Sudan University of Science and Technology

College of Graduate Studies

Development of Bayesian Optimization Convolutional Neural Network model for Face Age Estimation

تطوير نموذج محسن لشبكة عصبية تلافيفية بيزية لتقدير عمر الوجه

Dissertation

Submitted in Partial Fulfilment of the requirements

for the degree of Doctor of Philosophy in Computer Science and Information Technology

By:

Marwa Jamal Eldein Salih Ahmed

Supervisor:

Prof. Serestina Viriri

August-2021

DECLARATION

I hereby declare that this thesis is the result of my own investigation, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at Sudan University of Science and Technology or other institutions.

Marwa Jamal Eldein Salih Ahmed

Signature **Marwa Jamal Eldein Salih Ahmed** Date **07/8/2019**

A Development of Bayesian Optimization Convolutional Neural Network model for Face Age Estimation

تطوير نموذج شبكة عصبية تلافيفية بيزية محسنة لتقدير عمر الوجه

BY

Marwa Jamal Eldein Salih Ahmed

Supervised by: **Prof. Serestina Viriri**

SUDAN UNIVERSITY FOR SCIENCE & TECHNOLOGY

2021

A Development of Bayesian Optimization Convolutional Neural Network model for Face Age Estimation

تطوير نموذج شبكة عصبية تلافيفية بيزية محسنة لتقدير عمر الوجه

BY

Marwa Jamal Eldein Salih Ahmed

A dissertation submitted in Partialfillment of the requirements for the degree of Doctor of Philosophy

in

(Computer Science)

College of Computer Science and Information Technology

Sudan University for Science & Technology

August 2021

ABSTRACT

Age estimation of unrestricted imaging circumstances has attracted an augmented recognition as it is applicable in several real-world applications such as security control, multimedia communication, human computer interaction, and surveillance. Age estimation is a process of determining the exact age or age group of a person depending on his biometric features. It is a challenging problem to effectively and automatically estimate ages of human. Recent research demonstrates that the deeply learned features for age estimation from large-scale data result in significant improvement of the age estimation performance for facial images.

This research proposes a Convolutional Neural Network (CNN) using Bayesian Optimization for facial age estimation. Bayesian Optimization is applied to minimize the classification error on the validation set for CNN model.

Also an enhanced model based on Gender Classification has been proposed as an extension to the previous model. As it is known, Males and Females have a variable type of appearance aging pattern; this results in age differently. This fact leads to assuming that using gender information may improve the age estimator performance for the preceding model. A Convolutional Neural Network (CNN) is used to get Gender Information, then Bayesian Optimization is applied to this pre-trained CNN when fine-tuned for age estimation task. Bayesian Optimization reduces the classification error on the validation set for the pre-trained model.

Extensive experiments are done to evaluate the proposed model on three datasets: MORPH, FG-NET and FERET. The results show that using Bayesian Optimization on CNN outperforms the state-of-the-arts on FG-NET and FERET datasets with a Mean Absolute Error (MAE) of 2.88 and 1.3, and achieves good results compared to most of the state-of-the-art methods on MORPH dataset with a 3.16 MAE. Also, Extensive experiments are done to assess the enhanced model on two data sets: FERET and FG-NET. The experiments' result indicates that using a pre-trained CNN

containing Gender Information with Bayesian Optimization outperforms the state-of-the-arts on FERET and FG-NET data sets with a MAE of 1.2 and 2.67 respectively.

مستخلص البحث

اجتذب تقدير العمر لظروف التصوير غير المقيدة اعترافًا متزايدًا لأنه مناسب في العديد من تطبيقات العالم الحقيقي مثل التحكم الأمني ، واتصالات الوسائط المتعددة ، والتفاعل البشري مع الكمبيوتر ، والمراقبة. تقدير العمر هو عملية تحديد العمر أو الفئة العمرية للشخص اعتمادًا على سماته الحيوية. إنها مشكلة صعبة لتقدير عمر الإنسان بشكل فعال وآلي. توضح الأبحاث الحديثة أن الميزات المكتسبة بعمق لتقدير العمر من البيانات واسعة النطاق تؤدي إلى تحسن كبير في أداء تقدير العمر لصور الوجه.

يقترح هذا البحث نهج شبكة عصبية تلافيفية - باستخدام التحسين البيزي لتقدير عمر الوجه. يتم تطبيق التحسين البيزي لتقليل خطأ التصنيف في مجموعة التحقق الخاصة بنموذج شبكة عصبية تلافيفية.

