعددة ملامح الوجه العميق لتقدير العمر. واستند هذا . نقترح خط تنفيذDRF المخطط على الغابات العشوائية العميقة دّ جديد يدمج التعلم الفضاء الفرعي المتشد قبل تطبيق ملفات. يتم تمثيل ميزات الوجه العميق لمجموعة التدريب على أنها موتر ثلاثي الأبعاد. يتم استخدام مكون مبيض متعدد الخطي لاستخراج معظم (TEDA) والموتر الأسي المميز (MWPCA) المعلومات غير المميزة. ثم يتم تغذية ملامح الفضاء الفرعي الموتر إلى من أجل التنبؤ بالعمر DRF .التجارب التي أجريت على خمس قواعد بيانات الوجه العامة تبين أن الأسلوب يمكن أن تتنافس مع العديد من أساليب الأعمال المنشورة .

الكلمات المفتاحية :

تقدير العمر، ميزات عميقة، أشجار القرار، التعلم العميق، الغابات العشوائية، الغابات العشوائية العميقة، دمج الميزة ،MWPCA، TEDA، فضاء فرعي قائم على موتر.

Contents

List of Figures

List of Tables

Main acronyms and notations used in the work.

Acronym and Notation	Description
AAM	Active Appearance Model
BIF	Bio Inspired Feature
CNN	Convolutional Neural Network
DEX	Deep Expectation
DCNN	Deep Convolutional Neural Network
DRF	Deep Random Forest
DMTL	Deep Multi Task Learning
ERT	Ensemble of Regression Trees
GEI	Gait Energy Image
HCI	Human Computer Interaction
LAP	Looking At People
LBP	Local Binary Pattern
LPQ	Local Phase Quantization
MAE	Mean Absolute Error
MSG	Multi Grained Scanning
LSDML	Label Sensitive Deep Metric Learning
ODLA	Ordinal Deep Learning Approach
PCA	Principal Component Analysis
MPCA	Multilinear Principal Component Analysis
MWPCA	Multilinear Whitened Principal Component Analysis
LDA	Linear Discriminant Analysis
SVM	Support Vector Machine
SVR	Support Vector Regression
GB	Gabor Wavelets
C	Class number
D	Original Feature vector size
d	Number of features
F	Number of forests
V	Number of feature vectors
n	Number of samples
L	Number of layers (levels)
n_{trees}	Number of trees

List of Algorithms

Chapter 1

Overall introduction

1.1 Introduction

In recent years, the researchers' main area of inclination is towards biometric authentication owing to its possible use in surveillance systems, social security. Biometrics consists of technologies that are utilized in measuring and analysing the particular and unique features of an individual. Tow kinds of biometrics can be counted: behavioral and physical. The behavioral on is basically used for verification where physical biometrics may be employed for both verification identification. Human being, have certain features that are unique and can be used as biometrics, such as the fingerprint, the eye retina and many more. Biometrics technologies consist of several varieties of methods used in order to identify individuals in addition to automate the authentication of identity by making use of either the physical or behavioral charachters of the person involved. Among the many biometric techniques. Biometric methods involve the use of fingerprints, voice, iris, or face... Etc. They each have their advantages and limitations. Some methods are rigorous but are also very restrictive (high

cost, the collaboration of the indispensable person in the majority of cases, etc.) while others are more user-friendly face from precision problems. For the characteristics, specific to each individual, to be classified as biometric modalities, they must be:

- Universal (exist in all individuals.),
- Unique (possibility of differentiating one individual from another),
- Permanent (may evolve over time),
- Recordable (possibility to record the characteristics of an individual with his agreement),
- Measurable (possibility of future comparisons).

Study on the numerical analysis of human faces (including object identification, face recognition, gender classification, recognition of facial expression and age estimation) has drawn attention in the machine vision and pattern recognition communities, with a growing interest in social robotics and video-based surveillance systems [2, 3].

1.2 Age estimation : Motivation

The quality of interactions with an individual can be improved through a prior-knowledge of his\her age. This was a major reason for researchers to take part in the age estimation research field. The facial age estimation via face images considered a new-born field in the recent years. Due to its importance and wide applicability in the modern society, this subject flourishes and grows, year by year, to be a salient one.

Automatic age estimation in face images has proven to be of a great value. It was thus adopted in many areas like Human Computer Interaction (HCI), security and management applications. Furthermore, many companies and advertisers rely on age categories when they recommend products and services to their clients.

A set of utilities can be offered by automatic age estimation. It can boost surveillance systems and facilitate the investigations made by the police

1.3 Age Estimation: Main challenges

Age estimation in face images encounters a variety of challenges, some of which are caused by human aging process itself that cannot be governed by the control process (see Figure 1.1). Where the advancement of aging is uncontrollable and no one is able to age whenever and wherever, it is also exceedingly laborious to obtain enough training data for age estimation. While a few other challenges can be grouped in two different categories of factors: intrinsic and extrinsic factors [4]. Intrinsic factors are related to health conditions. However, the extrinsic factors are external to the health conditions. These factors can be related to the living style and the working environment. Other challenges include the sample origin (i.e., society and region) which can influence the estimation precision.

Furthermore, the automatic age estimation should overcome the challenge of close cross age correlation, i.e., a man of 40 years of age looks almost the same as in his 39 and 41 years of age. The latter encouraged a lot of researchers to lean toward the regression solutions [5], [6], [7]. Other works view age estimation

either as a multi-class classification problem [8], [9], or as a combination of regression and multi-class classification [10], [11].

Figure 1.1: Example of aging effect on a subject from FG-NET database. The appearance of the subject's face was affected considerably by aging.

1.4 Proposed Methods

the work in the third Chapter 3 : Feature Fusion via Deep Random Forest For Facial Age Estimation encouraged by the proposed method by Zhou et al. [12]. The authors in the mentioned work come up with a new decision trees ensembles Approach to a wide variety of classification tasks. The interest with this method is its nearest performance and accuracy with those of Deep Neural Networks. The gcForest or Deep Random Forest is the named of this method. the principal of gcForest is the sequence structure generated by an ensemble of forests, started by the Multi Grained Scaning as a feature extraction. As in the deep neural networks, layer-by-layer processing is the manner of learning representation for the raw features. The authors asserted that the training of these models is a lot simpler than training deep CNNs. They applied their examples to classic recognition problems. the original gcForest doesn't suit the age estimation problems. In gcForest structure, the first layer takes as input rawbrightness patches (obtained by sliding windows as it appears in Figure 4.3) in the original image. The patches are fed to the random forests to get an encoding. While this stunt delivered astounding outcomes for classic object recognition and classification problems, it neglected the issue of age estimation

in a satisfactory way. Which motivated us to propose a method that can be more adequate with the concerned problem. We introduce two main modifications: (i) we use image descriptors that can be either hand-crafted or provided by a pretrained deep CNN, and (ii) we propose a novel architecture for the deep random forests allowing to fuse any types of image descriptors. In the current work, although we use the DEX Chalearn pretrained CNN model to extract image's features, any other types of image features can be used and fused. For each type of image features, we create an ensemble of Random Forests as in [12] (that will be later called Deep Random Forest-Fusion (DRF-Fusion)) to extract a representation vector with more information. In a first phase, we carried out a fusion on the ouputs of each type to get one fused representation vector. In a final stage, we applied a different form of DRF, namely fd-DRF (final Decision-Deep Random Forest) to the fused representation vector and generated the predicted age. In fd-DRF, a modified decision function is adopted by imitating some deep learning based models. This decision concerns the final output of the proposed architecture and uses the N_{max} probabilities parameter (provided by the user), to select the N_{max} ages having the largest probabilities. It then calculates the final prediction age through their arithmetic mean. The main contributions of this work are summarized as follows:

- A novel deep architecture for Random Forests that is applied to the facial age estimation problem.
- The architecture contains two main parts:
    1. The first part encodes and fuses the features of data representations.
    2. The second part is a Deep Random Forest structure that provides final age prediction using the largest N_{max} probabilities.

- The proposed architecture allows the integration and fusion of different types of image descriptors.

In the Fourth chapter we propose a novel approach able to reduce many of the above limitations. In our proposed second approach we use pre-trained CNN models in order to extract features from face images. These features provided by different nets will be used as input features to our estimator. The latter is composed of tensor transformations and Deep Random Forests.

Thus far, subspace transformation is the furthermost utilized dimensional reduction techniques [13, 14]. Various reduced dimensional algorithms have been proposed in the preceding period that have suited the feature extraction. The Principal Component Analysis (PCA) [15] and Linear Discriminant Analysis LDA [16] are frequently used. They are linear subspace techniques. Mainly, an image face is a matrix of m by n pixels, which is treated as a 1 D feature vector of size $m \times n$. Unfortunately, this process involves losing the pixels' position information [17]. Recently, multilinear subspace techniques based on tensor analysis of data in high dimensional spaces is regarded as a remarkable multi-linear technique [18]. These approaches authorize the conservation of the important face structure information. Multilinear transformations analyze the multifactor structure of image face sets over n different index number.

The common linear subspace methods PCA and LDA are extended to Multilinear PCA (MPCA) [17] and Multilinear Discriminant Analysis (MDA) [19] that allow the mathematical of tensors to be manipulated. The high tensor order (i.e., ¿ 2) are presented in a normal form to show the set of face images without collapsing the initial structure and correlation of data [20]. In [1], the authors propose a new use of an adopted MPCA, this latter is named Multilinear

Whitened Principal Component Analysis (MWPCA), which can deal with the small sample size issue in high dimensional space and can enhance the tough discrimination obtained by classical MPCA. The multilinear varied analysis MDA was also extended to Tensor Exponential Discriminant analysis TEDA so as to improve the discriminant data included in the null space of the within class scatter matrix of each tensor's mode. TEDA increases the margin amidst samples belonging to multiple classes by distance diffusion mappings. The main contributions of this work are the following:

- We propose a multiview feature fusion that enhances the performance of our previous proposed method in [21] and the techniques in [1].
- We fuse the deep features using the Whitened principal component analysis (MWPCA) and Tensor exponential discriminant analysis (TEDA), respectively.
- Once the face image features are represented in the tensor subspace, the final age is estimated using our recent Deep Random Forests (DRF) [21].

1.5 Benchmark databases

1.5.1 MORPH

This database contains images of 13,618 individuals (males and females). It contains more than 55000 unique images. Each facial image is annotated with chronological age. Ages are between 16 and 77 years. MORPH can be divided into more than ethnicity: African, European, and others. Following

[22, 23, 24], we use the Caucasian subset, which contains 5492 images from the original MORPH database. We use the random split evaluation protocol on the Caucasian images. The 5,492 images are randomly partitioned to 80% training set and the other 20% testing set. It is repeated five times. The average of the five different splits will be the final performance.

1.5.2 FG-NET

This is a widely known database in age estimation. This database has a large variation in lighting conditions, pose and expression. FG-NET contains 1002 facial images associated with 82 individuals. Each individual has more than 10 photos taken at different ages. The FG-NET age range is from zero to 69. As in [25, 6], we use the "Leave One Person Out" cross-validation on FG-NET. We leave one individual image out for testing and the other 81 individuals images for training.

1.5.3 PAL

The Predictive Aging Lab face is another database from Texas university. It contains 1046 frontal face images (430 males, 616 females). PAL contains faces with different expressions. We perform the random partition as in [26, 7], where we randomly partition images in 80% training and the other 20% for testing. It is repeated five times. The average of the five different splits will be the final performance.

1.5.4 LFW+

The MSU LFW+ database [27] was created by extending the LFW database to study the joint attribute learning/estimation (age, gender, and race) from unconstrained face images "the images were taken in different positions and conditions and that what makes this database hard in the test." The extended LFW database (LFW+) contains 15,699 unconstrained face images of about 8,000 subjects. For each face image, three MTurk workers were asked to provide their estimates of age, gender, and race. The apparent age is determined as the average of the three estimates. we use the five-fold cross validation used in [27].

1.5.5 FACES

The FACES database contains 2052 face images from 171 persons. For each person, there are 6 expressions: neutral, sad, disgust, fear, angry, and happy. For evaluation, we used the five random split protocol as in MORPH, Caucasian, and PAL. We conducted the experiments on image subset having the same facial expression [25, 6].

1.5.6 APPA-REAL

The APPA-REAL database [28] contains 7591 images. It has a default split into train, test, and validation. We used the same setting that is described in [28]. This database contains two types of age labels: Real Age and Apparent Age label. The Apparent Age labels are gathered from around 300,000 votes. On average, around 38 votes per each image, and this makes the average apparent age very stable.

1.6 Evaluation Metric

To evaluate the performance of the proposed age estimation method, we used the Mean Absolute Error (MAE). It is one of the most known indicators for age estimator performance evaluation in literature. MAE calculates the average of absolute error between the predicted and the ground truth ages. It is given by:

$$MAE = \frac{1}{n}\sum_{t-1}^{n}|p_t - g_t|, \tag{1.1}$$

where n is the number of tested images, p_t is the predicted age of image t, and g_t is the ground-truth age of this image.

1.7 Thesis structure

The rest of this thesis was organized as follows: Chapter 2 is brief presentation of facial age estimaion exisiting techniques by order of appearence of this field. In Chapter 3 we present a literature review to facial age estimation, firstly we introduce the deep learning methods and their importance to biometrics. Then we summarize the performance terms for facial age estimation. Moreover, we give a brief of existing and a comparison between the latest works is given. Chapter 4 and 5 are the description of every methods and steps used in our approach. The last chapter is a general conclusion about our work and an envision for some future works.

Chapter 2

Overview of existing techniques

2.1 Introduction

Estimating the age automatically via facial images process is considered as an age guesstimate (age approximation) that is based on the numerical person's facial image analysis. The age approximation is done by estimating the exact age value of the given face image of an individual or by determining the age group.

Our work focuses on the estimation of the exact age value, in which a face image was automatically labeled with the estimated age through a learning process. The figure 2.1 shows the automatic facial age estimation system from a face image.

2.2 Overview of existing techniques

Pioneering works on facial age estimation faced a lot of challenges due to the scarcity of annotated image databases. The scientific community has put a

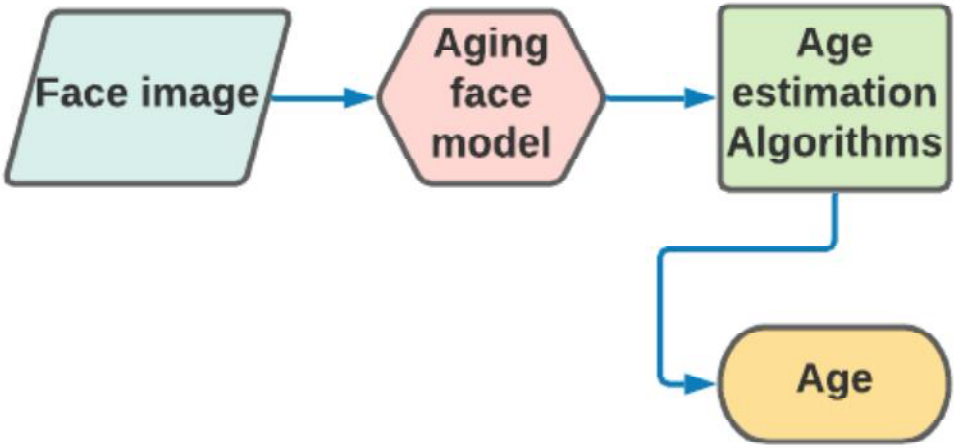


Figure 2.1: Flow chart presents the automatic age estimation systems.

great deal of work into developing models of age estimation based on the human face. For age estimation, the Image characterization can be warped up using several varied state-of-the-art models.

2.2.1 Anthropometric Models

This model was build on the individual faces measurements and proportions [29] as presented in Figure 2.2. The first work that showed interest in the subject was done by Kown and Lobo [30]. They classified subjects into babies, adults and senior adults. Their method utilizes the analysis of skin wrinkles and craniofacial shape evolution.

2.2.2 Active Appearnance Models

Lanitis et al. [32], later, used the Active Appearance Model (AAM) by considering both face anthropometrics and texture as presented in Figure 2.3. Many methods exploited hand-crafted features such as: Local Binary Pattern

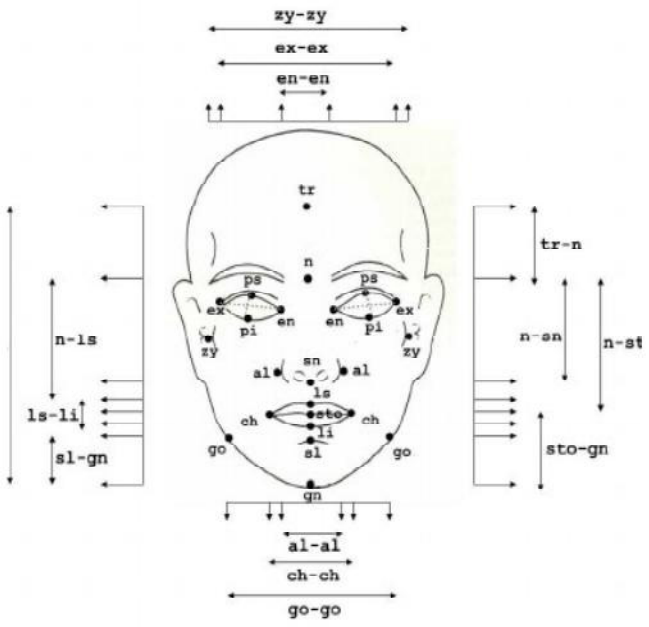
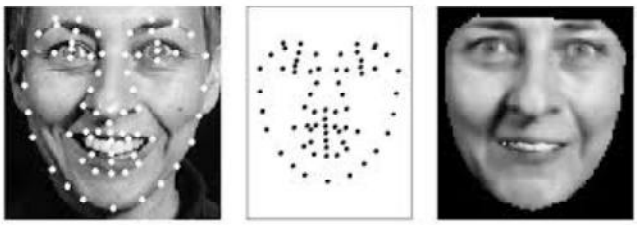


Figure 2.2: anthropometric models which based as illustrated on the measurements and proportions of the human faces [31]

(LBP) [33], Biologically Inspired Feature (BIF) [34] and Haar-like features [35]. Encouraged by the efficiency of gait representation [36] and Gait Energy Image (GEI) [37], Gabor features were utilized in [38]. A different approach that can be found in Gen et al. [39]. The authors considered each facial image as an instance linked with an age label distribution.



Labelled image Points Shape-free patch

Figure 2.3: Illustration of the face active appearance example [40]

2.2.3 Aging Patterns Subspaces models (AGES)

In this model the data structure , or the so called Aging pattern, deals with the sequence of a person aging face images as a whole instead of dealing with it [41, 42], it is based on the presentation of the sequence of facilal individual images which sorted in time order as illustrated in Figure 2.4. The main challenge in the aging pattern subspaces models is missing values, the aging patterns are always incomplete, there are many missing values in the aging pattern vector.

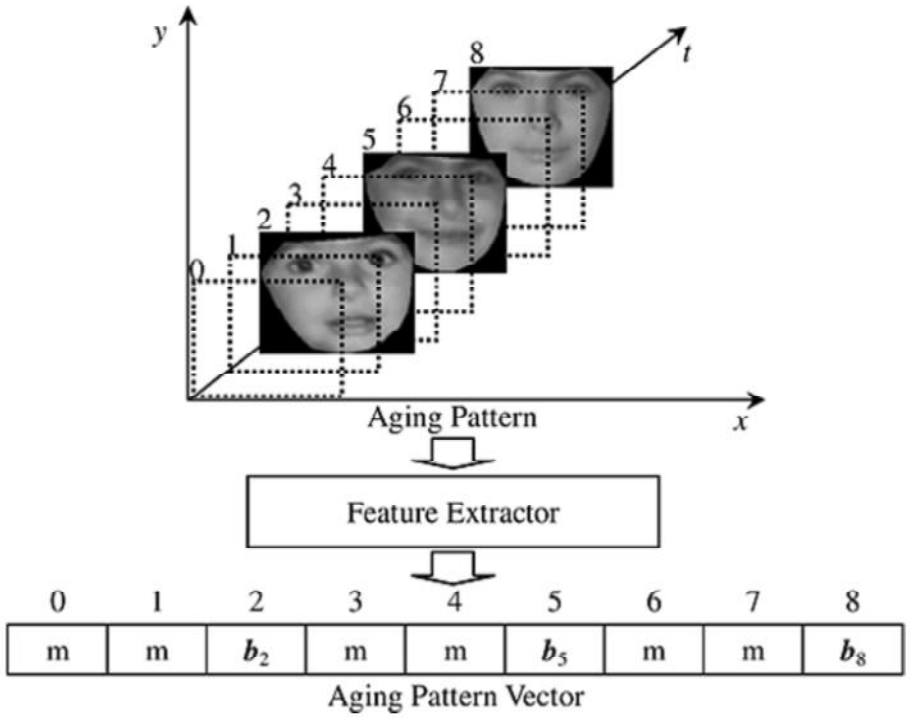


Figure 2.4: Presentation of the aging pattern subspace (AGES) where it presents as an sequence of individual face images sorted in time order [41]

2.2.4 Age manifold

Age manifold deals with age estimation problems as a special case of supervised manifold embeddind problems. Assume that the aging face images

are distributed on an intrinsic low-dimensional manifold (faces with close ages locate closely on the manifold)set in time order. Common approaches: Orthogonal Locality Preserving Projections (OLPP) [43, 44], Synchronized Submanifold Embedding (SSE)[45]

2.2.5 Hybrid Methods

Hybrid models are designed to achieve better age estimation accuracies as a combination of two distinct models. Many of the well-used facial features of human aging modeling extraction techniques are mentioned as Gabor filters [46], Linear Discriminant Analysis (LDA) [47], Local Binary Patters (LBP) [48], Local Directional Patterns (LDP) [49],Grassmann Manifold [50] and Biologically Inspired Features (BIFs) [51].

Chapter 3

State-of-the-Art

3.1 Introduction

The need to develop more accurate age estimation models is rendered necessary due to all of its recent various application areas. Numerous research works have been done in this area. However, more efficient and accurate results are achieved through the use of artificial neural networks for facial age estimation. In this chapter, a brief presentation of the deep learning methods that have recently been developed, their enhancement and evolution will be given, also their applications in the field of facial age estimation as in the work of Puniany et al. [52] where they detailed the differents types of CNN for facial age estimation based on facial image.

3.2 Deep Learning

In the last decade, Deep Learning (DL), a sub-field of machine learning, has witnessed a great interest within the artificial intelligence community.