اقترح هذا البحث أيضًا نموذجًا جديدًا محسنًا يعتمد على تصنيف الجنس كامتداد للنموذج السابق. كما هو معروف أن الذكور والإناث لديهم نوع متغيرمن نمط شيخوخة المظهر ؛ ينتج عن هذا إختلاف في العمر. تؤدي هذه الحقيقة إلى افتراض أن استخدام معلومات الجنس قد يحسن أداء مقدر العمر للنموذج السابق. تستخدم الشبكة العصبية التلافيفية لاستخلاص معلومات الجنس ثم يتم تطبيق التحسين البيزي على الشبكة العصبية التلافيفية المدرب مسبقًا عند ضبط مهمة تقدير العمر بدقة. يقلل التحسين البيزي خطأ التصنيف على مجموعة التحقق من العمر للنموذج المدرب مسبقًا.

تم إجراء تجارب مكثفة لتقييم النموذج المقترح على ثلاث مجموعات بيانات: MORPH و FG-NET و FERET.أظهرت النتائج أن استخدام التحسين البيزي على الشبكة العصبية التلافيفية يتفوق على الأبحاث السابقة في مجموعات بيانات FG-NET و FERET بمتوسط خطأ مطلق يبلغ 2.88 و 1.3 ، ويحقق نتائج جيدة مقارنة بمعظم الحالات في الطرق السابقة على مجموعة بيانات MORPH ب 3.16 متوسط خطأ مطلق.
أيضًا أجريت تجارب مكثفة لتقييم النموذج المقترح الجديد على مجموعتي البيانات FERET و FG-NET. تشير نتائج التجارب الى أن استخدام الشبكة العصبية التلافيفية المدرب مسبقًا باحتواء معلومات الجنس مع التحسين البيزي يتفوق على الدراسات السابقة في FERET و FG-NET بمتوسط خطأ مطلق يبلغ 1.2 و 2.67 على التوالي.

ACKNOWLEDGEMENTS

I would like to express my gratitude, and appreciation to my gracious and patient teacher Serestina Viriri Who has made his great and tireless support to help me during these hard times.

I dedicate this research to my sweeten hearts, my lovely daughters who have always been an inspiration and motivated me to peruse my goals.

I would like to thank my friend for their academic, and moral support during the journey especially senior mate, Waleed.

Also, million thanks to my best friends Nada, and Yusra who are the scent of my life.

The completion of this research could not have been accomplished without the help of my mother who has given me her golden times.

Also, I would like to thank my brothers who have always been beside me to give me the confidence, and determination till the end.

Very special thanks to my caring, loving, and supportive husband Zuhair who has always been with me like a shadow, and without his support and encouragement I couldn't be able to complete my research.

DECLARATION

I hereby declare that this thesis is the result of my own investigation, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at Sudan University of Science and Technology or other institutions.

Marwa Jamal Eldein Salih Ahmed

Signature: _______________ Date:____________________

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure No. Page No.

LIST OF ABBREVIATIONS

BO	Bayesian Optimization
CACD	Cross-Age Celebrity Dataset
CNN	Convolutional Neural Network
CS	Cumulative Score
DEX	Deep Expectation
DL	Deep Learning
DLBO	Deep Learning using Bayesian Optimization
FC	Fully Connected
FERET	Face Recognition Technology dataset
FG-NET	Face and Gesture Recognition Network
GA-DFL	Group-aware Deep Feature Learning
GPU	Graphical Processing Unit
HAP	Hybrid Ageing Patterns
MAE	Mean Absolute Error
MR-CNN	Metric Regression with CNN
ODFL	Ordinal Deep Feature Learning method
OR-CNN	Ordinal Regression with Multiple Output CNN
VGGNET	Visual Geometry Group Network
RBM	Restricted Boltzmann Machine

PUBLICATIONS

Survey paper was accepted at the 2017 Sudan Conference on Computer Science and Information Technology (SCCSIT), and is already published by IEEE. It can be accessed from this link:
https://ieeexplore.ieee.org/document/8293051/?anchor=authors

M. Ahmed, S. Viriri, "Deep Learning Using Bayesian Optimization for Facial Age Estimation", *Image Analysis and Recognition,* Springer LNCS, vol. 11663, pp. 243-254, **2019.** Published by Springer.

"Facial age estimation using transfer learning and Bayesian optimization based on gender information". Published by Signal & Image Processing: An International Journal (SIPIJ). It can be accessed from the link: https://aircconline.com/abstract/sipij/v11n6/11620sipij04.html

CHAPTER ONE

INTRODUCTION

1.1 INTRODUCTION

Age estimation is an important task in facial image classification. Age estimation is a process of determining the age of a person using his biometric features (Lin etal, 2012). The age definition by computer based on person's appearances called age estimation. The appearance of age is usually nearly close to the actual age. The objective of age estimation is that estimated age is as close to appearance age as possible.

Age classification from facial images has a series problem because the age of human differs based on many aspects which may be internal or external factors. Internal factors contract age include gender, genetic, etc. On the other hand, the external factors which affect the age include lifestyle, drugs, etc. These factors could make it problematic to frame the human growth pattern (Grd, 2013). In the automatic age classification, the main aim is to improve a scared algorithm which allows classifying the age based on features extracted from facial images. The accurateness level is one of the main challenges of the age classification which is caused by the difficulty of the human aging pattern. So it is not sufficient to classify the human age, but also important to predict it as correctly as possible. Another problem related to the age prediction is the age groups range and this limitation is a key factor as different characteristics of aging pattern appear in different groups. Therefore the system is trained to cope with definite range might not be suitable for a various range of age groups (Grd, 2013).

1.2 MOTIVATION

In the area of computerized analysis of facial images for recognition, gender recognition, ethnicity classification, etc. the age estimation is a barely explored issue. In recent times the concentration in age estimation has increased significantly, due to its several real-world applications. For example, there are age limitations for buying alcohol, video games, films, cigarettes, and for driving a car, which should be followed.

Nevertheless, the skills of a human to estimate age are very limited (Ahmed etal, 2017). Therefore, a well computer system that supports human ability to estimate age would be helpful. For the reason that the human-computer interaction differs for different age groups, a system's interface to the current user could be adapted to clear this problem. There are several known world applications that deal with facial age estimation directly or indirectly (Fu etal, 2010). These are some examples of the applications:

1. **Age-based retrieved for image faces:**

   Depending on basics of age, an indexing for the database of face images can be done. It seems useful in recalling of images based on age, for example, to discover what percentage of teenagers prefer laptops over desktops (Fu etal, 2010).

2. **Security Control:**

   To identify a user's face is the easiest biometric trait to be used in Security Systems. Face recognition is commonly used to secure private systems from unauthorized users. However, the quality of security provided by face recognition systems is hindered by time which is the only obstacle. This can be reduced by the combination of facial age estimation and face recognition (Fu etal, 2010).

3. **Forensic Art:**

   The knowledge of the shape of faces, aging of the human body, psychology, and perception is the forensic art. A basic practice found in this art is that; changes in face age progression are projected in the photograph by automated systems or artists to include age effects. Forensic artists use age estimation by automated systems to draw a face sketch for lost person identification and to identify possible suspects (Ahmed etal, 2017).

4. **Surveillance:**

   These days surveillance monitoring is considered as one of the difficult tasks comparing with the easily accessible information. Immaturity cannot be allowed to enter mature places or use tobacco by using monitoring camera and age estimation.

Also, children are forbidden to access adult websites by the same way (Ahmed etal, 2017).

1.3 CHALLENGES

The age estimation problem presents unique challenges that contain:

1. **Dependence on external factors**: some external factors that affect the age estimation process include psychology, health conditions, and lifestyle (Reade etal, 2015).
2. **Limited inter-age group variation**: Sometimes finding differences in the appearance is a challenging task for adjacent age groups, this leads to difficulties in estimating ages. This difficulty is increased when dealing with mature subjects (Reade etal, 2015).
3. **Data availability**: Developing accurate systems for age estimation needs the availabilities of suitable datasets for training and testing. These datasets must consist of many images for the same subject covering a wide range of age. Collecting such datasets necessitates images taken from the past. The currently available datasets FG-NET and MORPH that support facial aging experimentation have a limitation in that the FG-NET database consists of images showing significant non-aging related variation and the MORPH database includes for each subject only few samples (Reade etal, 2015).
4. **The diversity of aging variation:** The amount of facial wrinkles may differentiate between individuals related to the same age group. As a result of that for different groups of subjects, different approaches for age estimation may be required (Reade etal, 2015).

1.4 PROBLEM STATEMENT AND ITS SIGNIFICANCE

Currently, Aging affects the human face appearance. Skin related deformations and bone movement and growth affect the facial aging related to reduction of muscle strength and the introduction of wrinkles. During childhood the bones grow while texture changes happen in adult ages. Humans can estimate the age of other people when looking at their faces. However, age estimation's researchers conclude that the ages estimated by humans are not accurate .Hence the need of improving methods for automatic facial age estimation takes an attractive direction.