Figure 3.1 depicts how the term 'Deep learning' has been traded in the last decade reaching its trending popularity peak. DL is based on layer-by-layer cascade structure, each layer consists of several nonlinear modules, which are called neurons. (i.e. neurons). Through the network layers, the information is passed from where the information at each hidden layer is hierarchically transformed to reach a higher abstraction level at the output layer. The multiple

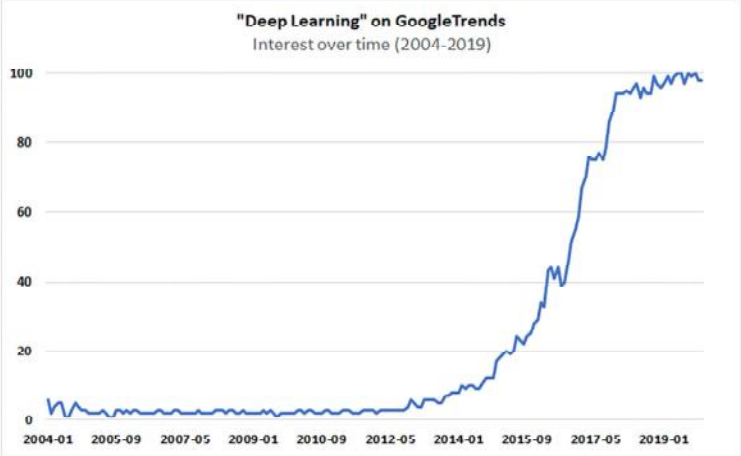


Figure 3.1: Trending of the term "Deep Learning" over the time period (2004-2019) estimated using Google Trends.

levels in the cascade structure make these deep models more generalized giving them the ability to automatically generate powerful features that can be applied directly on new domains without previous knowledge of them. Each level of deep models has a considerable number of trainable weights which can be learned using supervised, semi-supervised, or unsupervised learning procedure. , in which a large training dataset is required for the training. ([53]).

Even though the origin of DL backs to ([54], [55], [56]) , it has started attracting more attention by ([57], [58], [59]). Also, one of the key factors that have revived the research in DL is when the training of deep network architectures was accelerated using Graphics Processing Units (GPUs) in late

2006. Next, the CUDA programming platform was launched by NVIDIA company , which allowed for an enhanced exploitation of GPUs capabilities for parallel processing. ([60], [61]).

Next, Deep Learning has seen an exponential growth starting by 2012 when [62] have won the annual contest of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) for the year of 2012[1]. They proposed a Deep Learning approach based on a convolutional neural network model that is named 'AlexNet'. In the purpose of the classification task for a large hand-labelled image data-set (ImageNet ([63])), they drastically reduced the error rate to 17% compared to 27% of error rate using handcrafted engineered features. In the next year, only three teams proposed non-Deep Learning methods from a total of 24 participated teams in ILSVRC-2013, which was an omen sign to the imminent rise of Deep Learning techniques.

In summary, the success of Deep Learning is a consequence of its observable results. Where Deep Learning had outperformed the traditional handcrafted feature methods. This breakthrough was the result of three main factors: Firstly, the availability of large-scale datasets used to train the models used to train millions of the models parameters. Secondly, the great advancement of computational power hardware (Graphics Processing Units (GPUs)), which become faster. Finally, the considerable number of the open-source libraries and frameworks that are supported the Deep Learning technologies, but still there are many challanges to overcome such as the computations costs and the expensive hardware used.

[1] http://www.image-net.org/challenges/LSVRC/2012/

3.2.1 Artificial neural networks

Neural networks are defined as a structure that can copy and imitate how the human brain learns, in other words it perfectly works as the biological diagram of human neural system. As displayed in Figure 3.2, these systems include input layers, output layers and some sets of hidden layers which make appropriate connection between input and output layers. Each connection reports a weight associated with it and alterations in these weights directs the learning process. The latter comes in two types, supervised or unsupervised. As the name suggests, in supervised learning the output of the network is compared to the ground-truth (output labels) in the training phase. ANN compares its guess answers with correct answers to make modifications in the weights. In contrast, the unsupervised learning process is done in the absence of any supervision. Herein, clustering works best in which we divide the set of samples into groups based on some unknown pattern. There are essentially two types of Artificial neural networks named as Feed-forward and Feedback ANN.

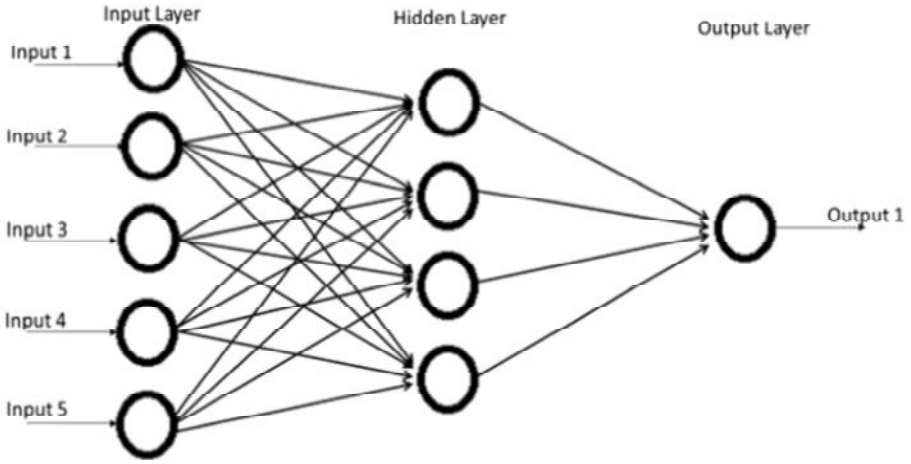


Figure 3.2: Basic architecture of Artificial Neural Networks. [64]

Both these types are depicted in Figure 3.9 a and b. As shown in this fgures,

feed-forward ANNs does not contain any feedback loops and are unidirectional in nature. They contain fixed number of inputs and outputs. On the other hand, Feed-back ANNs are made up of feedback loops which act as memory elements.

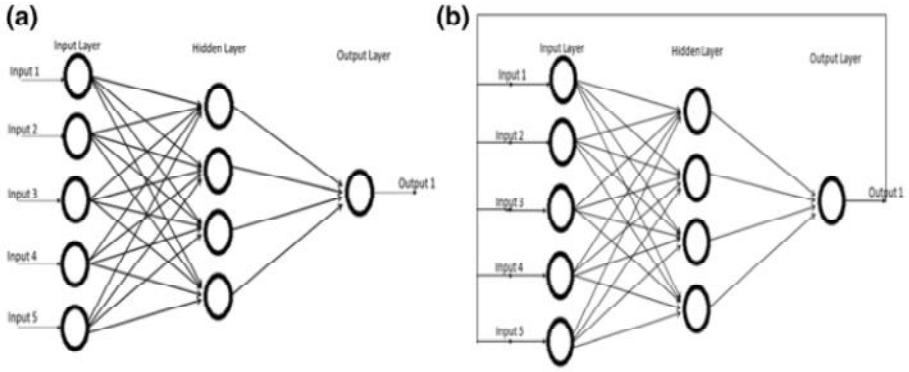


Figure 3.3: a) Feedforward ANN. b) Feedback ANN. [65]

3.2.2 Deep Neural Networks

Deep Neural Network (DNN) is a neural network with several layers. Each layer is comprised of several neurons, which form the basic of a computational unite in the network, with a specific activation function and parameters $\Theta = \{W, \gamma\}$ ([66]). The term "deep" is referred to using multiple hidden layers instead of one hidden layer. Generally, most of DNN architectures are based on feed-forward neural network in which there are no loops between units (no feedback connections). where the information is propagated from the input layer to the output via the hidden layers, where the network learns progressively the high-order features. More precisely, each hidden layer aims to learn from the outputs of the previous layer, and it generates an output to be the input of the next layer. Generally, the weights of the parameters in each layer are learned using a learning algorithm by optimizing by optimizing an objective function.

The network output is relative to the addressed task such as classification or regression.

Several deep neural network architectures have been proposed for different tasks. In this this section, we introduce the common Deep Learning architectures.

Convolutional Neural Network Convolutional Neural Network (CNN) is one of the most effective and important DNN architectures [67]. They have been shown an impressive performance on a wide range of applications such as image analysis and recognition, voice recognition, natural language processing, and recommendation systems.

The CNN architecture is typically composed of several successive layers (Figure 3.7). There are three main types of layers that are generally observed in the CNNs architectures: Convolutional Layers (Conv), the Pooling Layer (Pool), and the Fully-Connected Layers (FCLs):

1) The Convolutional Layer: performs a specific function of transformation on local regions in the input (receipt field) to obtain a useful representation. It functions as a feature extractor. An input image is passed through a series of sliding learnable convolution kernels (filters), creating as result 3-dimensional convolution feature maps (Fmaps) (see Figure 3.4). The feature maps values are produced using the neuron activation function that can be defined as:

$$f(x) = \sum_{i=1}^{s} w_i x_i + b \tag{3.1}$$

where x_i ,w_i and b are the convolutional input values (receipt field), the weights (filter values), and the bias, respectively. s represents the filter size.

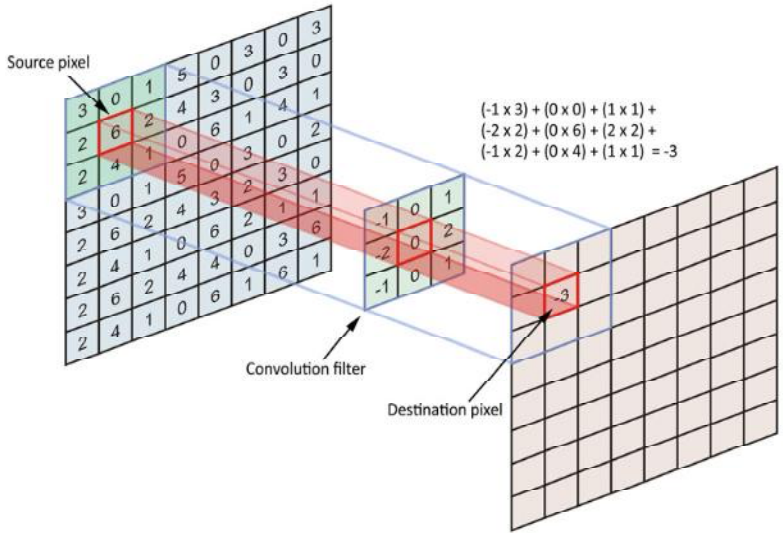


Figure 3.4: Principle of filter sliding in the convolution layer over an image [68].

In addition, a correction operation called Rectified Linear Unit (ReLU) is also applied to the obtained feature maps. ReLU is an element-wise operation defined by (Equation 3.2). The output feature maps have non-negative values.

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

2) The Pooling Layer: performs a sub-sampling operation, by shrinking the spatial dimensions (i.e the height and the width) of the intermediate feature maps and retaining the most important information. The pooling is an important concept for the CNNs since it aims to reduce the size of the feature maps in order to minimize the number of parameters and the computation operations in the network. The pooling is generally operated as a max, average, or sum function on every depth slice of the input feature maps independently. Whereas the depth dimension d is still unchanged, the height and the width dimensions of the depth slice are down-sampled using pooling filters (Figure 3.5). The output of this layer produces typically a 3-dimensional feature maps of

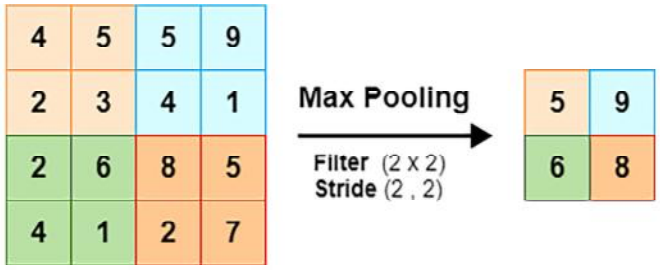


Figure 3.5: Pooling layer principle: example of performing Max pooling function [69]

the dimensions $(w \times h \times d)$ which can be defined by:

$$w = \frac{w_1 - F}{S} + 1 \tag{3.3}$$

$$h = \frac{h_1 - F}{S} + 1 \tag{3.4}$$

Were w_1, h_1, and d are the input width, height, and depth respectively, S is the stride, and F is the spatial extent. For the two pooling hyper-parameters F and S, they are commonly used in two variations: $F = 2$ and $S = 2$ as well as the overlapping pooling in which $F = 3$ and $S = 2$.

3) Fully-Connected Layers (FCLs): as its name indicates, is a feed-forward neural network in which all neurons are connected to all the neurons of the next layer and have connections with all previous layer neurons (Figure 3.6. As showed in Equation 3.1, the FCLs neuron activation function can be computed using matrix multiplication adding to bias offset.

Adding FCLs able the CNN model for end-to-end learning ([70]). More precisely, after feature generation, we need to classify the data into various classes. Thus, the obtained high-level features from the convolutional layer are fed into the FCLs structure that learns the non-linear combinations in that

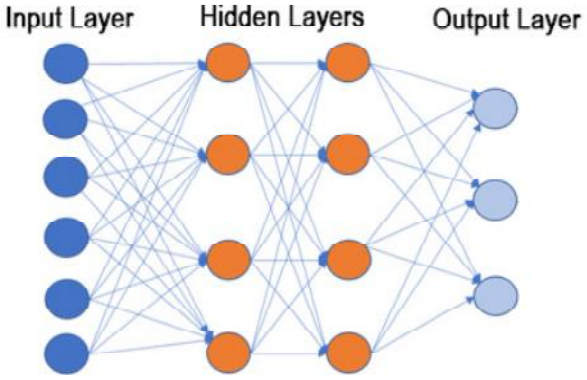


Figure 3.6: Illustration of Fully-Connected Layers structure.[69]

feature space. Over a series of back-propagation epochs, the weights of the neurons are updated progressively for optimising the loss function. Finally, the last FCL output the final classification decision.

VGG-16 Simonyan and Zisserman [71] designed the VGG-16 architecture which is a very uniform architecture consisting of 16 convolutional layers, 5 Max-pooling layers, 3 fully connected layers and a SoftMax layer at the output. A ReLU layer is provided after all hidden layers. The essential engineering of VGG-16 is introduced in Figure 3.7 of the article. Perhaps the most engaging point about this engineering is its effortlessness. An aggregate of 138 boundaries are utilized in this enormous organization design and still it is generally utilized for research works in view of its straightforwardness and consistency. Top 5 mistake of VGG-Net is 7.32% which is tremendously decreased in contrast with that of Alex-Net.

Pre-trained Convolutional Neural Network The successful results of CNNs on image recognition tasks have motivated further the research in network architecture design. Since 2012, several CNN architectures are proposed

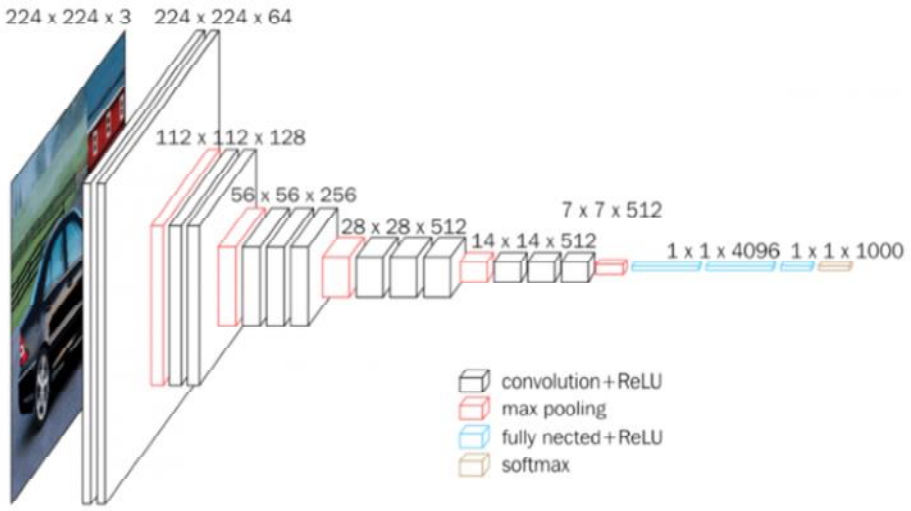


Figure 3.7: VGG-16 architecture [71]

achieving the stat-of-the-art performance on ImageNet data-set for different ILSVRC competition tasks, whereby each CNN architecture has tried to address the shortcomings of previous CNN architectures adding new structural reformulations or by exploring different strategies for parameter optimization in order to improve the CNNs performance and reduce the computational cost. Figure 3.8 summarizes the history of the CNN architectures evolution [72].

The full training of CNN models is a computationally expensive process and requires a huge amount of labelled data. Thus, several studies have examined the generalization power of the CNN architecture, demonstrating the transferability of the CNN models that are trained upon ImageNet data-set. In which these pre-trained CNN models are able to serve as the backbone for other recognition tasks on other datasets. In the literature, the most cited CNN networks are belonging to the three families: VGG, Resnet, and Inception [72].

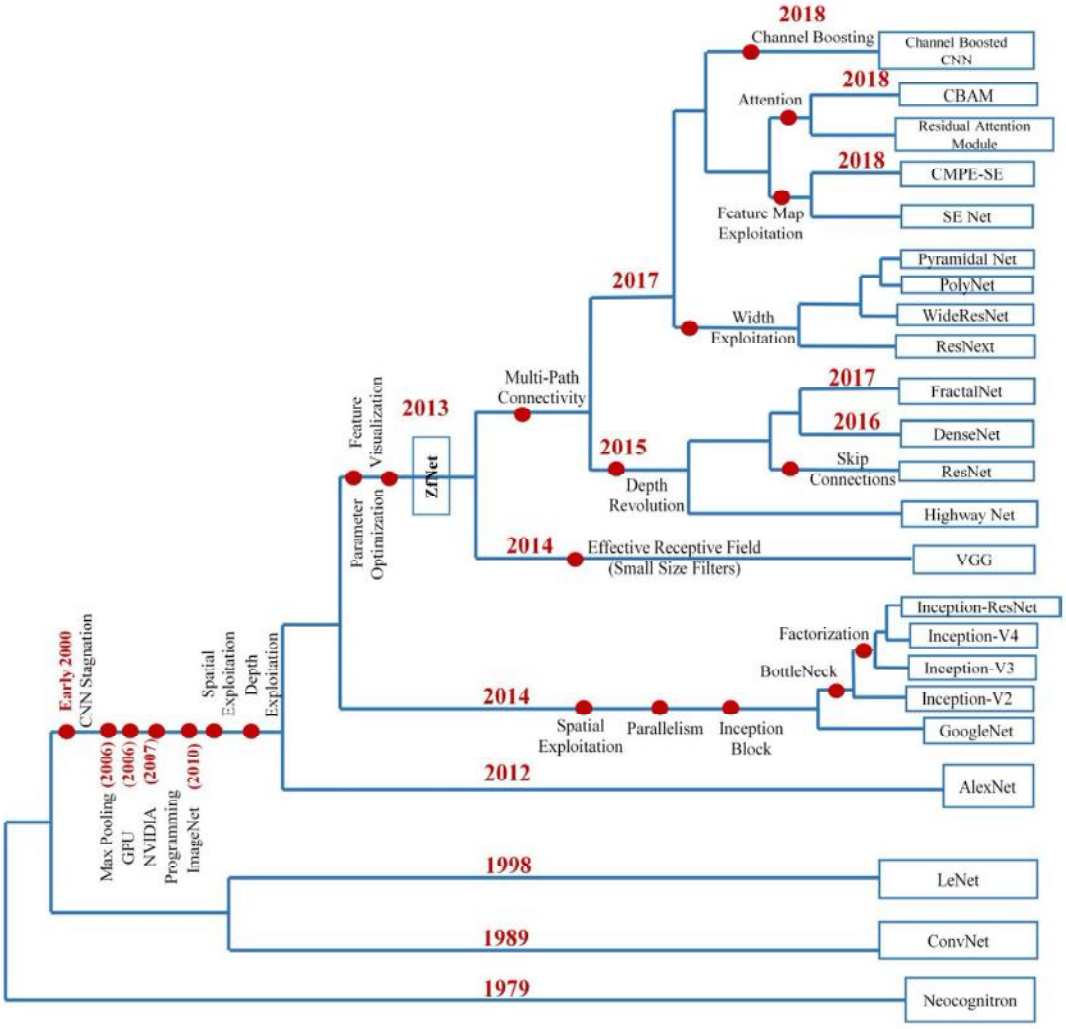


Figure 3.8: History of The CNN architectures evolution [72].

3.2.3 Decision Trees

Decision tree is a classifier in the form of a tree structure, where each node is either: Decision node specifies some test to be carried out on a single attribute-value, with one branch and sub-tree for each possible outcome of the test Leaf node indicates the value of the target attribute (class) of examples Decision trees attempt to classify a pattern through a sequence of questions. CART (Classification And Regression Trees)

3.2.4 Random forest

Combination of the bagging (random selection of examples) and random selection of features. Random Forest grows many classification trees. Each tree gives a classification, i.e., it -votes- for that class. The forest chooses the classification having the most votes (over all the trees in the forest). More details will on random forest will be presented in Chapter 4.

3.3 Recent advances on neural networks based facial age estimation

We have done an intensive examination of moste neural networks based facial age estimation methods proposed in the literature. This section gives a brief summary of the examined works in a chronological order

3.3.1 Feed-forward back propagation artificial neural network (FFBPANN)

Dehshibi et al. [73] proposed the first work of FFBPANN in face age estimations tasks in 2010. The proposed algorithm based in general on the anthropometric model, their method classifies the input images of frontal face in four age groups. Authors gathered for their proposed method a database named the Iranian Face Database. Their algorithm based on a Neural Network, the later use the computed facial features and rankles densities for the classification. Dehshipi et al. technique obtains an accuracy of 82.28%. Izhadpanahi et al. [74] designed an age classification system based on the same geometric ratio and rankles analysis but they used an SVC (support vectors classification) and

they obtain an accuracy of 92.62% better than Dehshibi [73] without using any ANN. In general classifying face images in age groups is considered a limited goal for the real word applications.

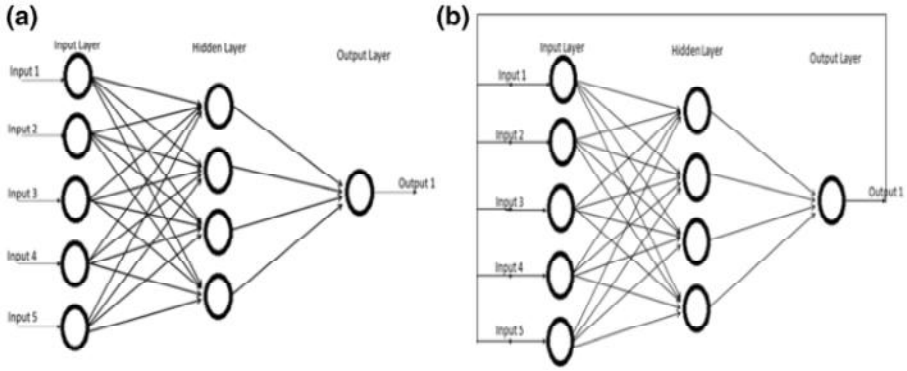


Figure 3.9: a) Feedforward ANN. b) Feedback ANN.

3.3.2 Deep learned Ageing Pattern

Deep Learned Ageing (DLA) Pattern is structure created by Wang et al. [75] for age estimation.They created six layers which make up the entire architecture of the Convolutional Neural Network (CNN). Three convolutional layers, two sub-testing (pooling) layers, one fully connected layer with Multiple Logic Perception (MLP) are utilized. Training of the CNN is trailed by features extraction. Dimensionality decrease of features is accomplished utilizing Principle Component Analysis (PCA), which improves significantly the training speed of the the model. ageing patern is made by connecting these features extracted from different layers. Authors have utilized Manifold Learning for catching the aging example and face maturing structure. They have compared the Marginal Fisher Examination (MFA) [76], Locality Sensitive Discriminant Analysis (LSDA) [77] and Orthogonal Locality Preserving Projections (OLPP) [43] in their work. Results are assessed on MORPH and FG-NET datasets.

Support vector regression (SVR) proves to be the best among all the classifcation and regression schemes. The proposed DLA pattern has likewise outperformed all the existing state-of-the-art methods. The values of MAE obtained from MORPH and FG-NET database are 4.77 and 4.26 years respectively.

3.3.3 AgeNet

In [78], Liu et al. used the score fusion of regression and classification models to handle the age estimation tasks. AgeNets are Enormous scale deep convolutional neural networks trained on age regression model and age classification depending on real-value, Gaussian label distribution respectively. A general deep transfer learning plan is likewise conveyed by the authors which comprised of two phases. This plan firstly pre-trains the multi-class face classification network and conveys it into fine-tuned apparent age estimation network. Thus, this procedure overcomes the overfitting issue. This methodology of apparent age estimation has indicated its accomplishment in ICCV-2015 *look for apparent age estimation challenge*.In a part of the evaluations, authors compared age classification and age regression, the comparison of those different transfer learning stages and the performance in the final evaluation. The eight models fusion revealed in the research paper plays out the best. For this best outcome, MAE is 3.33.

3.3.4 VGG-16 Architecture and the Fine-Tuning Methods

In such techniques researchers have ameliorated a multiple CNN models which where already trained for the sake of enhancing a better efficiency in

age estimation task using the fine tuning procedure. In fine tuning techniques the research are generally based on weights tuning based in general on weights tuning, the pre-trained CNN model weights are kept constant in some layers while the others are trained. Mostly, the layer that preserve their weight are the first layers, this latter features are generic and can be acceptable to other tasks. A more specific features that can benefit from the fine tuning approach in the last layers. For the age estimation issue, the CNN model are fine tuned on the individual age. Several works take a part in this techniques. Mali et al. in [79] used a multiple labels for image instead of using average age of the annotated face image. They employed and fine-tuned a CNN model that were based on VGG-16 architecture [71] and pre-trained on the IMDb-WIKI database [80]. In [80], authors presented another method named Deep Expectation (DEX) of apparent age to tackle the estimation of apparent age in face images. Their proposed DEX method uses also CNNs based VGG-16 architecture pre-trained on ImageNet. Also authors gathered face images with the available age in the internet, they created the moste largest age estimation database IMDb-WIKI. The proposed model was fine-tuned on the IMDb-WIKI which they created it, after that the authors fine-tuned in addition the resulting network on 20 different splits of the Chalearn LAP dataset, which is a famous competition and DEX win the first place in 2015 LAP challange [81] against 115 registered teams.

other technique under the axe of Deep Age Distribution Learning was taken a part in this field. In general, there are two types of Age Estimations: Chronological Age Estimation and Apparent age estimation.

Apparent age Estimation differs from actual chronological age in the way it is

decided. The description of each face picture is finished by numerous people and the estimation of the relative multitude of ages is taken as the right age and standard deviation is utilized to foresee the uncertainty in this age. Deep Age Distribution Learning (DADL) is the used algorithm by Huo et al. (2016) they presented their work and they utilized again the idea of multiple output CNN introduced by Malli et al. (2016). VGGFace is deployed for pre-training of deep convolutional neural network (DCNN) and for the fine-tuning of the age dataset. Instead of a single age, the authors used embraced Gaussian age distribution of a facial image as the training set. This Gaussian age distribution is created from the mean age and standard deviation. Datasets utilized for estimation are IMDB-WIKI dataset, ICCV ChaLearn taking a gander at individuals workshop 2015 and 2016 datasets. DADL technique is positioned second out of 105 contending members in track 1 of ChaLearn taking a look at People 2016: Age estimation. Deep Age Distribution Learning (DADL) is a combination of deep learning (Geng and Ji 2013) and label distribution learning (Geng 2016). Engineering of the presented profound CNN contains five convolutional layers, two fully connected and ReLU layers and one output layer. For intrigued analysts, measurements, pixel size and number of neurons for each layer are clarified in detail in the research paper. Various graphical interpretations, mean absolute error (MAE) and € error are used for evaluation and validation of the proposed study. € error and MAE result values for the proposed method are 0.1341 and 1.7569.

3.3.5 CNN for age estimation

Cascaded CNN

Age estimation utilizing proposed cascaded CNN is intended to manage unconstrained face images of Adience dataset, FG NET dataset and ICCV 2015 Chalearn challenge [80] dataset. The methodology utilized by Chen et al. [82] is finished in three stages. In the first one, a face picture is classified into an age range utilizing age group classifier. In stage two, apparent age is assessed by calculating the average age anticipated from the mean of each age gathering. This is done by utilizing apperent age regressor. In the final stage, an age correcting system is followed to address any of the anticipated age mistakes. The authors portrayed a clear pipeline of the whole methodology by explaining an example of a toy. The results indicated a better execution of this techniques compared to state-of-the-art methods. The Mean absolute error and Gaussian error was chosen as the evaluation metrics. Experiments show the value of Gaussian error as 0.297, Exact accuracy as 52.88 ± 6 and 1-OFF accuracy as 88.45 ± 2.2.

CNN for age estimation with age difference

Motivated by the issue of weakly labeled and non-labeled training samples, which can be found in social media platforms as Puniany et al. declared in [52], Hu et al.[83] ameliorated a Deep Convolutional Neural Network model suits the montioned issue. Age diference can be determined from a couple of pictures of a particular individual. This CNN is first intended for ageing dataset with age labels and afterward finally improved to work for non-labeled againg dataset calculation utilized for implanting age data is Kullback–Leibler

(K–L) divergence algorithem. Three sorts of loss functions are planned on top of the SoftMax layer. The mentioned functions are named as entropy loss, cross entropy loss and Kullback–Leibler (K–L) divergence loss [84]. CNN models attempt to decipher the right age range from the combination of the loss functions. Right age range is set apart by the single peak estimation of the probability distribution of the age classes. This model has demonstrated its benefits in real time applications to make successful age estimation of human face images with arbitrary age, arbitrary ethnicity and arbitrary poses. All the tests are done of FG-NET, MORPH and year labeled LFW [85] dataset.
The authors have also developed a dataset that contains approximately one faced images marked with the dates on which they are clicked. Each face image is marked with the people identity and timestamp. Cumulative score and mean absolute error are used as the evaluation metrics. The authors intend to exploit other soft biometric traits like gait, height and hair style for the or a similar assignment in the coming future. MAE acquired for FG-NET data set is 2.8 and that for MORPH is 2.78. Cumulative curve for the proposed technique shows significantly better outcomes than that for kNN, SVR and SVM.

3.3.6 Other facial age estimation methods

Aside from the previously mentioned different models of Neural Networks for facial age estimation proposed by researchers, there are numerous other classifications, regression and hybrid techniques used for age estimation.
Several works on facial age estimation have been proposed. The age estimation error in terms of mean absolute error (MAE) metric has been decreased by a massive margin from the appearance of this task. where, In [30], the au-

thors proposed to predict age category in images where three categories were considered: babies, adults, and senior adults. In [32], the authors used both anthropometries and texture as the main cues. The works described in [33],[34], and [35] used hand-crafted features, and achieved modest results.

In recent times, deep learning has bloomed significantly and gained popularity after being validated experimentally in a variety of fields in artificial intelligence, mainly in image recognition. Researchers have used Convolutional Neural Networks (CNN) extensively in different image-based tasks. The excellent performances in pose invariant face recognition tasks have led to its adoption in many demographic attributes estimation studies dealing with ethnicity, gender, and age estimation. In [86] the problem of age estimation through deep learning techniques was investigated. The diagnosis included three different kinds of formulations for the age estimation problem. They used the five most representative loss functions. In the work done by Huerta et al. [9] a deep learning scheme have been proposed to upgrade the state-of-the-art. A robust deep feature encoding-based discriminative model for age-invariant face recognition has been suggested in [87]. Researchers in this paper used a pre-trained Deep CNN model to extract high-level deep features. The extracted features were then encoded by learning a codebook, which converts each of the features into a discriminant S-dimensional code-word for image representation. They used canonical correlation analysis to fuse the pair of training features. For the recognition purposes, they uses a linear regression-based classifier. The authors in [88] used a multitask CNN model to extract features corresponding to attributes in images before the application of the SVM models. In other studies, an end-to-end solutions have also emerged in age estimation. For

instance, the works described in [89] and [6] relied on the use of a CNN and decision trees. Tree-based models treated as a chart-topping model due to its natural interpretability property. It is considered as a powerful method in decision tasks. In [89] presented Deep Neural Decision Forests as a novel approach, which brings a DNN representation learning functionality together with classification trees by training them in an end-to-end manner. This model differs from conventional deep networks because the final predictions are provided by a decision forest. In [6], an approach for age estimation under the name of Deep Regression Forest (DRFs) was implemented. In this endeavour, researchers connected the split nodes of a decision tree to a fully connected layer of a CNN, and dealt with heterogeneous data by jointly learning input-dependent data partitions at the split nodes and data abstractions at the leaf nodes. A new deep ranking framework for age estimation was proposed by Chen et al. in [90], in which they presented a model that included a set of basic CNNs, where each of these CNNs was initialized with the pre-trained base CNN and fine-tuned with ordinal labels. In order to provide the final age prediction, the authors aggregated the binary output of the basic CNNs. Standing on the fact that age labels are chronologically correlated, the age estimation is an ordinal learning problem. In [91], the authors has presented a method to learn feature descriptors for face representation directly from raw pixels. Their method is termed Ordinal Deep Learning approach (ODFL). In ODFL, two criteria were enforced on the descriptors, which were learned at the top of the deep networks. These criteria are: topology-preserving ordinal relation, which was used to exploit the order of information in the learned feature space and age difference cost information.

In [25], the authors have also considered age estimation as an ordinal learning problem. They exploited the label correlation among face samples in the transformed subspace. Their approach was named Label Sensitive Deep Metric Learning (LSDML) for facial age estimation. LSDML differs from the recent deep metric methods [92] and [93], which used hand-crafted feature to feed deep network, LSDML leverages deep residual network to learn series of nonlinear features transformation, where the feature similarity is smoothly sensitive to the degree of age difference.

In [94], the authors have introduced a new graphical model where age is jointly learnt with expression, in comparison to expression-independent age estimation. The proposed model aims to learn the relationship, which ties the age and the expression, by including a latent layer between the age expression's labels and features. The efforts in [27] have been focused on the attribute correlation and heterogeneity. The authors included an estimation of the multiple face attributes, in the form of deep multi-task learning approach in age estimation problem. They allowed shared feature learning among all attributes, and category-specific feature learning for heterogeneous attributes, by modeling all attributes in a single network.

Fusion strategies were considered as a popular technique in biometrics. They were used in some facial age estimation works. The basic idea is to fuse decisions or features in a hierarchical learning system. A typical example is given by the Deep EXpectation (DEX) of apparent age method [95]. The authors detected the facial images first prior to the extraction of CNN prediction from a network ensemble as a fusion method. In 2015 DEX won the apparent age Chalearn LAP competition.

Age estimation using linear binary patterns (LBP) was introduced by Gunay and Nabiyev [48]. The authors worked on FERET dataset images, and accomplished 80% precision using LBP. An assigned label by LBP operator for every pixel of the image was done by thresholding the middle pixel basing on all surrounding neighboring pixels. LBP histograms so created are utilized as pixels for classifcation. Multi-level local binary Pattern (MLBP) was clarified by Nguyen et al. [96]. MLBP gives better outcomes contrasted with LBP for human age estimation. This multi-level structure is comprised of several single level LBPs. These several LBPs have different estimations of radius, encompassing pixels and different number of sub-blocks. Each LBP histogram is finally connected together to get both local and global texture informations. Age ranking based Linear Binary Patterns (arLBP) are another type of LBP highlight which have signifcantly outperformed the traditional LBP strategy. It was tested by Onifade and Akinyemi [97] on FERET and FACE database. A reference set of numerous people was considered to make a reference image set. Age groupings in this reference image set is done based on the ages of diferent people in the reference set. At last, the age ranks are obtained from these age gatherings. Age-rank based local Binary Pattern (arLBP) features are built from this reference set. arLBP along with age ranks are utilized to anticipate the age of an individual using age estimation function LBP. MLBP and arLBP have performed very well for age estimation tasks.

Speed Up Robust Features (SURF) have additionally acquired a lot of consideration in the age estimation tasks. SURF was clarified by Bay et al. [98] in his research paper. SURF descriptor comprises of a couple steps. Initially, some interest points like T-junctions, corners and blobs are located. Secondly a feature

vector is assigned to neighborhood of every interest point. Finally, matching of these feature vectors is done between different images. SURF descriptor is remarkable in terms of its robustness, repeatability and distinctiveness. Another notable descriptor, which has been pursued for age estimation, is Histogram of Oriented Gradients (HOG) [99]. The image window is partitioned into small special regions called "cell" and afterward edge orientations are adjusted over the pixels of these cells. Contrast normalization is done by directing a small amount of local histogram energy over larger special regions called "blocks". Final normalized blocks are Histogram of Oriented Gradients (HOG).

Guo et al. [10, 100] explained linear Support vector machine (SVM) and Support Vector Regressor (SVR). Direct SVM can be utilized in the event that the quantity of training samples are extremely restricted. SVR can be used for global age prediction however it doesn't function well for exact age prediction. Support Vector Machines (SVM) is found to outperform Support Vector Regressor (SVR) for age prediction when tried regarding MAE on FG-NET information base. One reason behind this could be the varieties in the age because of different elements. Another explanation could be the tendency of SVR to find a fat curve inside a small tube. Cumulative scores which appeared with Pure SVM was lower less than pure SVR.

Table 3.1: Overview of some facial age estimation methods [52]

References	Year	Method	Databases	Contribution	Evaluation metric	Results
Xiaolong Wang et al.	2015	Deep Learned Ageing Pattern	MORPH II FG-NET	New CNN using aging features evaluations for manifold learning, classifcation and regression approaches	MAE Cumulative score (CS)	FG-NET 4.26 MORPH 4.77
Van Huerta et al.4	2015	CNN model by Ivan Huerta et al.	MORPH FRGC v2.0	Examination on texture and appearance-based descriptors and their fusion investigation of deep learning schemes for age estimation Comparisons with state-of-the-art studies on 2 large databases	MAE CS	One scale LBP 6.66 Three scale LBP 6.13 Single scale SURF 6.09 Multiscale SURF 5.59 Fusion of HOG, LBP and SURF 4.27 Cumulative score 71.2% on fusing features (best)
Zengwei Huo et al.	2016	Fine-tuned CNN based on VGG-16	IMDB-WIKI Chalearn LAP 2015 Chalearn LAP 2016	Deep age distribution model. VGG face based deep CNN for age estimation.	MAE	ChaLearn LAP 2016 MAE is 3.1182

Continued on Next Page...

Table 3.1 – Continued

References	Year	Method	Databases	Contribution	Evaluation metric	Results
Zhenzhen Hu et al	2016	CNN for age estimation with age difference	FG-NET MORPH Year labelled LFW	Age estimation by adopting age diference Study on K-L divergence loss, Ertropy loss and Cross-entropy loss A new face dataset.	MAE CS curves	MAE on FG-NET 2.8, MORPH 2.78 and Year labelled datasets 2.78
Shixing Chen et al.	2017	CNN for age estimation with age difference	MORPH-II	Age estimation using Ranking CNN model Announcement of a much tighter error bound for age ranking Proof that Ranking CNN gets smaller estimation errors	Accuracy MAE Cumulative score T-set outcomes	MAE Ranking CNN 2.96

Continued on Next Page...

Table 3.1 – Continued

References	Year	Method	Databases	Contribution	Evaluation metric	Results
Kai Li et al.	2017	D2C for age estimation	MORPH-2 WebFace	Novel cumulative hidden layer to improve age estimation and mitigate sample imbalance problem Novel comparative ranking layer to improve age estimation by aging feature learning Testing on two large datasets.	MAE	Average MAE of 3.16 is achieved by Network using Cumulative Hidden Layer on Morph-II database Average MAE of 6.12 is achieved by Network using Cumulative Hidden Layer on WebFace database MAE results using T values 6, 5 and 4 are 3.33, 3.37 and 3.50 respectively on Morph-II dataset MAE results using T values 6, 5 and 4 are 6.39, 6.50 and 6.70 respectively on WebFace dataset

Continued on Next Page...

Table 3.1 – Continued

References	Year	Method	Databases	Contribution	Evaluation metric	Results
S Taheri O Toygar	2018	CNN for Fusion based Multi-stage age estimation	FG-NET MORPH II	Combining of locally handcrafted features with multi-scaled learned features Features-level fusion for handcrafted features and Score-level fusion for multi-scale features Comparison with the results obtained on in-the-wild AgeDB database	MAE CS	MAE MORPH II 3.17 MAE FG-NET 3.29
Zeng at al.	2019	CNN Resnet32 model Soft-Ranking	MORPH II AgeDB	A novel age encoding method (Soft Ranking). that simultaneously encodes both ordinal informationand the correlation between adjacent ages	MAE	Soft-Ranking MAE MORPH II 2.89 MAE AgeDB 5.74

Continued on Next Page...

Table 3.1 – Continued

References	Year	Method	Databases	Contribution	Evaluation metric	Results
Guchairia et al.	2020	cascade trees ensembles	MORPH,LFW+ FG-NET,FACES PAL,APPA-REAL	A novel proposed scheme based on trees ensembles named (DRF) Deep Random Forest, the scheme suits the problem of age estimation, it compose of two principle part a DRF to extends featur vector and anotrherone for the last prediction after a fusion process on the extended feature vectors	MAE	MORPH Caucasian 3.67 FG-NET 3.65 FACES Avg 1.23 PAL 2.73 APPA-REAL Real age 5.25 APPA-REAL Apparent age 3.36

Chapter 4

Feature Fusion Via Deep Random Forest for Facial Age Estimation

4.1 General introduction

In this chaper, we propose a new architecture for age estimation based on facial images. It is mainly based on a cascade of classification trees ensembles, which are known recently as a Deep Random Forest. Our architecture is composed of two types of DRF. The first type extends and enhances the feature representation of a given facial descriptor. The second type operates on the fused form of all enhanced representations in order to provide a prediction for the age while taking into account the fuzziness property of the human age. While the proposed methodology is able to work with all kinds of image features, the face descriptors adopted in this work used off-the-shelf deep features allowing to retain both the rich deep features and the powerful enhancement

and decision provided by the proposed architecture. Experiments conducted on six public databases prove the superiority of the proposed architecture over other state-of-the-art methods.

4.2 Introduction

Convinced by the advantages of the classification and regression trees, Zhou et al. [12] have proposed a new approach (named gcForest) for image classification. Their method consists of a decision tree ensemble arranged in a cascade of layers where each layer (level) is composed of several random forests. The resulting performance is highly competitive to that of deep neural networks in a broad range of classification tasks [12].

The differences between our proposed method and the existing works are that we use deep features from pre-trained models as input features, and we integrate the paradigm of deep learning that is similar to the deep neural networks, by cascading a simple machine learning tool that is provided by Random Forests (RF).

The work described in [12] was dedicated to generic classification problems. The proposed classifier is a cascade of tree ensembles.

The major differences between the approach of [12] and the CNN methods concern the hyper-parameters and the training process. Indeed, the approach of [12] needs much fewer hyper-parameters than deep neural networks, and its model complexity can be automatically determined in a data-dependent way. The applications shown in [12] targeted classical recognition problems.

The main similarity between our work and the work described in [12] regards the use of Random Forests that generate feature representations. But,

our proposed approach contains several novel modules which are different from [12]. These are as follows: the use of different deep feature vectors as input, the use of a mid-level fusion module, and the targeted application (facial age estimation), which is different from the classic classification tasks. In detail, we were inspired by the main idea of [12] to create random forest ensembles with a cascade structure. However, this structure will be used twice in our proposed architecture. First, it is used for encoding and fusing the individual input features (i.e., generating a fused-representation). Second, the proposed architecture exploits the generated vector (fused-representation) for the final decision. More importantly, the work in [12] does not contain a fusion module, and its input is composed of raw brightness of a sliding window (Multi-Grained-scanning). Random forest ensembles with cascade structures are then used for the final classification.

Table 4.1 summarizes the similarities and differences between our proposed method and the one presented in [12] .

Table 4.1: A comparison between our work and the method in [12].

Phases	Method in [19]	Our method
Multi-Grained Scanning of raw images	✓	×
Ensembles of random forests	✓	✓
Cascade structure using random forests	✓	✓
Fusion representations using random forest ensembles.	×	✓
Final decision based on the max probabilities class	✓	✓
Final decision using the average of first largest probabilities	×	✓
Problem tackled	Classic classification	Facial age estimation

It is worth noting that our proposed method and the work presented in [6] are not similar. In [6], the authors proposed a method where the split nodes of a regression tree are directly linked to a fully connected layer of a convolution

neural network.

In their work, DRF refers to Deep Regression Forest and the deep concept is tied to the use of deep Convolutional Neural Networks.

they used trees model conditional probability over the ages where each leaf node can store a given trainable probability distribution.

They described how to learn a single differentiable regression tree. Also, they described how to learn an ensemble of trees to form a forest. In our work, for facial image features, we use adequate deep features (retrieved via some pre-trained CNNs) and integrate the deep concept by deploying a cascade of classic Random Forests ensembles. Furthermore, our architecture allows the fusion of different types of features. We created many RFs with different settings. Those random forests (the ensemble) will be constructed layer by layer (level). The training in [6] alternates between learning a CNN and learning a set of differentiable trees, which increases the computational complexity of the algorithm.

Moreover, the major difference between our proposed method and the methods that use Deep CNNs (e.g., [6],[91],[25],[95]) is the training time complexity. Indeed, our proposed method has less computational cost than that of the CNN based approaches.

4.3 Review of Deep Random Forest

Some recent works used a Probabilistic Random Forest to tackle the age estimation problem [6]. They showed some interesting results. The method based on Random Forests RF [12] was considered as a good competitor to Deep

Neural Networks for classic classification problems. Deep Random Forests have many advantages such as the implementation simplicity and the reasonable time complexity associated with the training phase. These have encouraged us to explore and to adopt this strategy to the facial age estimation problem. To the best of our knowledge, our work, which is partially inspired by the idea presented in [12], is the first work that addresses the age estimation task using deep Random Forests.

The specific characteristic that makes RF suitable for such applications is the reasonable cost of training and the robustness to over-fitting. Besides, RF has the advantage that it's few parameters are easy to set. We can use many RFs with different parameters. Diversity enhances the final performance. We can create a cascade structure with RF to create more layers like in deep neural networks where each layer can produce a piece of different information.

4.3.1 Random Forest

Random Forest (RF) is a method that provides predictive models for classification and regression operations. RF uses binary decision trees that include CART trees proposed by Breiman et al. [101]. The general idea behind the RF method is to generate several predictors before pooling their different predictions instead of trying to get an optimized procedure at once (see Figure 4.1). More details about Random Forests can be found in [102].

4.3.2 Deep Random Forest for classification

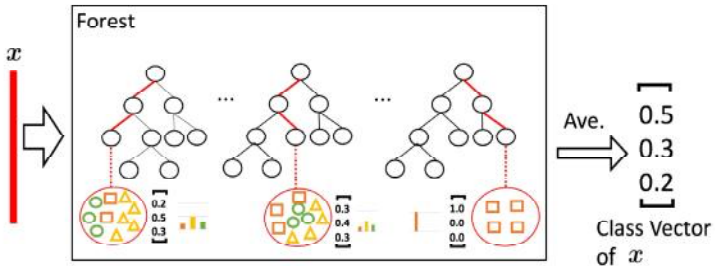


Figure 4.1: Illustration of Random Forest classifier. Each class vector is generated by counting the percentage of different classes of training examples at the leaf node where the concerned instance falls and then averaging across all trees in the same forest [12].

Relying on the advantages of random forest, Zhou et al. [12] have introduced a new approach that included many ensembles of random forests. By creating more than one level, the ensembles of random forests act as a cascade structure. In this work, we will not distinguish between "level" and "layer". In this structure, each level is composed of an ensemble of random forests as illustrated in Figure 5.2.

This structure was partially inspired by the layer-by-layer (or level-by-level) processing of a learning representation in the deep neural networks. Each level (or layer) of the Deep Random Forest is an ensemble of forests, precisely an ensemble of decision trees ensembles. The first level receives the feature vector as a given input, each forest of the same level will generate a class probability distribution as in Figure 5.2. Suppose there are C classes to predict, then a C-dimensional class vector will be the output of every single forest. The input vector for the next levels is obtained by concatenating the original input vector with the generated class vectors of each forest (resulting from the previous level). The dimension of the representation vector will be given by Eq. (4.1).

$$Dim = D + F \times C \tag{4.1}$$

where:

- D: the original feature size.
- F: the number of forests.
- C: the number of classes.

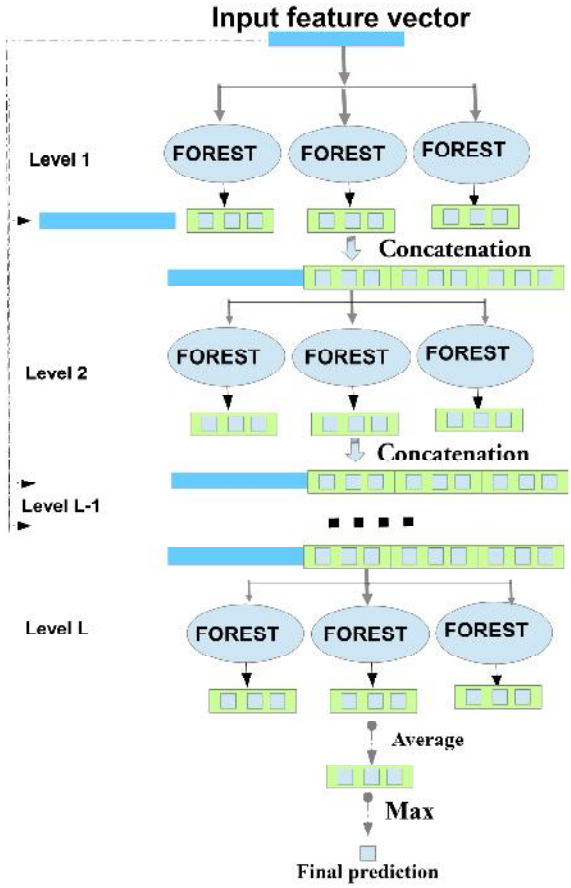


Figure 4.2: Illustration of the deep random forest (DRF) structure where each level of the cascade receives feature information processed by its preceding level and outputs its processing results to the next level. Assume that each level of the cascade consists of three forests, and that there are three classes to predict. Thus, each forest will output a three-dimensional class vector, which is then concatenated for re-representation of the original input.

In the L-level (the last level), the RF generated class vectors will be averaged via arithmetic mean to produce the final class vector, the max value index of which, will be the prediction class.

$$FinalClass = \frac{1}{F}\sum_{f-1}^{F} Class(f) \tag{4.2}$$

where:

- *FinalClass*: the final probabilities class vector.
- *Class (f)*: the probabilities class vectors of a single forest f.
- F: the number of forests.

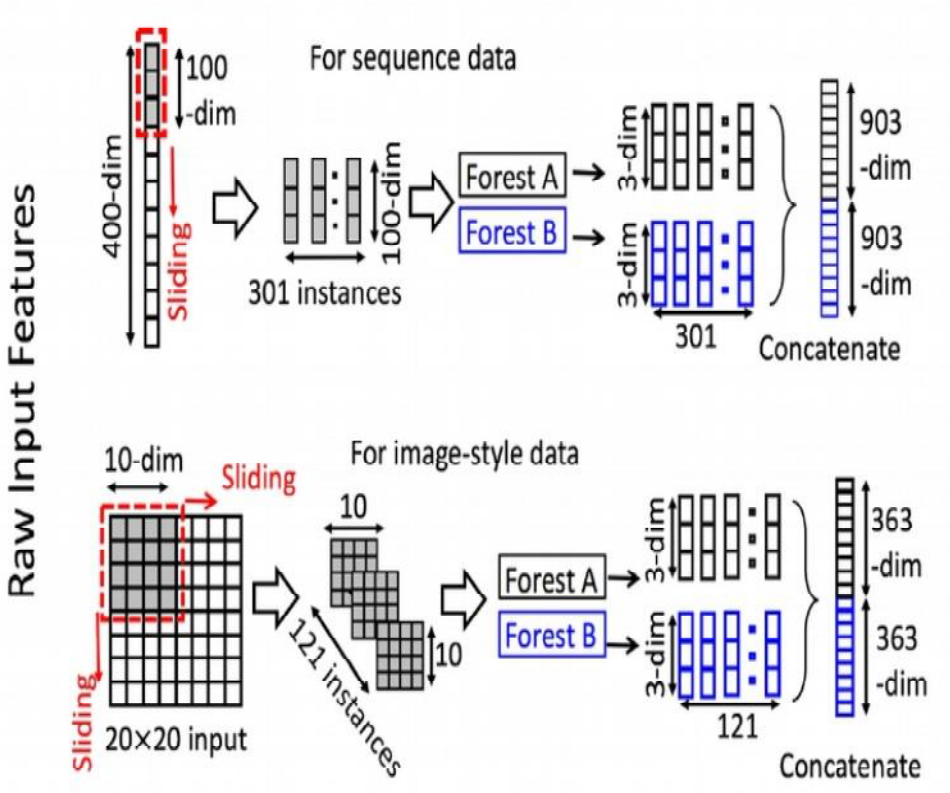


Figure 4.3: Illustration of Multi-grained scanning in both case sequence data and image style data [12].

The capability of treating feature relationships using Deep neural networks encouraged the authors in [12] to create a procedure for feature re-representation. This procedure aims to replace the convolution operation applied to the pixels of an image. It also aims to enhance the cascade forest and is named Multi-grained scanning (MGS) (see Figure 4.3). The MGS uses a sliding window to scan the raw feature of sequence data or image style data, and it creates an ensemble of instances that have the same scan window size, those instances will be used to train two different types of forests to generate a class vector (for each) as elucidated before. The resulting class vectors will be concatenated to be transformed features.

4.4 Proposed approach

The DRF introduced in [12] was proposed and used for the classic recognition and classification tasks. It was used for identity discrimination and object identification. The method in [12] inspired us to develop an approach that will be applied to the problem of age estimation. The original method might face some difficulties if the human age nature is not taken into account. Indeed, the human age follows a uni-modal distribution, and the associated classes (if each year is considered as a class) can be fuzzy. We included an arithmetic function to improve the original final decision. This function influences the final decision to be more suitable to the nature of the human age. Although the DRF has proved its good results [103, 104] in classic recognition problems, we think that there is still a room for better results. The estimation can be improved through the enhancement of either the prediction criteria, the initial input features, and the intermediate fusion scheme. We propose a method that uses all these three items and applies the resulting architecture to the problem of age estimation. Figure 5 illustrates the overall architecture of our proposed model. This architecture is composed of two principal parts in addition to the pre-processing and feature extraction phases. The first part is represented by several individual DRF whose output is fused and handed out to the second part represented in the bloc fd-DRF (final decision-Deep Random Forest).

First, a process that serves as an enrichment of the initial input vectors should be added. We propose a different use of the original DRF that aims, in addition to final class prediction, to extract a vector through the previously explained concatenation of the original input vector with the random forest generated class vectors of a chosen level. The resulting vector will be larger and richer in

information than the original one.

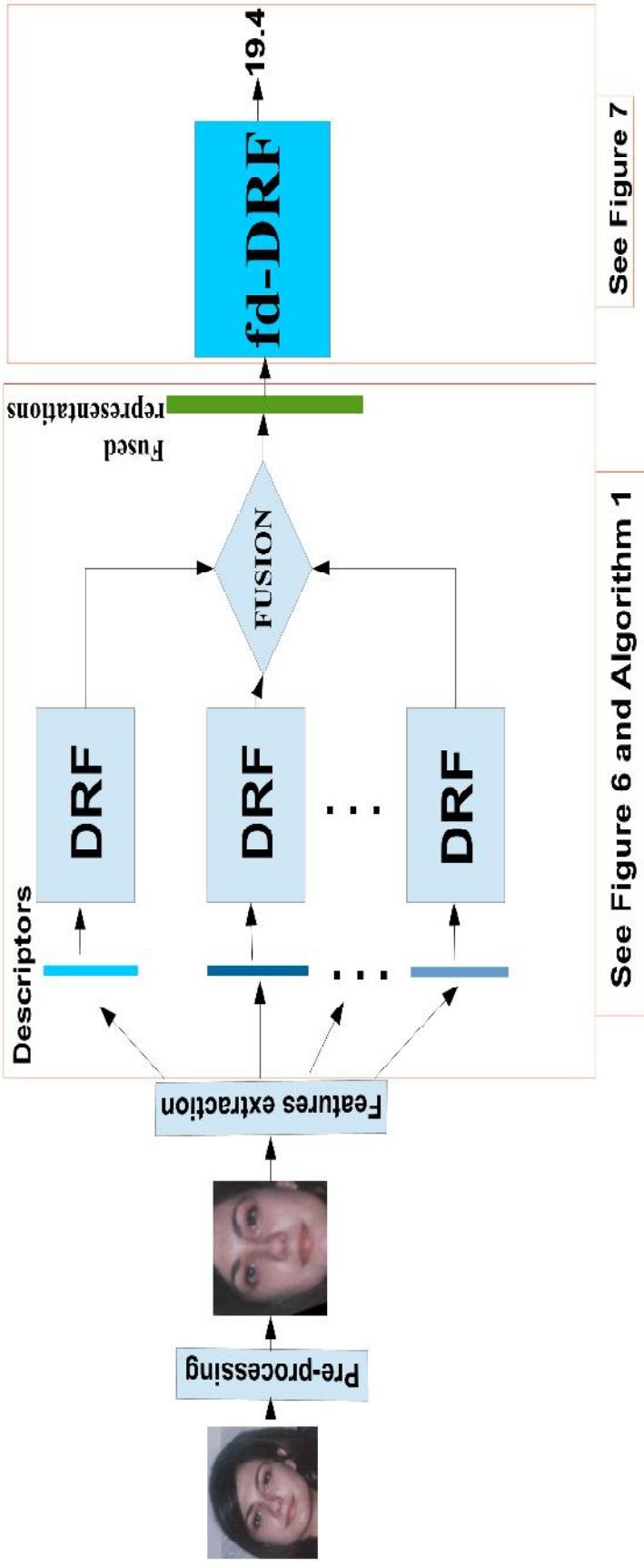


Figure 4.4: Illustration of the overall architecture of our method.

To enrich the input feature vector, even more, we took advantage of the efficiency of the feature fusion, which is considered as a popular technique in biometric that attracted the focus of researchers. We opted to fuse feature vectors obtained from several original descriptors in the DRF architecture arithmetically. Thus, for each type of original features, one DRF is designed to process it and to produce its DRF representation. Those original descriptors can be of any type: hand crafted, deep features, and scanned windows like in gcForest [12]. Algorithm 1 and Figure 6 show how the new input vector is computed.

$$Fused - representation = \frac{1}{V}\sum_{v=1}^{V}\mathbf{DRFF}(v) \tag{4.3}$$

where:

- **DRFF**(v) is the output of each individual DRF.
- *Fused-representation* is the output of the DRF-Fusion.
- V is the number of original input feature vectors.

The proposed fusion (average of all DRF representations) assumes that the original features input in DRF has the same dimension. To overcome the case of different sizes of input vectors, we use zero paddings for the shortest vectors before the averaging process. The proposed fusion scheme aims to provide an averaged input vector for the final prediction process to minimize the influence of extreme values.

Algorithm 1 takes as input the face descriptor vectors $\mathbf{FV}(v)(v = 1, ..., V)$ (the feature vectors), the number of cascade levels (layers) L (level or a layer contains several random forests), the number of input feature vectors V and the number of forests F in a given level. For the first level or if the number

of levels is set to one, the input feature vectors to this level are the original feature vectors, else it will be another vector generated by the previous level as follows: Each forest $f \in 1, ..., F$ in the current level $level \in 1, ..., L$ will generate a probabilities class vector; those class vectors will be concatenated with the original input vector, as illustrated in Figure 4.5. Each original feature vector has been encoded using the deep Random Forests. We call this generated code **DRFF**. The V **DRFF** representations will be fused using the arithmetic average. Other fusion schemes can be adopted. In our work, we have tested two fusion schemes: the average and the concatenation. We have found that the average fusion has provided almost the same performance that is obtained with the concatenation (see section 5.3.4.), yet the average scheme provided much more compact representations.

Algorithm 1 DRF-Fusion

Input:
Face descriptors: $\mathbf{FV}_1, \mathbf{FV}_2, ..., \mathbf{FV}_V$;

Number of input feature vectors V;
Number of levels L;
Number of forests F.

Output:
Fusion: **Fused-representation**

For $v = 1 : V$

- **For** *level*=1 : L
  if (*level*=1) **input**=**FV**(v)
  Else input=current-input
  End if
  - **For** each *FOREST* in *level* $f = 1 : F$
    class (f) =generate-class-probabilities (*FOREST* (f), **input**)
    End for

  current-input = concatenate (**FV**(v), **class**(1), ..., **class**(F))
  End for

- **DRFF**(v)= **current-input**

End For

Fused-representation = $\frac{1}{V}\sum_{v=1}^{V}$ **DRFF**(v)

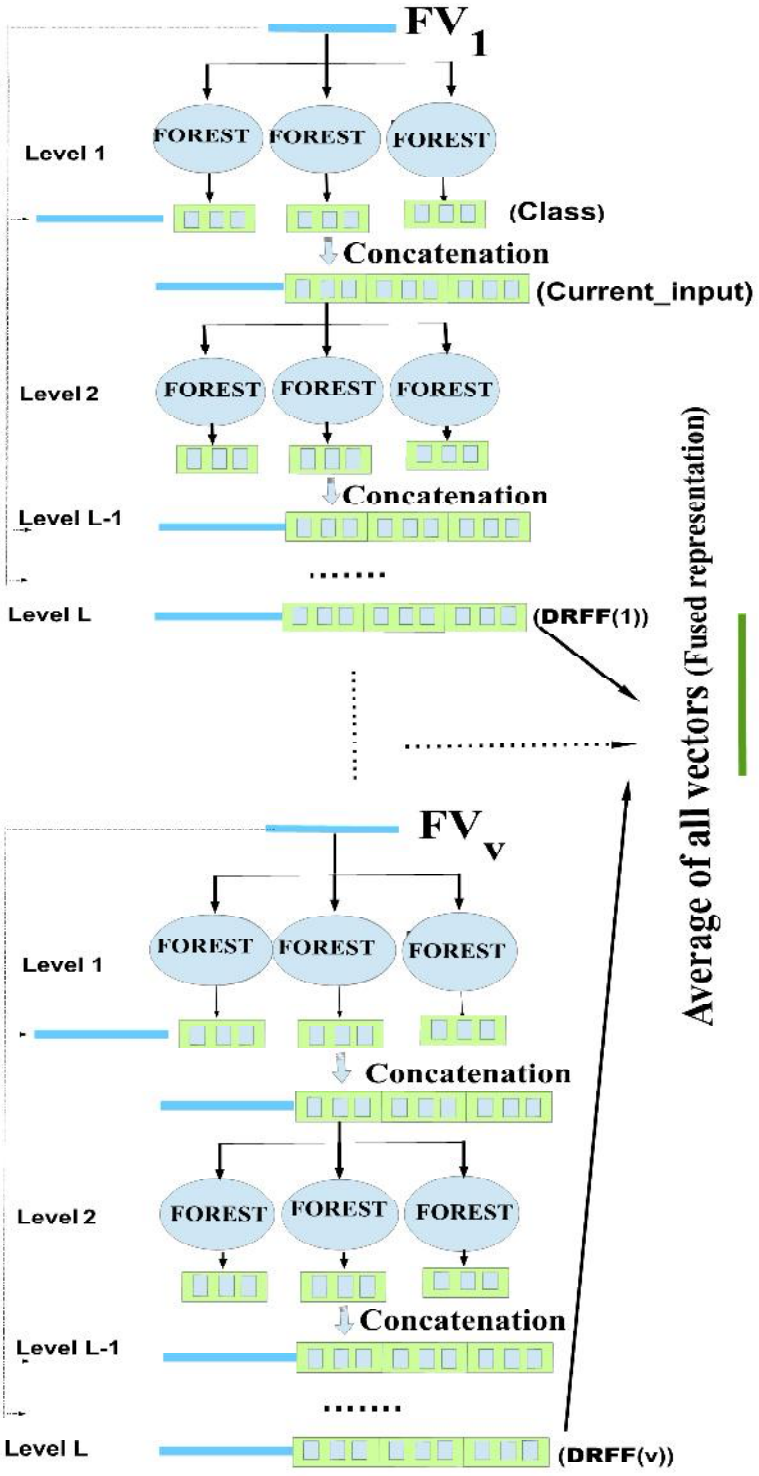


Figure 4.5: An illustration of the DRF-Fusion scheme. Many DRFs with various input feature vectors are used to produce richer representations, which are later fused to obtain the Fused-representation.

Second, another process that enhances the prediction criteria is also integrated and is called fd-DRF. This one is similar to the cascade structure presented in the section of Random forest only in the last level (level L) which contains the final class vectors probabilities. We propose two ways for the final decision (Ages having the largest probability and Ages with N_{max} highest probabilities). Ages with N_{max} highest probabilities considered as new decision function , distinguished from the original work. Instead of picking the age having the highest probability, the new decision function takes in consideration the other ages having high probability (chosen in descending order) values. The new decision function uses the N_{max} probabilities and their associated ages to produce the final age prediction (The mathematical process is the arithmetic mean of the N_{max} ages having the highest probabilities where N_{max} is a given parameter.). Figure 7 illustrates the fd-DRF.

4.5 Experiments

In this section, we will present the details of the algorithm implementation. We also provide a comparison against other similar studies. Our implementation contains many parts in which our main goal is to test various methods on a few given feature vectors. This allows us to assess the performance of the proposed model. We used the original Deep Random forest algorithm. We then compare its results with both the proposed method and the SVM classifier after the fusion phase. More explanation will be provided in the following subsections.

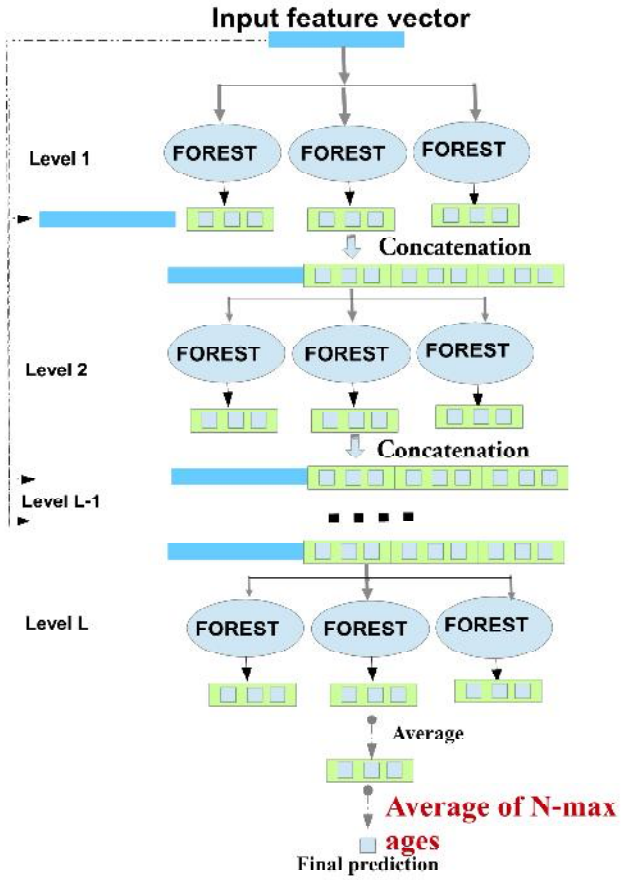


Figure 4.6: Illustration of the final decision method fd-DRF.

4.5.1 Implementation Details

Preprocessing

In this work, we localized the facial landmarks using the Ensemble of Regression Trees (ERT) algorithm [105] which is a robust and very efficient algorithm for facial landmarks localization. Facial landmarks help us to get eyes coordination, building on those points, we applied the face alignment, which is considered an important step in image-based age estimation. After performing the alignment, the face region should be cropped (aligned face).

Face Features Extraction

For face feature, we use the Deep Expectation (DEX) –Chalearn ICCV2015. DEX-Chalearn ICCV2015 is the winner (1st place) of the Chalearn LAP 2015 challenge on apparent age estimation. More than 500,000 images of celebrities from IMDb and Wikipedia labeled with age were assembled by authors of DEX-Chalearn to fine-tune the VGG-16 architecture used in DEX, the VGG-16 pre-trained on ImageNet for image classification. DEX-Chalearn is a powerful deep learning model for age estimation. It provides tools to generate deep features suitable for age characteristics due to the large data used to fine-tune it. We extract the last two fully connected layer vectors of DEX-Chalearn pre-trained model FC6 and FC7 of the input preprocessed images with a size of 224×224. The vectors FC6 and FC7 are later considered as the input features to the proposed architecture.

Parameter Setting

Each level of the Deep Random Forest (DRF) (in both DRF-Fusion and fd-DRF) contains 10 forests. To encourage diversity, we used two types of forests. Thus, we used 5 completely random trees forests and 5 random forests. For both types, the five forests contain 500, 1000, 1500, 2000, 2500 trees, respectively. Selecting a feature at each tree node was randomly generated.

Evaluation Metric

To evaluate the performance of the proposed age estimation method, we used the Mean Absolute Error (MAE). It is one of the most known indicators for age estimator performance evaluation in literature. MAE calculates the

average of absolute error between the predicted and the ground truth ages. It is given by:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|p_t - g_t|, \tag{4.4}$$

where n is the number of tested images, p_t is the predicted age of image t, and g_t is the ground-truth age of this image.

4.5.2 Experimental Results

Performance evaluation

In this section, we will quantify the performance of the proposed method as a function of different factors. These include: (i) fused and non-fused features, (ii) type of features, (iii) number of layers, and (iv) number of highest probabilities.

Table 4.2: MAE (years) obtained with two different hand-crafted features (HOG and LBP) using DRF on the MORPH Caucasian dataset. We used L_2 normalization for LBP vector in the fusion part.

DescriptorNumber of layers	1 layer	2 layers
LBP	7.56	7.98
HOG	5.20	4.93
LBP+HOG	6.11	5.23
LBP+HOG+FC6	5.24	5.13

Table 4.3: MAE obtained with two different hand-crafted feature using the DRF method with $N_{max} - 5$ (N_{max} is the number of the ages having the highest probabilities) of MORPH Caucasian dataset. We used L_2 normalization for LBP vector in the fusion part.

DescriptorNumber of layers	1 layer	2 layers
LBP	6.23	6.12
HOG	5.83	5.27
LBP+HOG	5.12	5.98
LBP+HOG+FC6	4.86	5.37

We have used the hand-crafted features (LBP and HOG) as presented in Tables 4.2 and 4.3 with the MORPH Caucasian dataset. We relied on the same fusion strategy presented in our work. Tables 4.2 and 4.3 show the results of LBP, HOG, a fusion of HOG and LBP (LBP+HOG) and finally the fusion of HOG, LBP and FC6 feature vectors. The results presented in Table 4.2 correspond to the use of the age associated with the largest probability. The results in Table 4.3 were obtained using the mean of the five ages associated with the largest probabilities. The third row presents the fusion of the two hand-crafted features (LBP+HOG). In the fourth row, we included the deep feature vector FC6 of the DEX-Chalearn pre-trained model. The use of the deep feature aims to demonstrate its impact on the results when performing such tasks. As it can be seen, the MAEs depicted in Tables 4.2 and 4.3 are different. This is not surprising since the image feature type and the number of layers all affect the final performance. We remind that the proposed architecture is composed of two modules (fusing process and final decision process) that cannot be separated. The depicted architecture in Figure 5 presents the general case where we have more than one type of features. The fusion module has several DRFs whose output are fused. The resulting fused vector feeds another DRF (named fd-DRF) for the final decision (see Figure 4.4).

Table 4.4: MAE (years) obtained by the proposed architecture on seven datasets.

DescriptorDatabases	FG-NET	PAL	MORPH Caucasian	LFW+	APPA-REAL Real Age	APPA-REAL Apparent Age	FACES
FC6	3.80	3.54	4.50	6.00	5.25	3.11	1.49
FC7	4.00	4.80	4.26	6.16	5.71	3.60	2.77
Fused-representation1	3.77	3.07	4.07	5.99	5.39	3.47	1.35
FC6	3.84	3.68	5.80	6.12	5.96	3.12	1.30
FC7	4.20	4.71	6.66	6.46	6.30	3.16	2.67
Fused-representation2	3.90	3.09	6.11	6.11	6.52	3.57	1.85

Table 4.5: MAE (years) obtained by the proposed architecture on the FACES dataset.

DescriptorFace Expression	Neutrality	Happiness	Disgust	Fear	Sadness	Angry	Average
FC6	1.10	1.26	1.92	1.52	1.34	1.82	1.49
FC7	2.21	2.54	3.38	2.87	2.65	3.01	2.77
Fused-representation1	0.86	1.15	1.73	1.47	1.18	1.71	1.35
FC6	0.90	1.14	1.71	1.29	1.16	1.60	1.30
FC7	2.11	2.43	3.25	2.83	2.51	2.91	2.67
Fused-representation2	0.88	1.69	2.63	2.10	1.51	2.29	1.85

We emphasize that the presented comparisons in Tables 4.2 and 4.3 aim at studying several cases (fused features vs. individual features) as well as several types of features. The presented comparison aims to observe the advantages of the fusion process itself in the final stage (fd-DRF). This comparison elucidates what we can gain with such fusion processes. In some cases in which the decision is based on the highest probability (see Table 4.2), the fusion has not given an MAE that is better than the best one obtained by the individual features. This is the case where the LBP and HOG descriptors were used. The explanation can be as follows. Since the used LBP descriptor is not very relevant to the problem of age estimation, its fusion with HOG and FC6 features was not able to get a better result than what can be obtained by HOG alone. On the other hand, when the decision is based on the use of the highest probabilities (Table 4.3), the fusion of LBP and HOG gave better results than that of the individual features. Moreover, as it will be shown in Table 4.4, the fusion

scheme of FC6 and FC7 features has not given the best results for some datasets. Based on the above observations, we can see clearly that the performance of the fusion depends on many factors that include the image feature type, the number of layers, the decision scheme, and the dataset. Thus, future work would investigate the fusion of many types of features as well as automatic feature weighting.

We have used the deep features FC6 and FC7 as input vectors to the DRFs (first part of Figure 4.4). We perform two groups of experiments. In the first group, the DRF adopts one layer. In the second group, the DRF adopts two layers. The final output vector of this process is named $Fused-representation1$ when using DRF with one layer and $Fused-representation2$ when using the output DRF with two layers.

For the representations $Fused-representation1$ and $Fused-representation2$, we evaluated two solutions: (i) the first one is given by the DRF (i.e., the predicted age is estimated by the full architecture of Figure 4.4), the second one uses the SVM multi-class classifier which is applied on the representation generated by the DRF module. We emphasize that, for the individual features, the SVM is applied to the output of the DRF module.

We compare them with the original FC6 and FC7, which allowed us to evaluate the possible benefits offered by the fused representations generated by the proposed architecture.

Tables 4.4 and 4.5 summarize the results obtained with the proposed architecture using the deep features FC6 and FC7. Table 4.4 contains all used datasets and

Table 4.6: MAE (years) obtained by the proposed architecture (without the fd-DRF) and the SVM multi class classification on seven datasets.

DescriptorDatabases	FG-NET	PAL	MORPH Caucasian	LFW+	APPA-REAL Real Age	APPA-REAL Apparent Age	FACES
FC6	3.80	3.54	4.50	6.00	5.25	3.11	1.49
FC7	4.00	4.80	4.26	6.16	5.71	3.60	2.77
Fused-representation1	4.67	3.11	4.99	8.02	6.89	4.30	1.08
FC6	3.84	3.68	5.80	6.12	5.96	3.12	1.30
FC7	4.20	4.71	6.66	6.46	6.30	3.16	2.67
Fused-representation2	4.33	2.99	4.05	5.95	5.31	3.20	1.03

Table 4.7: MAE (years) obtained by the proposed architecture (without the fd-DRF) and the SVM multi class classification on the FACES dataset.

DescriptorFace expression	Neutrality	Happiness	Disgust	Fear	Sadness	Angry	Average
FC6	1.10	1.26	1.92	1.52	1.34	1.82	1.49
FC7	2.21	2.54	3.38	2.87	2.65	3.01	2.77
Fused-representation1	0.90	1.07	1.25	1.17	1.05	1.09	1.08
FC6	0.90	1.14	1.71	1.29	1.16	1.60	1.30
FC7	2.11	2.43	3.25	2.83	2.51	2.91	2.67
Fused-representation2	0.85	1.01	1.20	1.06	0.96	1.15	1.03

Table 4.5 presents the detailed results on the FACES dataset with all facial expressions. The first three rows of each table present the results obtained with DRFs adopting one layer (one level), where the remaining three rows present results obtained with DRFs adopting two layers (2 levels) . In those two tables, we can see that the best results were obtained by $Fused - representation1$.

Tables 4.6 and 4.7 summarizes the results obtained by the SVM multi-class classifier. For the five datasets, the use of SVM with $Fused - representation2$ gives better results than the other representations. The SVM classifier with $Fused - representation2$ gives more accurate results than the SVM classifier that used the DRF representation of the individual FC6 or FC7 except for the FG-NET database.

Using SVM with the fused representations (provided by the first part of the proposed architecture) can reduce the final MAE in particular when two layers are used. This demonstrates the efficiency of the fusion method.

Tables 4.4 and 4.6 summarize the results of three types of comparisons:

(i) individual feature vs. fused features; (ii) one layer vs. two layers for the individual DRFs, and (iii) SVM classifier on fused representations versus the proposed architecture.

The results have shown that whenever SVM is used the fusion has not improved the results compared with individual features (in particular in the case of one layer). On the other hand, when the proposed architecture is used, the fusion scheme adopting one layer for the individual DRFs has improved the performance with respect to the individual features.

Actually, the effectiveness of DRF depends on the number of layers. There is no evidence that by increasing the number of layers in the individual DRFs the final performance would necessarily increase. As in the original work that proposed the DRF for object recognition and classification the number of layers should be determined by a cross-validation scheme, and adopt the one that provides the best performance.

Thus, it is normal that results can be influenced by the number of layers and by the final classifier that output the predicted age (SVM or DRF). We recall that our method that we compared its results with the state of the art results (Table 4.14) is the fd-DRF (averaged predictions of several ages).

In the remainder of this section, we will present the results of the proposed architecture when the predicted age is set to the mean of ages having N_{max} highest probabilities. We used the full proposed architecture with $Fused-representation1$ and $Fused-representation2$. The final age prediction is given as the average of N_{max} ages that have the N_{max} highest probabilities in the final output as explained in Figure 4.6. We studied the effect of several values of N_{max}. Tables 4.8 and 4.9 illustrate the MAEs obtained by the DRF estimator on

the vector $Fused-representation1$. Tables 4.10 and 4.11 illustrate the MAEs obtained by the fd-DRF estimator on the vector $Fused\ \ representation2$. The results depicted in Tables 4.8 and 4.9 shows the benefit of using N_{max} ages with the highest probabilities. We can observe that the MAE decreases in all datasets as N_{max} increases from one to six. In our work, the best results were, in general, obtained with the six highest probabilities.

Table 4.8: MAE obtained with the DRF using the highest probabilities method with Fused-representation1.

Databases N_{max} Probabilities	1	2	3	4	5	6
FG-NET	3.77	3.90	3.81	3.72	3.70	3.67
PAL	3.07	2.79	2.78	2.80	2.73	2.80
MORPH Caucasian	6.11	5.51	5.16	4.98	4.86	4.78
LFW+	5.99	5.86	5.82	5.82	5.82	5.83
APPA-REAL Real Age	5.39	5.25	5.28	5.30	5.33	5.34
APPA-REAL Apparent Age	3.47	3.36	3.37	3.39	3.40	3.43
FACES	1.35	1.24	1.21	1.24	1.24	1.24

Table 4.9: MAE obtained with the DRF using the highest probabilities method on FACES database with Fused-representation1.

Face expressionN_{max} Probabilities	1	2	3	4	5	6
Neutrality	0.86	0.75	0.73	0.73	0.72	0.72
Happiness	1.15	1.02	0.98	1.01	1.03	1.02
Disgust	1.73	1.637	1.58	1.62	1.65	1.64
Fear	1.47	1.38	1.43	1.46	1.45	1.44
Sadness	1.18	1.07	1.00	0.98	1.02	1.04
Angry	1.71	1.60	1.56	1.57	1.57	1.57
Average	1.35	1.24	1.21	1.24	1.24	1.24

Table 4.10: MAE obtained with the DRF using the highest probabilities method with Fused-representation2.

DatabasesN_{max} Probabilities	1	2	3	4	5	6
FG-NET	3.90	3.80	3.79	3.78	3.79	3.86
PAL	3.09	2.86	2.85	2.86	2.85	2.86
MORPH Caucasian	4.07	3.98	3.93	3.89	3.88	3.88
LFW+	6.11	5.96	5.92	5.90	5.89	5.89
APPA-REAL Real Age	6.52	6.45	6.45	6.50	6.60	6.63
APPA-REAL Apparent Age	3.57	3.50	3.53	3.53	3.56	3.60
FACES	1.85	1.63	1.59	1.58	1.56	1.55

Table 4.11: MAE obtained with the DRF using the highest probabilities method on FACES dataset with Fused-representation2.

Face expressionN_{max} Probabilities	1	2	3	4	5	6
Neutrality	0.88	0.76	0.75	0.73	0.72	0.70
Happiness	1.69	1.35	1.35	1.34	1.29	1.27
Disgust	2.63	2.42	2.25	2.20	2.16	2.15
Fear	2.10	1.86	1.88	1.89	1.88	1.88
Sadness	1.51	1.34	1.32	1.29	1.30	1.30
Angry	2.29	2.02	2.01	2.01	2.02	2.00
Average	1.85	1.63	1.59	1.58	1.56	1.55

Figures 4.7.(a), 4.7.(b), 4.7.(c), and 4.7.(d) illustrate graphically the MAE as a function of N_{max} (the results were also depicted in Tables 11, 12, 13, and 14). Using the average of N_{max} ages allowed the reduction of the final MAE by exploiting the strength of decision trees that can provide a distribution of the estimates. Thus, this scheme helped to get more accurate age prediction.

In Tables 4.10 and 4.11, the obtained MAEs are better than those obtained by many existing methods. In Tables 4.8 and 4.9, we can observe a constant decrease of the MAE as N_{max} increases. However, in Tables 4.10 and 4.11, there is no constant decrease. Nevertheless, the averaging process show that the

optimal N_{max} is either 5 or 6. For the PAL database, results obtained with *Fused-representation2* were better than those obtained with *Fused-representation1*.

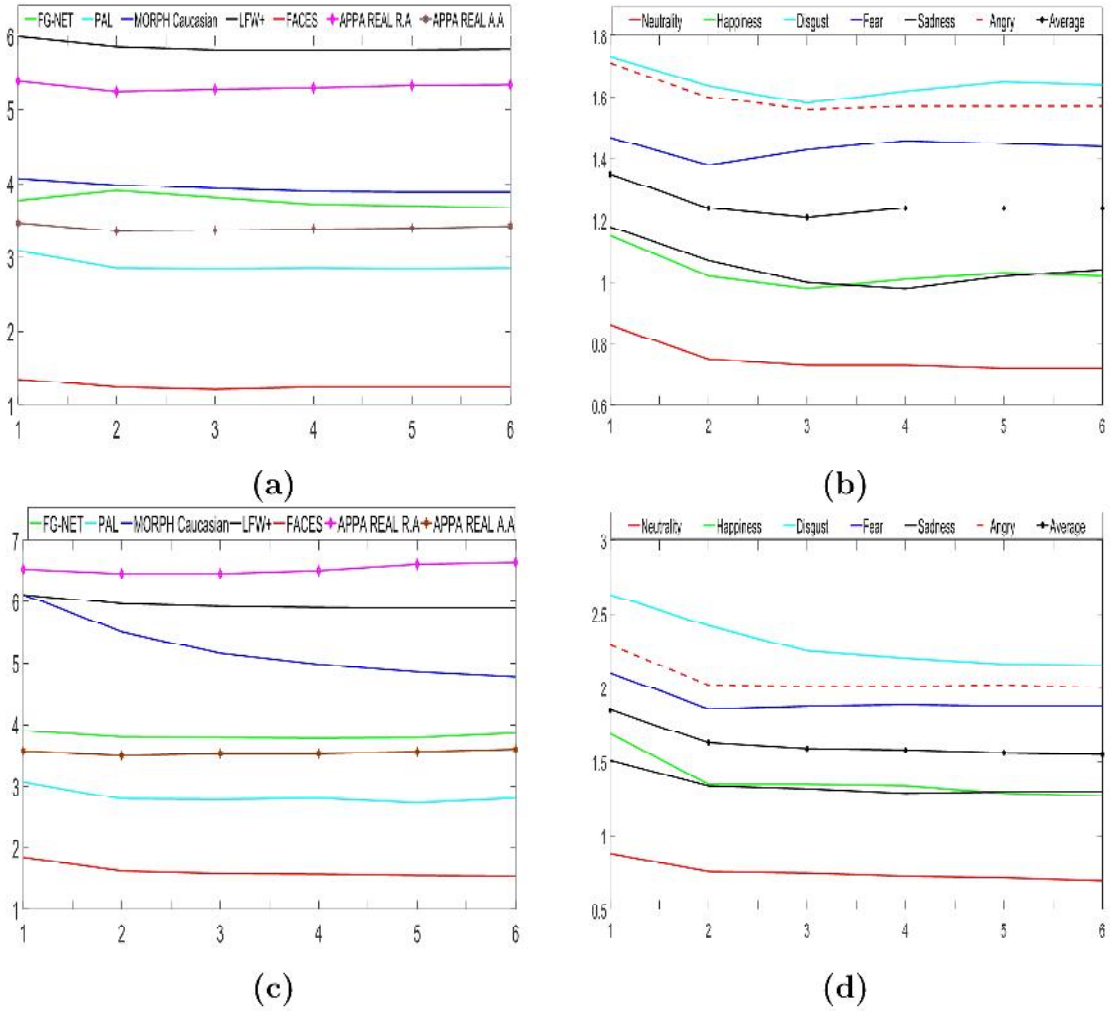


Figure 4.7: Performance as a function of N_{max}. (a): MAE variation with DRF using the input Fused-representation1 on six databases,(b): Six subsets of the FACES dataset,(c) MAE variation with fd-DRF using the input Fused-representation1 on six databases, (d): Six subsets of the FACES dataset.

Fusion schemes

We have also used the concatenation method in the intermediate fusion stage to show the differences between various fusion strategies. The MAE obtained by the concatenation (concatenation of FC6 and FC7) vectors with DRF and SVM are shown in Tables 4.12 and 4.13, respectively. The *fused-*

representation1 and the *fused-representation2* refer to the fusion of FC6 and FC7 obtained by DRF with one level and two levels, respectively. Results indicate that the intermediate fusion by concatenation or by average gives almost the same results. However, using the average method leads to less memory space use and less computational cost since the size of the fused vectors is half that obtained with the concatenation.

Table 4.12: MAE with DRF of fused representation vectors obtained using two fusion strategies.

	Concatenation method		Average method	
SchemeDatasets	PAL	FG-NET	PAL	FG-NET
Fused-representation1	2.92	3.81	3.07	3.77
Fused-representation2	3.35	4.11	3.09	3.90

Table 4.13: MAE with SVM of fused representation vectors obtained using two fusion strategies.

	Concatenation method		Average method	
SchemeDatasets	PAL	FG-NET	PAL	FG-NET
Fused-representation1	3.08	4.73	3.11	4.67
Fused-representation2	3.14	4.52	2.99	4.33

4.5.3 Comparison with state-of-art methods

We compared our method, in term of the MAE, with the state-of-the-art methods depicted in Table 4.14. This table presents the results associated with MORPH Caucasian, FG-NET, PAL, LFW+ , FACES and the APPA-REAL database with both label types (real age and apparent age). Table 4.15 shows the comparison of our method with state of the art with FACES database in detail of expression folds. Our work outperforms all the state of the arts in FACES and PAL with a large difference and in FG-NET too. Table 4.15 shows

that our method is better than the compared methods in any face expression fold.

Table 4.14: Comparison of our method with some of state-of-the-art method using six datasets FG-NET, MORPH Caucasian, PAL, LFW+, FACES and APPA-REAL.

MethodDatabase	FG-NET	MORPH Caucasian	PAL	LFW+	FACES	APPA-REAL Real Age	APPA-REAL Apparent Age
Human workers [106]	4.70	6.30	/	/	/	/	/
Rank [107]	5.79	/	/	/	/	/	/
DIF [106]	4.80	/	/	7.8	/	/	/
AGES [41]	6.77	8.38	/	/	/	/	/
IIS-LLD [39]	5.77	/	/	/	/	/	/
CPNN [39]	4.76	/	/	/	/	/	/
CA-SVR [23]	4.67	5.88	/	/	/	/	/
OHRank [22]	4.48	6.07	/	/	7.14	/	/
Pontes et al. [108]	4.50	/	/	/	/	/	/
CAM [109]	4.12	/	/	/	/	/	/
Rothe et al. [24]	5.01	3.45	/	/	/	/	/
Liu et al. [110]	3.93	/	/	/	/	/	/
LSDML [25]	3.92	/	/	/	4.14	/	/
Liu et al. [111]	3.92	/	/	/	/	/	/
DRFs [6]	3.85	**2.91**	/	/	/	/	/
Gunay and Nabiyev [26]	/	/	5.40	/	/	/	/
Nguyen et al [112]	/	/	6.50	/	/	/	/
Luu et al [109]	/	/	6.00	/	/	/	/
Bekhouche et al. [113]	/	/	5.00	/	/	/	/
Dornaika et al. [7]	/	/	3.79	/	/	/	/
(DMTL) Han et al. [27]	/	/		**4.50**	/	/	/
Structured learning [94]	3.89	/	/	/	7.40	/	/
Agustsson et al. [28] (DEX)	/	/	/	/	/	5.46	4.08
Agustsson et al. [28] (Residual DEX)	/	/	/	/	/	5.35	4.45
Proposed method	**3.65**	3.67	**2.73**	5.82	**1.24**	**5.25**	**3.36**

LFW and LFW+ exist in [22, 27]. The work done by Chang et al. [22], they used just the frontal face images (4211 images) of the LFW. The other work [27] authors create the LFW+ and they find the best results due to the many advantages offered by their Multi-task learning approach in which the age and other face attributes are simultaneously predicted. The training of their method requires more auxiliary attributes in addition to the age labels. They proposed Deep Multi-Task Learning (DMTL) network and they use a modified layer, with batch normalization (BN) layer inserted after each Convolution layer for shared feature learning. In [28], the authors introduced the APPA-REAL database that contains both real and apparent age labels. We emphasize that

the work of [28] presented for each type of ages two solutions: (i) Fine-tuned DEX (DEX), and (ii) Fine-tuned DEX followed by a network-based residual estimation (Residual DEX). Experimental results in the two last columns, in table 4.14, have shown that our proposed method outperforms the work of [28] on both types of ages. It also outperforms the two solutions. For the APPA-REAL dataset, although the DEX CHALEARN was fine-tuned, the final performance is still inferior to that obtained by our proposed scheme.

It is obvious that our proposed method is neither a CNN-based approach nor a hand-crafted approach. This improvement is determined by various factors in our architecture.

Table 4.15: Comparison of our method with some state-of-the art methods on FACES database detailed in facial expression.

MethodFace expression	NEUTRAL	HAPPY	DISGUST	FEARFUL	SAD	ANGRY	Average
BIF+OHRANK [22]	5.16	7.64	8.31	7.00	6.87	7.87	7.14
LBP+OHRANK [22]	6.36	8.88	9.20	7.30	9.09	8.86	8.28
BIF [114]	9.50	10.70	13.26	12.65	10.78	13.26	11.69
BIF+MFA [114]	8.14	10.32	12.24	10.73	10.66	10.96	10.50
CS-LBFL [115]	5.06	6.53	7.15	6.32	6.27	6.94	6.46
CS-LBMFL [115]	4.84	5.85	5.70	6.10	4.98	5.50	5.49
DEEPRANK [116]	5.99	7.12	8.15	6.35	7.77	6.68	7.01
DEEPRANKER+ [116]	5.86	7.87	7.80	6.66	7.49	6.59	7.04
LSDML [25]	3.88	3.49	4.41	5.10	4.09	3.87	4.14
MLSDML [25]	3.83	3.11	4.16	5.01	3.67	3.16	3.82
Structured learning [94]	5.97	6.77	8.17	8.25	7.07	8.21	7.40
Proposed method	**0.72**	**1.02**	**1.64**	**1.44**	**1.04**	**1.57**	**1.23**

Table 4.16: Comparison of our method with the results obtained using the well-known DEX-CHALEARN network. The comparison is carried out with six databases.

MethodDatabase	FG-NET	MORPH Caucasian	PAL	LFW+	FACES (Average)	APPA-REAL Real Age	APPA-REAL Apparent Age
DEX-CHALEARN	4.12	4.54	6.71	7.61	7.73	9.57	5.11
Proposed method	**3.65**	**3.88**	**2.73**	**5.82**	**1.24**	**5.25**	**3.36**

Since our proposed method uses the pre-trained DEX-Chalearn network for extracting two types of image features, it would be interesting to compare

the performance of our proposed approach with that obtained by the direct use of this network. Table 4.16 presents the results obtained from a direct use of the DEX-Chalearn network. The pre-trained CNN model is used for a direct age prediction of the face images.The comparison results show that our method provides better results than those obtained by the DEX-Chalearn estimator. Special attention can be drawn to PAL and FACES databases, where we get a significant difference in performance.

4.6 Complexity and running time

The computational complexity for training the proposed architecture which is composed of two parts based DRF will be split in two parts. The first part corresponds to the computational complexity of the DRF Fusion part. This is $V*F*L*\mathcal{O}(n^2 d\ n_{trees})$ where: n is the number of face images, d is the average number of features, n_{trees} is the number of trees per forest, V is the number of original input feature vectors, L is the levels number in the DRF and F is the number of forests in each level. The second part corresponds to the computational complexity of the fd-DRF part (in our work it contains one single level) is given by $F*\mathcal{O}(n^2 d\ n_{trees})$. Finally, for the training phase, the computational complexity of the total architecture is $V*F*L*\mathcal{O}(n^2 d\ n_{trees}) + F * \mathcal{O}(n^2 d\ n_{trees})$.

For the test phase, the computational complexity is $V*F*L*\mathcal{O}(d\ n_{trees}) + \mathcal{O}(d\ n_{trees})$.

Our experiments use a PC equipped with Intel(R) Core(TM) i7-4702MQ cpu @2.20GHz and 8Go of RAM. Table 4.17 depicts the running times (in

ms) associated with the extraction and age estimation for one face image. It offers a good guess for the total running time of the proposed method with any complete database. Table 4.17 also details the running time of every sub-process using DRF-Fusion with one layer.

As can be seen, the running time of the DRF-Fusion with one layer is 10.6 ms. We note that V is equal to two corresponding to the use of the input vectors FC6 and FC7 in our experiments. Table 4.18 depicts the total running time when the DRF-Fusion used 2 layers. The running time associated with the DRF fusion increased due to the use of more than one layer, **that influence the prediction running time**. In Table 4.17 and 4.18, feature extraction running time has the highest running time compared with other processes, especially when we compared it with the DRF-Fusion running time. The deep feature extraction influences the total time of the proposed architecture. That fact encourages to envision the use of other types of features that are much faster to extract.

This can be given by the hand-crafted features.

Table 4.17: Running time (in ms) of the different phases of the proposed approach (extraction and age prediction) for one face image. Two types of features were used FC6 and FC7. The architecture adopted one layer for DRF-Fusion.

Phase	Pre-processing	Feature extraction	DRF-Fusion 1 layer	Prediction	Total time
Time ms	11.6	370.1	10.6	0.35762	392.65

Table 4.18: Running time (in ms) of the different phases of the proposed approach for one face image. Two types of features were used FC6 and FC7. The architecture adopted two layers for DRF-Fusion.

Phase	Pre-processing	Feature extraction	DRF-Fusion 2 layers	Prediction	Total time
Time (ms)	11.6	370.1	30.1	0.3077	412.1

The main advantage of the proposed method is its training's cheap com-

putational cost, even when deep features, like DEX-Chalearn, are used. The complexity of the training stage is lower than that of classic deep learning approaches. Table 4.19 shows the running time for the training phase using the overall PAL dataset witch contains 1046 images. When dealing with such tasks, the computational toll of the training phase is lighter than the commonly used deep learning method.

Table 4.19: Running time (in seconds) when the PAL dataset is used as a training set. It includes the feature extraction (using the pre-trained model DEX-Chalearn) and the learning phase of the DRF in both cases one layer and two layers.

Phase	Feature extraction	Training (1 layer)	Training (2 layers)
Time (s)	493.865	169.0506	186.947

4.7 Conclusion

Throughout this work, we have proposed a new architecture for age estimation based on facial images. This architecture stands on a recently proposed classification method, currently known as Deep Random Forest. Our architecture is mainly built on a cascade of classification forests ensembles similar to those found in the DRF method and is composed of two types of DRFs. One seeking the enrichment of the feature representation of a given facial descriptor followed by a fusion of the enriched (high level) feature vectors. The other operates on the fused form of all of the enhanced representations in order to estimate the age. Experiments were conducted on different public databases: FG-NET, MORPH Caucasian, PAL, LFW+, FACES, and APPA-REAL. These experiments demonstrate the outperformance of the proposed architecture over many existing state-of-the-art methods. The main limitation

of the proposed method is the space complexity, where we will include space reduction techniques in the future work.

Chapter 5

Facial Age Estimation Using Tensor Based Subspace Learning and Deep Random Forests

5.1 General introduction

scheme [21] for fusing multiple deep face features for age estimation. This scheme was based on Deep Random Forests. We propose a new pipeline that integrates tensor based subspace learning before applying the DRFs. Deep face features of a training set are represented as a 3D tensor. Multi-linear Whitened Principal Component (MWPCA) and Tensor Exponential Discriminant (TEDA) are used to extract the most discriminant information. The features of the tensor subspace are then fed to DRFs in order to predict the age. Experiments conducted on five public face databases show that the method can compete with many state-of-the art methods.

5.2 Introduction

In this chapter, we propose a novel approach able to reduce many of the above limitations. In our proposed approach we use pre-trained CNN models in order to extract features from face images. These features provided by different nets will be used as input features to our estimator. This latter is composed of tensor transformations and Deep Random Forests.

Thus far, subspace transformation is the furthermost utilized dimensional reduction techniques [13, 14]. Various reduced dimensional algorithms have been proposed in the preceding period that have suited the feature extraction. The Principal Component Analysis (PCA) [15] and Linear Discriminant Analysis LDA [16] are frequently used. They are linear subspace techniques. Mainly, an image face is a matrix of m by n pixels, which is treated as a 1-D feature vector of size $m \times n$. Unfortunately, this process involves losing the pixels' position information [17]. Recently, multilinear subspace techniques based on tensor analysis of data in high dimensional spaces is regarded as a remarkable multi-linear technique [18]. These approaches authorize the conservation of the important face structure information. Multilinear transformations analyze the multifactor structure of image face sets over n different index number.

The common linear subspace methods PCA and LDA are extended to Multilinear PCA (MPCA) [17] and Multilinear Discriminant Analysis (MDA) [19] that allow the mathematical of tensors to be manipulated. The high tensor order (i.e., ¿ 2) are presented in a normal form to show the set of face images without collapsing the initial structure and correlation of data [20]. In [1], the auhors propose a new use of an adopted MPCA, this latter is named Multilinear Whitened Principal Component Analysis (MWPCA), which can

deal with the small sample size issue in high dimensional space and can enhance the tough discrimination obtained by classical MPCA. The multilinear varied analysis MDA was also extended to Tensor Exponential Discriminant analysis TEDA so as to improve the discriminant data included in the null space of the within class scatter matrix of each tensor's mode. TEDA increases the margin amidst samples belonging to multiple classes by distance diffusion mappings. In [1] authors present The MWPCA as an extension of MPCA to improve the data representation in the tensorial space. To achieve this, the training tensor data set are centered by subtracting the average tensor from the training sample in a preprocessing step. Subsequently, in an initialization step, the covariance matrix and its eigen-decomposition are computed, this allows for the whitening to be performed on each of tensors, which consists of normalizing each eigenvector by the square root of its corresponding eigenvalue. As a consequence, the data become less correlated and with a uniform variance on all directions. After the initialization whitening of each mode of the tensor sample, an iterative local optimization step of the projection matrices is carried out until the maximum number of iterations is reached or the difference of the projected tensors between two successive iterations becomes less a predefined threshold. Less than a predefined threshold. The process is taken as input upon the set of tensor sample $\mathbf{A}_i \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_m}$ $(i = 1, \ldots, N)$, The number $n_{(k)}$ of selected eigenvectors for each k-mode, the itr_{max}, which is the maximal number of iterations and the threshold η. MWPCA produce the projection tensor $\widetilde{\mathbf{A}} \in \mathbb{R}^{I'_1 \times I'_2 \times \cdots I'_m \times N}$ For the TEDA presented in the same work of ouamane et al. [1] it takes as an input a tensor produced by the previous MWPCA which is defined as $\widetilde{\mathbf{A}} \in \mathbb{R}^{I'_1 \times I'_2 \times \cdots I'_m \times N}$ of the N training samples belonging to L classes,

each class $\widetilde{\mathbf{A}_j}$ contains n_j samples, the itr_{max} is maximal number of iterations and the final lower dimensions,which is $I_1^{"} \times I_2^{"} \ldots I_m^{"}$. We obtained as a final output the projections U_k. TEDA includes the null space of the within-class scatter matrices of each tensor's mode. Additionally, TEDA enlarges the margin between samples belonging to different classes via distance diffusion mappings.

The main contributions of this work are the following:

- We propose a multiview feature fusion that enhances the performance of our previous proposed method in [21] and the techniques in [1].
- We fuse the deep features using the Whitened principal component analysis (MWPCA) and Tensor exponential discriminant analysis (TEDA), respectively.
- Once the face image features are represented in the tensor subspace, the final age is estimated using our recent Deep Random Forests (DRF) [21].

5.3 Building Blocks of the Proposed Method

This section describes the main modules of our pipeline. The latter is composed of two parts (see Figure 5.2). The first one consists of a Multilinear whitened PCA (MWPCA) followed by a Tensor Exponential Discriminant Analysis (TEDA) [1]. The second part performs regression-by-classification using Deep Random Forests [21] that map the Tensor space features to a predicted age.

5.3.1 Multilinear Whitened PCA (MWPCA)[1]

A tensor $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m}$ is defined as a multidimensional array [117] [118]. m is the order of the tensor and $\mathbf{A}$ is called an m^{th}-order tensor. I_k, $1 \leq k \leq m$, is the dimension of the k^{th} mode. Each element of the tensor $\mathbf{A}$ is denoted as $\mathbf{A}_{i_1 i_2 \cdots i_m}$, where $1 \leq i_k \leq I_k$, $1 \leq k \leq m$. A number of mathematical operations on tensors used in this work are presented hereafter.

The inner product of two tensors, $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m}$, with the same order and dimensions is given by:

$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i_1=1 \cdots i_m=1}^{I_1 \cdots I_m} \mathbf{A}_{i_1 \cdots i_m} \mathbf{B}_{i_1 \cdots i_m}$.

The norm of a tensor $\mathbf{A}$ is defined as:

$\|\mathbf{A}\|_F = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle}$.

The difference between two tensors $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m}$ is defined by $D(\mathbf{A}, \mathbf{B}) = \|\mathbf{A} - \mathbf{B}\|_F$.

The k-mode unfolding of a tensor $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_m}$ to a matrix $A^{(k)} \in \mathbb{R}^{I_k \times \prod_{i \neq k} I_i}$ is denoted by $\mathbf{A} \Rightarrow_k A^{(k)}$, where:

$$A^{(k)}_{i_k j} = \mathbf{A}_{i_1 \cdots i_m}, \quad j = 1 + \sum_{l=1, \; l \neq k}^{m} (i_l - 1) \prod_{o=l+1, o \neq k}^{m} I_o \tag{5.1}$$

The unfolding operation on a 3^{rd}-order tensor is illustrated by Fig. 5.1.

The k-mode product of a tensor $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_m}$ and a matrix $U \in \mathbb{R}^{I'_k \times I_k}$ $(k = 1, 2, \ldots, m)$ is an $I_1 \times I_2 \times \cdots I_{k-1} \times I'_k \times I_{k+1} \times \cdots \times I_m$ tensor denoted by $\mathbf{B} = \mathbf{A} \times_k U$, and:

$$\mathbf{B}_{i_1, \ldots, i_{k-1}, j, i_{k+1}, \ldots, i_m} = \sum_{j=1}^{I_k} \mathbf{A}_{i_1, \ldots, i_{k-1}, j, i_{k+1}, \ldots, i_m} U_{i,j} \tag{5.2}$$

where $j - 1, \ldots, I_k$ and $U_{i,j}$ denotes the (i, j)th element of the matrix U.[1]

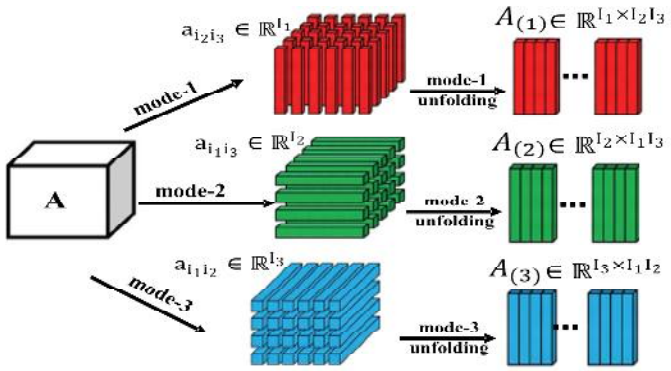


Figure 5.1: Example of tensor unfolding [117]

MPCA is regarded as an extension of PCA to a tensorial space. It permits the projection of tensors samples into a lower subspace, such that the maximum variance present in the original tensor set is captured.

Let $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_m}$ be an m^{th}-order tensor representing a data sample that is feature vectors of a face image. The $(m+1)^{th}$-order tensor, $\widetilde{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_m \times N}$, is the set of all training samples denoted N. Accordingly, the MPCA algorithm consists of the following computational steps [117] :

1. Center the input samples: $\widetilde{\mathbf{A}}_i = \mathbf{A}_i - \overline{\mathbf{A}}$, $i = 1, \ldots, N$. where $\overline{\mathbf{A}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{A}_i$ is the tensor mean of the training tensors.

2. Calculate the the eigenvalue decomposition of the co-variance matrix for each mode k:

   $C_{(k)} = \sum_{i}^{N} \tilde{A}_{i(k)} . \tilde{A}_{i(k)}^{T} = V_{(k)}{}^{T} \Lambda_{(k)} U_{(k)}$,

   with $\tilde{A}_{i(k)}$ is the k-mode unfolding matrix of tensor $\widetilde{\mathbf{A}}_i$, $V_{(k)}$ is the eigenvector matrix and $\Lambda_{(k)}$ is the eigenvalues matrix, let: $U_{(k)} = \left[u_1, u_2, \cdots, u_{n_{(k)}}\right]$, and $U_{(k)}$ contains the selected eigenvectors that corresponds to the largest $n_{(k)}$ eigenvalues.

3. Finally, the dimensionality reduction of a given tensor $\mathbf{A}_i$ is then achieved

as follows:

$$\mathbf{Y}_{\mathrm{i}} = \mathbf{A}_{\mathrm{i}} \times_1 U_{(1)}^{\mathrm{T}} \times_2 U_{(2)}^{\mathrm{T}} \ldots \times_{\mathrm{m}} U_{(\mathrm{m})}^{\mathrm{T}}.$$

Whitening: The whitening is a preprocessing step that consists of a linear transformation of the data such that its covariance matrix is the identity matrix. This step makes the data less correlated with a uniform variation in all directions by normalizing each eigenvector by the square root of its corresponding eigenvalue. Consequently, the feature discrimination of data with high noise is improved.

To improve the data representation using tensors, the MPCA is extended to whitened MPCA (MWPCA). For each mode in the tensor data, the whitening is performed as follows:

$$W_{(\mathrm{k})} = \Lambda_{(\mathrm{k})}{}^{-\frac{1}{2}} U_{(\mathrm{k})}, \Lambda_{(\mathrm{k})} = \left[{}_{12} \cdots {}_{\mathrm{n}_{(\mathrm{k})}} \right] \tag{5.3}$$

where $U_{(\mathrm{k})}$ are the selected eigenvectors that correspond to the $\mathrm{n}_{(\mathrm{k})}$ largest eigenvalues. After the initialization whitening of each mode of the tensor sample. The projection matrices $W_{(\mathrm{k})}$ are iteratively optimized until a maximum number of iterations is reached or the difference of the projected tensors between two successive iteration becomes less than a predefined threshold η. A detailed description of the MWPCA algorithm is found in Algorithm 2.

After the initial whitening of each mode of the tensor, the projection matrices $W_{(\mathrm{k})}$ are iteratively optimized until a number max of iterations is reached or the difference of the norm of the projected tensor between two iterations is less than a preset threshold η. The MWPCA is detailed in Algorithm 2 as in [1].

Algorithm 2 Multilinear Whitened PCA (MWPCA) [1]

Inputs:

- A set of tensor samples $\mathbf{A}_i \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_m}$ $(i = 1, \ldots, N)$
- The number $n_{(k)}$ of selected eigenvectors for each k-mode
- itr_{max} is the maximal number of iterations and the threshold η.

Outputs: The projection matrices $\tilde{W}_{(k)}$

1. **Preprocessing:** Center the input training samples $\widetilde{\mathbf{A}}_i = \mathbf{A}_i - \overline{\mathbf{A}}$, $i - 1, \ldots, N$, and $\overline{\mathbf{A}} - \frac{1}{N}\sum_i^N \mathbf{A}_i$.
2. **Initialization**: Compute the covariance matrix $C_{(k)}$, its eigen-decomposition, and the whitened eigenvectors as: $W_{(k)} = \Lambda_{(k)}{}^{-\frac{1}{2}} U_{(k)}$, sort the $n_{(k)}$ eigenvectors $W_{(k)}$ according to $_{i(k)}$ in decreasing order for $k = 1, \ldots, m$.
3. **Local optimization:**
   - Compute: $\widetilde{\boldsymbol{B}}_i = \widetilde{\boldsymbol{A}}_i \times_1 W_{(1)}^T \times_2 W_{(2)}^T \cdots \times_m W_{(m)}^T, i = 1, \cdots, N.$
   - Compute: $_{B_0} = \sum_{i=1}^N \left\| \widetilde{\mathbf{B}}_i \right\|_F^2$.
   - **For** t= 1 to itr_{max}
     - **For** k—1 to m

       Compute the covariance matrix $C_{(k)}$, its eigen-decomposition, then the whitened eigenvectors as: $W_{(k)} = \Lambda_{(k)}{}^{-\frac{1}{2}} U_{(k)}$, sort the $n_{(k)}$ eigenvectors $W_{(k)}$ according to $_{i(k)}$ in decreasing order.
     - Compute: $\widetilde{\mathbf{B}}_i$, $i = 1, \ldots, N$ and $_{B_k}$.
     - **If** $_{B_k} - _{B_{k-1}} < \eta$, break and go to step 4.
4. **Projection:** The projected tensor is $\mathbf{B}_i = \mathbf{A}_i \times_1 \dot{W}_{(1)}^T \times_2 \dot{W}_{(2)}^T \ldots \times_m \dot{W}_{(m)}^T$, $i = 1, \ldots, N$.

5.3.2 MDA [1]

MDA is created to detect a set of multiple interralted projection matrices $U * mathrmk$ that maximizes inter-class scatter while minimizing it in each mode of the training tensor:

Let the training samples represented as an m^{th}-order tensors $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_m}$ belonging to L different classes and each class j contains n_j samples. MDA seeks m interrelated projection matrices U_k^* by maximizing the inter-class scatter while minimizing the intra-class scatter in each mode of the training tensor:

$$U_k^*|_{k=1}^m = \underset{U_k|_{k=1}^m}{\operatorname{argmax}}$$

$$\frac{\sum_{j=1}^{L} n_j \left\| \overline{\mathbf{A}}_j \times_1 U_1 \cdots \times_m U_m - \overline{\mathbf{A}} \times_1 U_1 \cdots \times_m U_m \right\|^2}{\sum_{i=1}^{N} \left\| \mathbf{A}_i \times_1 U_1 \cdots \times_m U_m - \overline{\mathbf{A}}_{n_i} \times_1 U_1 \cdots \times_m U_m \right\|^2} \tag{5.4}$$

where N is the number of training samples, $\overline{\mathbf{A}}_j$ is the average tensor of each class j, and $\overline{\mathbf{A}}$ is the average tensor of all the training data.

Eq. (5.4) is equivalent to a higher order nonlinear constraint. Hence, finding a a closed-form solution is challenging if not difficult. Alternatively, an iterative optimization approach to estimate the interrelated discriminative subspaces can be applied [118]. Considering the optimization problem from each k-mode, the following objective function can be formulated:

$$U_k^* = \underset{U_k}{\operatorname{argmax}} \frac{\sum_{j=1}^{L} n_j \left\| \overline{\mathbf{A}}_j \times_k U_k - \overline{\mathbf{A}} \times_k U_k \right\|^2}{\sum_{i=1}^{N} \left\| \mathbf{A}_i \times_k U_k - \overline{\mathbf{A}}_{n_i} \times_k U_k \right\|^2} \tag{5.5}$$

The optimization problem is a special discriminant analysis where the sample tensors are unfolded into matrices in the k-mode and the column vector of the

unfolded matrices are labeled with the same class label as the original tensor is unfolded in the k-mode and the column vectors of the unfolded matrices are labeled with the tensor's original label. The problem in Eq. (5.5) is then reformulated as a special discriminant analysis as:

$$U_{\mathrm{k}}^{*}=\underset{U_{\mathrm{k}}}{\operatorname{argmax}}\ \frac{\operatorname{Tr}\left(U_{\mathrm{k}}^{\mathrm{T}} S_{\mathrm{b}} U_{\mathrm{k}}\right)}{\operatorname{Tr}\left(U_{\mathrm{k}}^{\mathrm{T}} S_{\mathrm{w}} U_{\mathrm{k}}\right)} \tag{5.6}$$

where $S_{\mathrm{b}}^{\mathrm{k}}=\sum_{\mathrm{j}=1}^{\mathrm{L}} \mathrm{n}_{\mathrm{j}}\left(\overline{A}_{\mathrm{j}}^{\mathrm{k}}-\overline{A}^{\mathrm{k}}\right)\left(\overline{A}_{\mathrm{j}}^{\mathrm{k}}-\overline{A}^{\mathrm{k}}\right)^{\mathrm{T}}$ and $S_{\mathrm{w}}^{\mathrm{k}}=\sum_{\mathrm{j}=1}^{\mathrm{L}} \sum_{\mathrm{i}=1}^{\mathrm{n}_{\mathrm{j}}}\left(A_{\mathrm{j,i}}^{\mathrm{k}}-\overline{A}_{\mathrm{j}}^{\mathrm{k}}\right)\left(A_{\mathrm{j,i}}^{\mathrm{k}}-\overline{A}_{\mathrm{j}}^{\mathrm{k}}\right)^{\mathrm{T}}$ are the between and within class scatter matrices, respectively. $A_{\mathrm{j,i}}^{\mathrm{k}}$ is the k-mode unfolded matrix of tensor $\mathbf{A}_{\mathrm{i}}$, $\overline{A}_{\mathrm{j}}^{\mathrm{k}}$ the average matrix on class j and $\overline{A}^{\mathrm{k}}$ is the average matrix of the whole training data[118].

5.3.3 Tensor Exponential Discriminant Analysis (TEDA) [1]

To address the problem of small sample size (SSS) as well as preserving the discrimination achieved by the null space of the within-class scatter matrix in LDA, the exponential discreminant analysis (EDA) has been proposed [119]. Taking advantage of this method, MDA [118] is extended to TEDA, by introducing the exponentiation for tensor dimensionality reduction and discrimination improvement. In what comes next a brief description of the MDA will be firstly given, before presenting its extension to TEDA.

TEDA

Using the eigenvalue decomposition, the kth-mode the projection matrix, U_{k}^{*} in Eq. (5.6), is rewritten as:

$$U_{\mathrm{k}}^{*} = \underset{U_{\mathrm{k}}}{\operatorname{argmax}} \frac{\operatorname{Tr}\left(U_{\mathrm{k}}^{\mathrm{T}}\left(\Upsilon_{\mathrm{b}}^{\mathrm{T}} \Lambda_{\mathrm{b}} \Upsilon_{\mathrm{b}}\right) U_{\mathrm{k}}\right)}{\operatorname{Tr}\left(U_{\mathrm{k}}^{\mathrm{T}}\left(\Upsilon_{\mathrm{w}}^{\mathrm{T}} \Lambda_{\mathrm{w}} \Upsilon_{\mathrm{b}}\right) U_{\mathrm{k}}\right)} \tag{5.7}$$

where $\Upsilon_{\mathrm{b}} = (\upsilon_{\mathrm{b}_1}, \upsilon_{\mathrm{b}_2}, \ldots, \upsilon_{\mathrm{b}_{\mathrm{m}}})$ is the eigenvector matrix of S_{b} and $\Lambda_{\mathrm{b}} = diag\left(_{\mathrm{b}_1}, _{\mathrm{b}_2}, \ldots, _{\mathrm{b}_{\mathrm{m}}}\right)$ represent the corresponding eigenvalues. $\Upsilon_{\mathrm{w}} = (\upsilon_{\mathrm{w}_1}, \upsilon_{\mathrm{w}_2}, \ldots, \upsilon_{\mathrm{w}_{\mathrm{m}}})$ is the eigenvector matrix of S_{w} and $\Lambda_{\mathrm{w}} = diag\left(_{\mathrm{w}_1}, _{\mathrm{w}_2}, \ldots, _{\mathrm{w}_{\mathrm{m}}}\right)$ represent the corresponding eigenvalues.

S_{w} is not full-rank matrix under the small sample size situation. In this case, the discriminant data related to the null eigenvalues of S_{w} has the best discriminant power [1]. However, in MDA, this data is discarded by the projection. To prevent this issue, in TEDA, we introduce the expectational by changing $_{\mathrm{w}_i}$ the eigenvalues of S_{w} by $\exp(_{\mathrm{w}_i})$. Hence, the objective function in Eq. (5.7) become:

$$U_{\mathrm{k}}^{*} = \underset{U_{\mathrm{k}}}{\operatorname{argmax}} \frac{\operatorname{Tr}\left(U_{\mathrm{k}}^{\mathrm{T}}\left(\Upsilon_{\mathrm{b}}^{\mathrm{T}} \exp(\Lambda_{\mathrm{b}}) \Upsilon_{\mathrm{b}}\right) U_{\mathrm{k}}\right)}{\operatorname{Tr}\left(U_{\mathrm{k}}^{\mathrm{T}}\left(\Upsilon_{\mathrm{w}}^{\mathrm{T}} \exp(\Lambda_{\mathrm{w}}) \Upsilon_{\mathrm{b}}\right) U_{\mathrm{k}}\right)} \tag{5.8}$$

Applying the property 8 in [1] of experiential matrix, Eq. (5.7) become:

$$U_{\mathrm{k}}^{*} = \underset{U_{\mathrm{k}}}{\operatorname{argmax}} \frac{\operatorname{Tr}\left(U_{\mathrm{k}}^{\mathrm{T}}\left(\exp(S_{\mathrm{b}})\right) U_{\mathrm{k}}\right)}{\operatorname{Tr}\left(U_{\mathrm{k}}^{\mathrm{T}}\left(\exp(S_{\mathrm{w}})\right) U_{\mathrm{k}}\right)} \tag{5.9}$$

Based on property 2 in [1], the matrix $\exp\left(S_{\mathrm{w}}\right)$ is a full-rank matrix. Consequently, the discriminant data included in the null space of S_{w} can be preserved by equation Eq. (5.9). The optimal projection matrix U_{k}^{*} for each k-mode and iteration comprises to the first leading $\mathrm{n}_{(\mathrm{k})}$ eigenvectors of

$\exp(S_{\mathrm{b}})\ \Upsilon = \Lambda \exp(S_{\mathrm{w}})\ \Upsilon$, where $_1 \geq _2 \geq \cdots \geq _{\mathrm{n}_{(\mathrm{k})}}$.

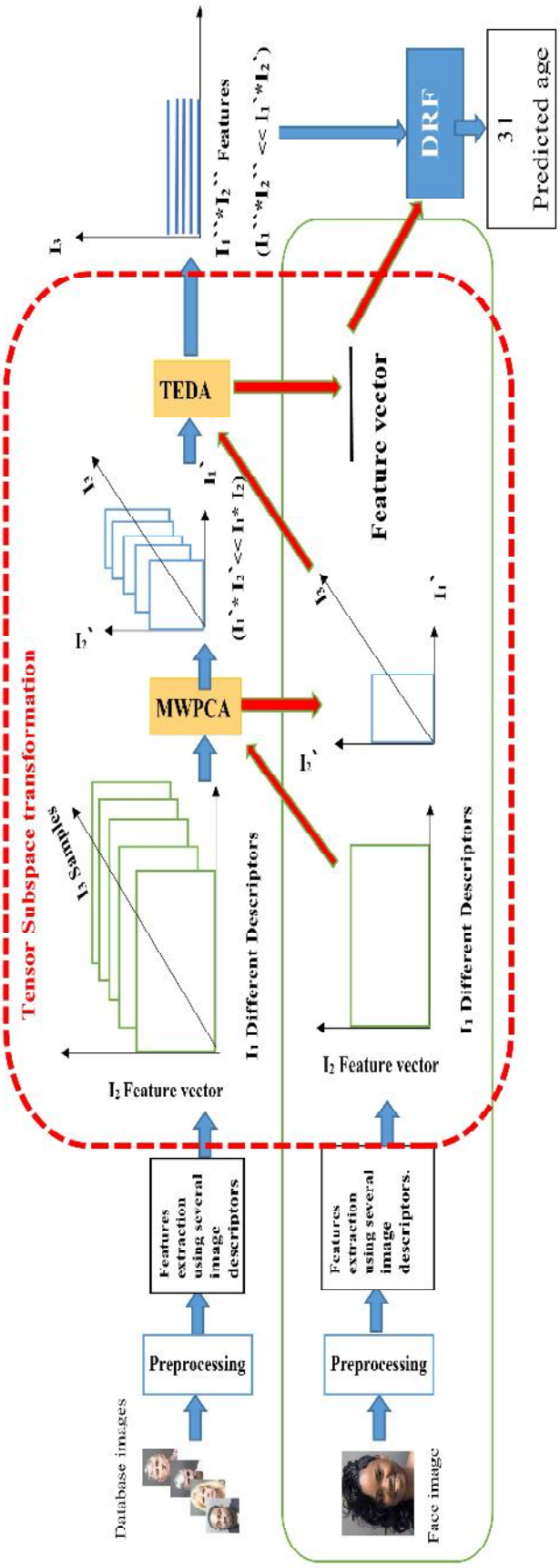


Figure 5.2: Illustration of the proposed architecture. The model is given by the MWPCA, TEDA, and DRFs. Any test image is fed to this pipline (red arrows) in order to get the associated age.

5.3.4 Deep Random Forests for age estimation [21]:

Our proposed method [21] targets the age estimation problem, it predicts the age from a single facial image. It performs regression-by-classification. In this subsection, we will briefly describe its principle.

The aim behind the RF method is the production of several predictors before combining their various predictions rather than attempting to get an optimized procedure at once. More details in relation to Random Forests can be found in [120]. In [21], we have brought into practice a new approach able to resolve the age estimation problem from face image. It showed a good performance compared with the state of the art. It consists of ensembles of Random Forests. The ensembles of Random Forests create a cascade structure by making more than one layer. An ensemble of Random Forests forms a layer in the structure. The feature vector is received by the first layer as a given input. A class probability distribution will be produced by every forest of the similar level. A C-dimensional class vector will be the output of every forest if there are C classes to predict. Concatenating the original input vector with the produced class vectors of every forest (coming from prior level) helps in obtaining the input vector for the subsequent layers.

5.4 Proposed approach

In this section, we will shed the light on the proposed architecture. It is composed of two main stages: the tensor dimensional reduction and Deep Random Forest for final age estimation. These two main parts are performed during a training stage in which the first part (tensor transformations) is

determined first, then the second part (DRFs) is estimated. We assume that each face images has several types of deep descriptors. These feature vectors form a 2D matrix that is considered as a feature matrix of one face image. Thus, a training set of face images can be represented by a 3D tensor. Next, the 3D tensor will be treated by the MWPCA and TEDA techniques successively to reduce the dimension with preserving the essential component. The output of TEDA will be the input to the Deep Random Forests method for the final estimation. We will present the details of the implementation of the proposed method in the next section.

We collect multiple types of descriptors for each face image. The purpose of using those different features types is to exploit the diverse types of information in order to enhance the age estimation process. After extracting the multiple face image feature vectors we will gather them in a 2D matrix of size $I_1 \times I_2$, where I_1 is the number of used feature vectors and I_2 is the dimension of the feature vectors.

The optimal multi-linear projection matrices are estimated in the training stage. After the model is computed, any new face image could be projected by the aforementioned tensor transformations in the test stage. The training 3^{rd} order Tensor $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is constructed using the feature vectors, which are extracted from the pre-processed face image of the training database. The feature vectors can be of different types. However, they should have the same dimension I_1. The tensor $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ modes are :

- I_1 denotes the ensemble of descriptors.
- I_2 denotes the feature vector dimension.
- I_3 denotes the number of training face samples.

The transformation of the input 3D tensor $\mathbf{X}$ are estimated based on the MWPCA/TEDA methods. According to I_1 and I_2 modes which are projected into another subspace. Hence, a new reduced tensor $\mathbf{Y} \in \mathbb{R}^{I_1'' \times I_2'' \times I_3}$, where $I_1'' \times I_2'' \prec I_1 \times I_2$. For further processing, these transformed features are reshaped into a vector of dimension $I_1'' \times I_2''$. The same process will be applied to the test samples without any change.

After that we proceed through the Deep Random Forests proposed by [21, 121] where an ensemble of random forests interacts in a form of a cascade structure. The input is a feature vector. It will be processed by multiple random forests. This collection of RFs is considered as a first layer, and the deep random forest consists of multiple layers.

Every RF of the same layer will generate a class probabilities vector. The generated class probabilities vectors will be concatenated with the input vector in order to form the input vector of the next layer. This aims to create a new feature vector with more information. The dimension of the new vector (at first layer) is given by:

$$Dim = D_1 + (F \times C) \tag{5.10}$$

where D_1 denotes the original feature size, F denotes the number of forests, and C is the number of classes.

The out of the first layer will be the input of the second layer, until the final layer (the number of layers is a parameter chosen by the user). In the final layer, the vectors of probabilities will be averaged, to get one final class probabilities vector. The final estimated class will be the averaged over the N biggest class probabilities. This suits the age estimation problem. For more

details see [21].

5.5 Experiments and implementation details

We used five datasets to test the performance of our proposed architecture: MORPH II (with 55,608 images and 5 random split protocol), FG-NET (with 1002 images and LOPO protocol), PAL (1,046 images with 5 fold random split), LFW+ (with 15,699 images and 5 fold cross validation), APPA-REAL (real age labels of 7591 images and 5 fold cross validation protocol).

5.5.1 Pre-processing

We used the ensemble of regression trees (ERT) algorithm [122] to localize the facial landmarks. This algorithm considered a good one for facial landmarks detection. The landmarks points serve for aligning the 2D face image by using the eyes coordination. After performing the 2D alignment, the face region is cropped.

5.5.2 Feature extraction

We used the pre-trained IMBD-WIKI and DEXchalearn models for deep facial feature extraction [80]. We extract the last two fully connected layer vectors of the mentioned pre-trained models FC6 and FC7 of the input preprocessed images with a size of 224×224. For each input face image, the vectors FC6 and FC7 of both models are later extracted as mentioned above to create a 2D matrix feature of size 4096 x 4.

5.5.3 Implementation

After we get the features vectors from the pre-trained model, we extract them to create a matrix of 4096×4. We then extract the matrices of all training samples to from a 3^{rd} order tensor. Each database's training data is used to estimate the matrices for subspace projection. There are two matrices for the MWPCA method and two matrices for the TEDA method. The dimension was automatically determined by retaining 97% of the eigenvalues' energy. Used to monitor the 3rd order tensor projection convergence, the maximum number of iterations is empirically set to 16. We set the convergence threshold to 10^6 for the MWPCA algorithm as done in [1]. For the TEDA method, we modified the class labels by gathering the closest labels (ages) in a unique class, which can be considered as grouping the ages. This was repeated several time times where at any time we try a new class interval, for example for the first run we used a class interval of one year. In the next run, we took 2 years as a class interval and we regenerate the labels and so on. The used class widths were set from $\{1, 2, 3, 4, 5\}$. Note that the DRF part uses the normal labels (i.e., the class is given by one year), and it used two layers (including the decision layer). The remaining settings of the DRFs are similar to those described in [21].

5.5.4 Results

Figure 5.3 shows all experimental results obtained with the five databases. Subplots (a),(b),(c),(d), and (e) depict the MAE in years as a function of two hyper-parameters: the number of highest probabilities and the class width in the TEDA method. The X axis corresponds to the number of the highest probabilities used by the final layer of the DRFs. The Y axis corresponds

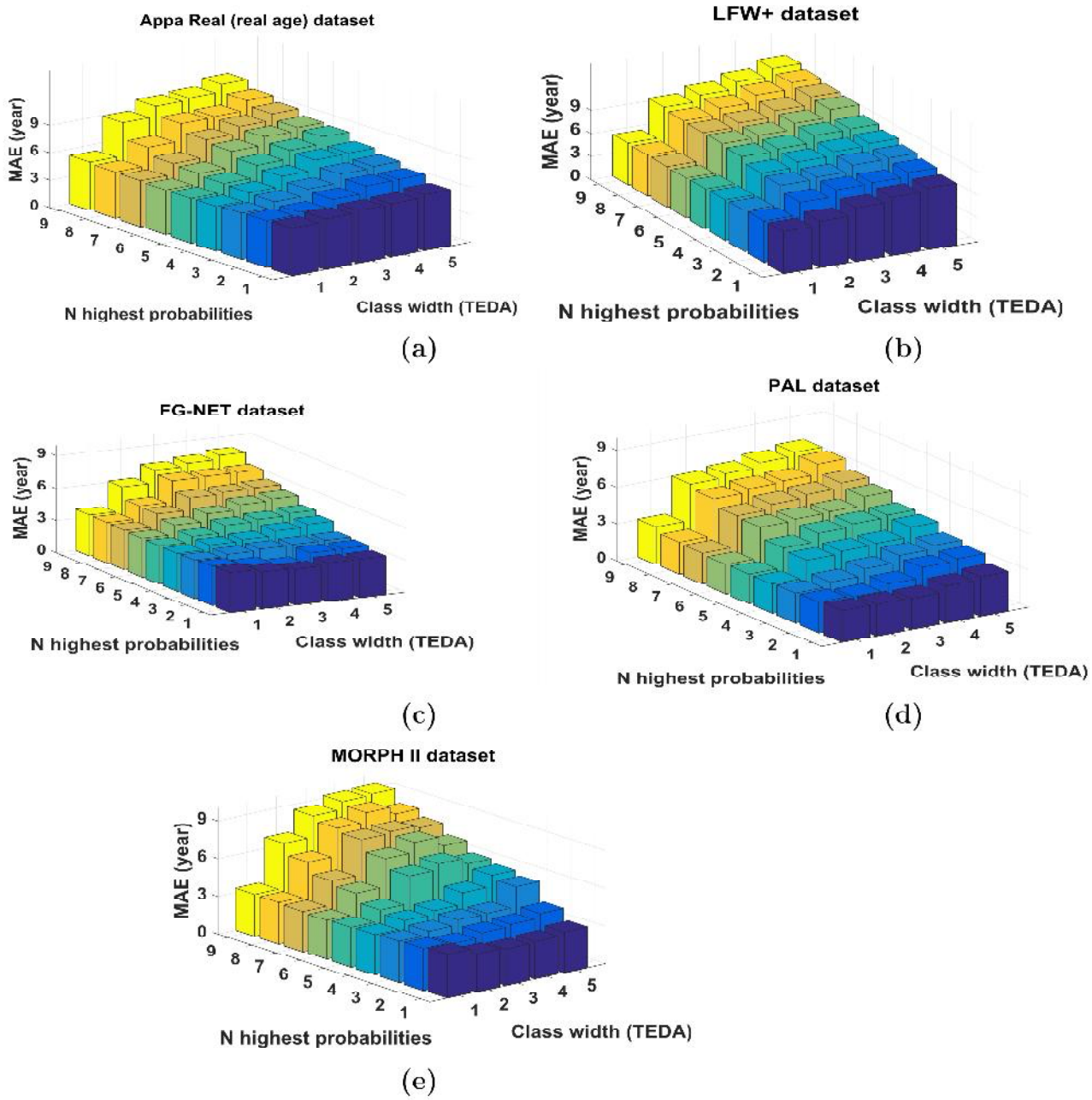


Figure 5.3: MAE of the proposed method as a function of two hyper-parameters: (i) the number of highest probabilities used by the last layer in the Deep Random Forest and (ii) the class width used by the TEDA method.

to the class width (in years) in the TEDA method. Each integer of this axis presents the number of years used for forming the age groups in the TEDA method. For instance, a value of 2 means that the groups of ages are formed by classes having a width of two years.

In Figure 5.3, we can see, in general, that TEDA works better when the class width is set to 2 years. We also observe that, in general, the best MAE was obtained when the number of highest probabilities was greater than one. In (a), it can be appreciated that the best MAE for APPA-REAL (real age) dataset was 4.89. This was obtained with a number N of probabilities equal to 5 and the original labels. For the LFW+ dataset, the best MAE was 5.21, obtained with N number of probabilities equal to 6. For the FG NET dataset, the best MAE was 3.05 obtained by N probabilities number equal to 2 and a class step of 2 years. For the PAL dataset, the best MAE was 2.39, obtained with the original labels and number probabilities equal to 4. For the MORPH II, the best MAE was 2.89 obtained with a number probabilities $N = 2$ and a class width of 2 years.

In Table 5.1, we compared our method, in terms of MAE, with some state-of-the-art methods. This table presents the results associated with the FG-NET, PAL, LFW+, APPA REAL (real ages), and MORPH II databases. Our work outperforms most of the state-of-the-art methods. The proposed method outperformed our previous DRF method. This is due to the use of tensor transformations applied to the deep features.

5.6 Conclusion

We enhanced our previous DRF method for facial age estimation. The current work combines two totally different methods: the first one is based on tensor subspace learning, and the second one is based on Deep Random Forests that performs regression-by-classification. Performances obtained on five public face databases are very promising. These results pave the way to

Table 5.1: Comparison of our method with some of state-of-the-art methods using five datasets FG-NET, MORPH II, PAL, LFW+ and APPA-REAL real age

MethodDatabase	FG-NET	PAL	LFW+	APPA REAL REAL AGE	MORPH II
Liu et al. [110]	3.93	/	/	/	/
LSDML [25]	3.92	/	/	/	3.08
DRFs [6]	3.85	/	/	/	**2.17**
Gunay and Nabiyev [26]	/	5.40	/	/	/
Bekhouche et al. [113]	/	5.00	/	/	/
Dornaika et al. [7]	/	3.79	/	/	3.67
(DMTL) Hun et al. [27]	/	/	**4.50**	/	3.0
Structured learning [94]	3.89	/	/	/	/
Agustsson et al. [28]	/	/	/	5.46	3.25
Agustsson et al. [28] (Residual DEX)	/	/	/	5.35	2.68
Olatunbosun et al [123]	3.56	/	/	5.31	2.72
Guehairia et al [21]	3.65	2.73	5.82	5.25	3.98
Proposed Approach	**3.05**	**2.39**	5.21	**4.92**	2.89

more investigation about enriching the face descriptors and integrating feature selection paradigms in the main parts of the proposed pipeline.

Chapter 6

Conclusion and future work

6.1 Conclusion and future work

In this research, we shed light on one of the biometrics tasks, the age estimation via face images. The latter is considered influential and valuable in several nowadays applications. We also discussed the most and well-known difficulties and challenges facing this field. Particular mention must be made that during our research in this field, we faced several valuable and important kinds of researchers, We have summarized and mentioned only what contributes to the enrichment of our research and what helps to understand it more.
Our work aims to take a part in such field, it focus on minimizing the Mean Absolute Error, using two original architectures, we have provide through the chapters of this thesis two contributions that have been combined to design an overall scheme to solve the estimation of the exact age value from face images.

we proposed a new architecture for age estimation based on facial images. It is mainly based on a cascade of classification trees ensembles, which are known recently as a Deep Random Forest, and the Multi-linear Whitened Principal Component (MWPCA) and Tensor Exponential Discriminant (TEDA). Throughout this work, we have proposed two new architectures for age estimation based on facial images.The first one in chapter three, This architecture stands on a recently proposed classification method, currently known as Deep Random Forest. Our architecture is mainly built on a cascade of classification forests ensembles similar to those found in the DRF method and is composed of two types of DRFs. One seeking the enrichment of the feature representation of a given facial descriptor followed by a fusion of the enriched (high level) feature vectors. The other operates on the fused form of all of the enhanced

representations in order to estimate the age.

Experiments were conducted on different public databases: FG-NET, MORPH Caucasian, PAL, LFW+, FACES, and APPA-REAL. These experiments demonstrate the outperformance of the proposed architecture over many existing state-of-the-art methods. Some of the highlights of the work can be summed up in the following points:

- The reduction of the mean absolute error shows the efficiency of the DRF based extended feature along with the fusion representation, compared to the original feature.
- An even further reduction of the age error was obtained by using the concept of N_{max} probabilities function that was a natural output of the proposed architecture. This concept has shown its superiority over the original decision process.
- The computational complexity of the training stage is cheaper than that of classic deep learning approaches.

In the fourth chapter We enhanced our previous DRF method for facial age estimation. The current work combines two totally different methods: the first one is based on tensor subspace learning, and the second one is based on Deep Random Forests that performs regression-by-classification. Performances obtained on five public face databases are very promising. These results pave the way to more investigation about enriching the face descriptors and integrating feature selection paradigms in the main parts of the proposed pipeline.

6.2 Perspective

In order to improve our proposal, we may pursue a number of research avenues in the future. Despite the proposed age estimation through facial images' good results, it aims to reduce the error gaps between the real and estimated ages.

Future work will focus on the fusion enrichment phase, which will make use of a variety of input features from both (deep features and hand-crafted features). Other new proposals can look into the decision-making process. To achieve this, all individual forests in the last layer can be subjected to a weighted average of ages using the highest probabilities.

Bibliography

[1] A. Ouamane, A. Chouchane, E. Boutellaa, M. Belahcene, S. Bourennane, and A. Hadid. Efficient tensor-based 2d+3d face verification. IEEE Transactions on Information Forensics and Security, 12:2751–2762, 2017.

[2] Ivan Huerta, Carles Fernández, Carlos Segura, Javier Hernando, and Andrea Prati. A deep analysis on age estimation. Pattern Recognition Letters, 68:239–249, 2015.

[3] 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.

[4] 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.

[5] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. Ordinal regression with multiple output CNN for age estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4920–4928, 2016.

[6] Wei Shen, Yilu Guo, Yan Wang, Kai Zhao, Bo Wang, and Alan Yuille. Deep regression forests for age estimation. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2304–2313. IEEE, 2018.

[7] F. Dornaika, I. Arganda-Carreras, and C. Belver. Age estimation in facial images through transfer learning. Machine Vision and Applications, pages 1–11, 2018.

[8] Zhenzhen Hu, Yonggang Wen, Jianfeng Wang, Meng Wang, Richang Hong, and Shuicheng Yan. Facial age estimation with age difference. IEEE Transactions on Image Processing, 26(7):3087 3097, 2017.

[9] Ivan Huerta, Carles Fernández, Carlos Segura, Javier Hernando, and Andrea Prati. A deep analysis on age estimation. Pattern Recognition Letters, 68:239–249, 2015.

[10] Guodong Guo, Yun Fu, Charles R Dyer, and Thomas S Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. IEEE Transactions on Image Processing, 17(7):1178–1188, 2008.

[11] Guodong Guo, Yun Fu, Charles R Dyer, and Thomas S Huang. A probabilistic fusion approach to human age prediction. In Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on, pages 1–6. IEEE.

[12] Leslie G Farkas. Anthropometry of the Head and Face. Raven Pr, 1994.

[13] Young H Kwon and Niels da Vitoria Lobo. Age classification from facial images. Computer vision and image understanding, 74(1):1–21, 1999.

[14] N. Ramanathan and R. Chellappa. Modeling age progression in young faces. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), volume 1, pages 387 394, 2006.

[15] 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.

[16] Sung Eun Choi, Youn Joo Lee, Sung Joo Lee, Kang Ryoung Park, and Jaihie Kim. Age estimation using a hierarchical classifier based on global and local facial features. Pattern Recognition, 44(6):1262 – 1281, 2011.

[17] 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.

[18] Shaohua Kevin Zhou, Bogdan Georgescu, Xiang Sean Zhou, and Dorin Comaniciu. Image based regression using boosting method. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, volume 1, pages 541–548. IEEE, 2005.

[19] Tai Sing Lee. Image representation using 2d gabor wavelets. IEEE Transactions on Pattern Analysis & Machine Intelligence, (10):959 971, 1996.

[20] Jinguang Han and Bir Bhanu. Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis & Machine Intelligence, (2):316–322, 2006.

[21] Jiwen Lu and Yap-Peng Tan. Gait-based human age estimation. IEEE Transactions on Information Forensics and Security, 5(4):761–770, 2010.

[22] 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.

[23] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active appearance models. IEEE Transactions on pattern analysis and machine intelligence, 23(6):681–685, 2001.

[24] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. Automatic age esti-

mation based on facial aging patterns. IEEE Transactions on pattern analysis and machine intelligence, 29(12):2234–2240, 2007.

[25] Xin Geng, Zhi-Hua Zhou, Yu Zhang, Gang Li, and Honghua Dai. Learning from facial aging patterns for automatic age estimation. MM '06, page 307–316, New York, NY, USA, 2006. Association for Computing Machinery.

[26] D. Cai, X. He, J. Han, and H. . Zhang. Orthogonal laplacianfaces for face recognition. IEEE Transactions on Image Processing, 15(11):3608–3614, 2006.

[27] Yun Fu, Ye Xu, and Thomas S Huang. Estimating human age by manifold analysis of face pictures and regression on aging features. In 2007 IEEE International Conference on Multimedia and Expo, pages 1383–1386. IEEE, 2007.

[28] Shuicheng Yan, Huan Wang, Yun Fu, Jun Yan, Xiaoou Tang, and Thomas S Huang. Synchronized submanifold embedding for person-independent pose estimation and beyond. IEEE Transactions on Image Processing, 18(1):202–210, 2008.

[29] Dennis Gabor. Theory of communication. part 1: The analysis of information. Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering, 93(26):429–441, 1946.

[30] Ronald A Fisher. The statistical utilization of multiple measurements. Annals of eugenics, 8(4):376–386, 1938.

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

[32] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. A generalized local binary pattern operator for multiresolution gray scale and rotation in-

variant texture classification. In International Conference on Advances in Pattern Recognition, pages 399–408. Springer, 2001.

[33] Tao Wu, Pavan Turaga, and Rama Chellappa. Age estimation and face verification across aging using landmarks. IEEE Transactions on Information Forensics and Security, 7(6):1780–1788, 2012.

[34] Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. Nature neuroscience, 2(11):1019–1025, 1999.

[35] Zhi-Hua Zhou and Ji Feng. Deep forest: Towards an alternative to deep neural networks. arXiv preprint arXiv:1702.08835, 2017.

[36] S. Wang, S. Yan, J. Yang, C. Zhou, and X. Fu. A general exponential framework for dimensionality reduction. IEEE Transactions on Image Processing, 23(2):920–930, 2014.

[37] T. Zhang, B. Fang, Y. Y. Tang, Z. Shang, and B. Xu. Generalized discriminant analysis: A matrix exponential approach. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 40(1):186–197, 2010.

[38] H. Abdi and Lynne J. Williams. Principal component analysis. WIREs Computational Statistics, 2(4):433–459, 2010.

[39] Yanwei Pang, Shuang Wang, and Yuan Yuan. Learning regularized lda by clustering. IEEE transactions on neural networks and learning systems, 25(12):2191–2201, 2014.

[40] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Mpca: Multilinear principal component analysis of tensor objects. IEEE Transactions on Neural Networks, 19(1):18–39, 2008.

[41] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Mpca: Multilinear principal component analysis of tensor objects. IEEE Transactions on Neural Networks, 19(1):18–39, 2008.

[42] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang. Multilinear discriminant analysis for face recognition. IEEE Transactions on Image Processing, 16(1):212–220, 2007.

[43] Haiping Lu, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos. A survey of multilinear subspace learning for tensor data. Pattern Recognition, 44(7):1540 1551, 2011.

[44] O Guehairia, A Ouamane, F Dornaika, and A Taleb-Ahmed. Feature fusion via deep random forest for facial age estimation. Neural Networks, 130:238–252, 2020.

[45] Kuang-Yu Chang, Chu-Song Chen, and Yi-Ping Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In Computer vision and pattern recognition (cvpr), 2011 ieee conference on, pages 585 592. IEEE, 2011.

[46] Ke Chen, Shaogang Gong, Tao Xiang, and Chen Change Loy. Cumulative attribute space for age and crowd density estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2467–2474, 2013.

[47] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Some like it hot-visual guidance for preference prediction. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5553–5561, 2016.

[48] Hao Liu, Jiwen Lu, Jianjiang Feng, and Jie Zhou. Label-sensitive deep metric learning for facial age estimation. IEEE Transactions on Information Forensics and Security, 13(2):292–305, 2018.

[49] 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.

[50] Hu Han, Anil K Jain, Fang Wang, Shiguang Shan, and Xilin Chen. Heterogeneous face attribute estimation: A deep multi-task learning approach. IEEE transactions on pattern analysis and machine intelligence, 40(11):2597–2609, 2018.

[51] E. Agustsson, R. Timofte, S. Escalera, X. Baro, I. Guyon, and R. Rothe. 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.

[52] Prachi Punyani, Rashmi Gupta, and Ashwani Kumar. Neural networks for facial age estimation: a survey on recent advances. Artificial Intelligence Review, 53(5):3299–3347, 2020.

[53] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. Nature, 521(7553):436–444, 2015.

[54] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5(4):115–133, 1943.

[55] Donald Olding Hebb. The organization of behavior, volume 65. Wiley New York, 1949.

[56] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review, 65(6):386–408, 1958.

[57] Geoffrey E Hinton. To recognize shapes, first learn to generate images. 165:535–547, 2007.

[58] Yoshua Bengio and Yann LeCun. Scaling learning algorithms towards ai. Large-scale kernel machines, 34(5):1–41, 2007.

[59] Marc'Aurelio Ranzato, Christopher Poultney, Sumit Chopra, and Yann L Cun. Efficient learning of sparse representations with an energy-based model. In Advances in Neural Information Processing Systems, pages 1137–1144, 2007.

[60] John Nickolls, Ian Buck, Michael Garland, and Kevin Skadron. Scalable parallel programming with CUDA. Queue, 6(2):40–53, 2008.

[61] Erik Lindholm, John Nickolls, Stuart Oberman, and John Montrym. Nvidia tesla: A unified graphics and computing architecture. IEEE micro, 28(2):39–55, 2008.

[62] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6):84–90, 2017.

[63] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 248–255. IEEE, 2009.

[64] Sun-Chong Wang. Artificial neural network. In Interdisciplinary computing in java programming, pages 81–100. Springer, 2003.

[65] Oludare Isaac Abiodun, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad. State-of-the-art in artificial neural network applications: A survey. Heliyon, 4(11):e00938, 2018.

[66] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E Alsaadi. A survey of deep neural network architectures and their applications. Neurocomputing, 234:11–26, 2017.

[67] S-CB Lo, S-LA Lou, Jyh-Shyan Lin, Matthew T Freedman, Minze V Chien, and Seong Ki Mun. Artificial convolution neural network tech-

niques and applications for lung nodule detection. IEEE transactions on medical imaging, 14(4):711–718, 1995.

[68] David Lopes de MACÊDO. Enhancing deep learning performance using displaced rectifier linear unit. Master's thesis, Universidade Federal de Pernambuco, 2017.

[69] Patrik Kamencay, Miroslav Benčo, Tomáš Mižduš, and Roman Radil. A new method for face recognition using convolutional neural network. 2017.

[70] Mariusz Bojarski, Philip Yeres, Anna Choromanska, Krzysztof Choromanski, Bernhard Firner, Lawrence Jackel, and Urs Muller. Explaining how a deep neural network trained with end-to-end learning steers a car. 2017.

[71] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[72] Asifullah Khan, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi. A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, 53(8):5455–5516, 2020.

[73] Mohammad Mahdi Dehshibi and Azam Bastanfard. A new algorithm for age recognition from facial images. Signal Processing, 90(8):2431–2444, 2010.

[74] Shima Izadpanahi and Önsen Toygar. Human age classification with optimal geometric ratios and wrinkle analysis. International Journal of Pattern Recognition and Artificial Intelligence, 28(02):1456003, 2014.

[75] Xiaolong Wang, Rui Guo, and Chandra Kambhamettu. Deeply-learned feature for age estimation. In 2015 IEEE Winter Conference on Applications of Computer Vision, pages 534–541. IEEE, 2015.

[76] Shuicheng Yan, Huan Wang, Xiaoou Tang, and Thomas S Huang. Learning auto-structured regressor from uncertain nonnegative labels. In 2007 IEEE 11th international conference on computer vision, pages 1–8. IEEE, 2007.

[77] Deng Cai, Xiaofei He, Kun Zhou, Jiawei Han, and Hujun Bao. Locality sensitive discriminant analysis. In IJCAI, volume 2007, pages 1713–1726, 2007.

[78] Xin Liu, Shaoxin Li, Meina Kan, Jie Zhang, Shuzhe Wu, Wenxian Liu, Hu Han, Shiguang Shan, and Xilin Chen. Agenet: Deeply learned regressor and classifier for robust apparent age estimation. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pages 16–24, 2015.

[79] Refik Can Malli, Mehmet Aygun, and Hazim Kemal Ekenel. Apparent age estimation using ensemble of deep learning models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 9–16, 2016.

[80] 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.

[81] Sergio Escalera, Junior Fabian, Pablo Pardo, Xavier Baró, Jordi Gonzalez, Hugo J Escalante, Dusan Misevic, Ulrich Steiner, and Isabelle Guyon. Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pages 1–9, 2015.

[82] Jun-Cheng Chen, Amit Kumar, Rajeev Ranjan, Vishal M Patel, Azadeh Alavi, and Rama Chellappa. A cascaded convolutional neural network for age estimation of unconstrained faces. In 2016 IEEE 8th International

Conference on Biometrics Theory, Applications and Systems (BTAS), pages 1–8. IEEE, 2016.

[83] Zhenzhen Hu, Yonggang Wen, Jianfeng Wang, Meng Wang, Richang Hong, and Shuicheng Yan. Facial age estimation with age difference. IEEE Transactions on Image Processing, 26(7):3087–3097, 2017.

[84] John R Hershey and Peder A Olsen. Approximating the kullback leibler divergence between gaussian mixture models. In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, volume 4, pages IV–317. IEEE, 2007.

[85] G Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments univ. massachusetts, amherst. Technical report, MA, Tech. Rep. 07-49, 2007.

[86] M. Saad Shakeel and Kin-Man Lam. Deep-feature encoding-based discriminative model for age-invariant face recognition. Pattern Recognition, 93:442 – 457, 2019.

[87] Junliang Xing, Kai Li, Weiming Hu, Chunfeng Yuan, and Haibin Ling. Diagnosing deep learning models for high accuracy age estimation from a single image. Pattern Recognition, 66:106 – 116, 2017.

[88] Abrar H Abdulnabi, Gang Wang, Jiwen Lu, and Kui Jia. Multi-task CNN model for attribute prediction. IEEE Transactions on Multimedia, 17(11):1949–1959, 2015.

[89] Peter Kontschieder, Madalina Fiterau, Antonio Criminisi, and Samuel Rota Bulo. Deep neural decision forests. In Proceedings of the IEEE international conference on computer vision, pages 1467–1475, 2015.

[90] Shixing Chen, Caojin Zhang, Ming Dong, Jialiang Le, and Mike Rao.

Using ranking-cnn for age estimation. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[91] H. Liu, J. Lu, J. Feng, and J. Zhou. Ordinal deep learning for facial age estimation. IEEE Transactions on Circuits and Systems for Video Technology, 29(2):486–501, Feb 2019.

[92] Junlin Hu, Jiwen Lu, and Yap-Peng Tan. Discriminative deep metric learning for face verification in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1875–1882, 2014.

[93] Junlin Hu, Jiwen Lu, and Yap-Peng Tan. Deep transfer metric learning. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 325–333, 2015.

[94] Zhongyu Lou, Fares Alnajar, Jose M Alvarez, Ninghang Hu, and Theo Gevers. Expression-invariant age estimation using structured learning. IEEE transactions on pattern analysis and machine intelligence, 40(2):365–375, 2018.

[95] 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.

[96] Dat Tien Nguyen, So Ra Cho, and Kang Ryoung Park. Age estimation-based soft biometrics considering optical blurring based on symmetrical sub-blocks for mlbp. Symmetry, 7(4):1882–1913, 2015.

[97] Olufade FW Onifade and Damilola J Akinyemi. Gwageer-a groupwise age ranking framework for human age estimation. International Journal of Image, Graphics and Signal Processing, 7(5):1, 2015.

[98] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up

robust features. In European conference on computer vision, pages 404–417. Springer, 2006.

[99] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), volume 1, pages 886–893. Ieee, 2005.

[100] Guodong Guo, Yun Fu, Charles R Dyer, and Thomas S Huang. A probabilistic fusion approach to human age prediction. In 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pages 1–6. IEEE, 2008.

[101] Leo Breiman. Random forests. Machine learning, 45(1):5–32, 2001.

[102] Gilles Louppe. Understanding random forests from theory to practice. PhD thesis, Liège, 2015.

[103] Xianghai Cao, Renjie Li, Yiming Ge, Bin Wu, and Licheng Jiao. Densely connected deep random forest for hyperspectral imagery classification. International Journal of Remote Sensing, 0(0):1–16, 2018.

[104] Ya-Lin Zhang, Jun Zhou, Wenhao Zheng, Ji Feng, Longfei Li, Ziqi Liu, Ming Li, Zhiqiang Zhang, Chaochao Chen, Xiaolong Li, et al. Distributed deep forest and its application to automatic detection of cash-out fraud. arXiv preprint arXiv:1805.04234, 2018.

[105] 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.

[106] 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.

[107] Kuang-Yu Chang, Chu-Song Chen, and Yi-Ping Hung. A ranking approach for human ages estimation based on face images. In 2010 20th International Conference on Pattern Recognition, pages 3396–3399. IEEE, 2010.

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

[109] Khoa Luu, Keshav Seshadri, Marios Savvides, Tien D Bui, and Ching Y Suen. Contourlet appearance model for facial age estimation. In 2011 International Joint Conference on Biometrics (IJCB), pages 1–8. IEEE, 2011.

[110] Hao Liu, Jiwen Lu, Jianjiang Feng, and Jie Zhou. Group-aware deep feature learning for facial age estimation. Pattern Recognition, 66:82 – 94, 2017.

[111] Kuan-Hsien Liu and Tsung-Jung Liua. A Structure-Based Human Facial Age Estimation Framework under a Constrained Condition. IEEE Transactions on Image Processing, 2019.

[112] Dat Tien Nguyen, So Ra Cho, Kwang Yong Shin, Jae Won Bang, and Kang Ryoung Park. Comparative study of human age estimation with or without preclassification of gender and facial expression. The Scientific World Journal, 2014, 2014.

[113] S. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, and A. Hadid. Pyramid multi-level features for facial demographic estimation. Expert Systems with Applications, 80:297 – 310, 2017.

[114] Guodong Guo and Xiaolong Wang. A study on human age estimation under facial expression changes. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 2547–2553. IEEE, 2012.

[115] Jiwen Lu, Venice Erin Liong, and Jie Zhou. Cost-sensitive local binary feature learning for facial age estimation. IEEE Transactions on Image Processing, 24(12):5356 5368, 2015.

[116] Huei-Fang Yang, Bo-Yao Lin, Kuang-Yu Chang, and Chu-Song Chen. Automatic age estimation from face images via deep ranking. networks, 35(8):1872–1886, 2013.

[117] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Mpca: Multilinear principal component analysis of tensor objects. IEEE Transactions on Neural Networks, 19(1):18–39, Jan 2008.

[118] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. J. Zhang. Multilinear discriminant analysis for face recognition. IEEE Transactions on Image Processing, 16(1):212–220, Jan 2007.

[119] S. J. Wang, S. Yan, J. Yang, C. G. Zhou, and X. Fu. A general exponential framework for dimensionality reduction. IEEE Transactions on Image Processing, 23(2):920–930, Feb 2014.

[120] Leo Breiman. Random forests. Machine learning, 45(1):5–32, 2001.

[121] O. Guehairia, A. Ouamane, F. Dornaika, and A. Taleb-Ahmed. Deep random forest for facial age estimation based on face images. In 020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP), pages 305 309, 2020.

[122] 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.

[123] O. Agbo-Ajala and S. Viriri. A lightweight convolutional neural network for real and apparent age estimation in unconstrained face images. IEEE Access, 8:162800–162808, 2020.