The main aim in automatic facial age estimation is to have an effective algorithm that facilitates estimating the person's age by using features extracted from the facial image. The problem of estimating age shares numerous similarities with other interpretation tasks of face image where the execution phase contains the process of detecting face, extracting features and classification. Depending on the application used for age estimation, the classification phase output can be the age group of a person or the exact age of a person.

The Research Questions are:

1. How to specify the parameters' values that controlling the network architecture and the training options to develop the deep learning methods that used in age estimation.
2. What is the impact of gender classification in age estimation?

1.5 RESEARCH OBJECTIVES

The main aim of this research is to specify the parameters' values that controlling the network architecture and the training options to develop the deep learning methods that used in age estimation. Also, using gender information to enhance the proposed model.

The objectives of this thesis are:

1. To identify good deep learning methods that improve the accuracy rate for age estimation.
2. To optimize the parameters of the training options and the network architecture using Deep learning with Bayesian Optimization.
3. To model a framework for accurate age estimation using robust convolutional neural network and Bayesian optimization.

4. To evaluate the Proposed Model.
5. To investigate the impact of Gender Classification in age estimation.

1.6 THESIS CONTRIBUTIONS

A detailed survey about age estimation based on hand crafted features and deep learning for facial images.

1. An improvement of the accuracy rate for age estimation using deep learning and Bayesian optimization.
2. Modelling a framework for accurate age estimation using deep learning and Bayesian optimization.
3. Investigating the impact of gender information in age estimation.

1.7 RESEARCH ORGANIZATION

The organization of the following chapters in this research is as follows:

Chapter 2: presents the literature review that is associated to the age estimation, provides more information about the research problem, and critically investigates the existing solutions, which were proposed to address the age estimation problem.

Chapter 3: presents the related works that are used for our proposed model in details

Chapter 4: presents the research methodology and the details of the proposed solution.

Chapter 5: presents the research results and discussion about the results.

Chapter 6: concludes the research besides the potential future works.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter presents a background about age estimation. Then different models for facial representation to extract useful features for age estimation purposes have been displayed. Also all datasets used by researchers in age estimation field are shown. Finally two concepts used by this research are presented which are: Deep Learning and Bayesian Optimization.

2.2 BACKGROUND

A face image of a human consists of rich information about personal characteristics, including emotional expression, identity, age, gender, etc. In general, an image of a human can be taken as a complex signal containing several facial attributes such as geometric facial features and skin color. These attributes have a critical role for facial image analysis in real applications. Attributes estimated in such applications from face image can infer additional system reactions. In particular, age is further significant among these attributes (Lin etal, 20120).

Age estimation is an important task in facial image classification. It plays a significant role in many applications such as security control, multimedia communication, human computer interaction, and surveillance. Age estimation is a process of determining the age of a person using his biometric features (Grd, 2013). The age definition by computer based on person's appearances is called age estimation. The appearance of age is nearly close to the actual age. The aim of age estimation is to estimate age closely to appearance as possible. Age classification from facial images has a series problem because the age of humans differs based on many aspects which may be internal or external factors. Internal factors contract age include gender, genetic, etc. On the other hand, the external factors which affect the age include lifestyle, drugs, etc. Those factors could make it problematic to frame the human growth pattern (Fu etal, 2010). In the automatic age classification, the main aim is to improve a scared algorithm which allows classifying the age based on features extracted from facial images. The accurateness level is on the main challenges of the age classification which is caused by the difficulty of the human aging pattern. So it is not sufficient to only classify the human age, but also important to predict it as accurately as possible. Another problem related to age prediction is the age groups range, this limitation is a key factor as different characteristics of aging patterns appear in different groups. Therefore the system trained to cope with definite range might not be suitable for a various range of age group (Fu etal, 2010).

The age estimation problem presents unique challenges that contains (Tin, 2012):

1. **Dependence on external factors:** Some external factors that affect the age estimation process include psychology, health conditions, and lifestyle (Tin, 2012).
2. **Limited inter-age group variation:** Sometimes finding differences in the appearance is a challenging task for adjacent age groups, this leads to difficulties in estimating ages. This difficulty is increased when dealing with mature subjects (Tin, 2012).
3. **Data availability:** Developing accurate systems for age estimation needs the availabilities of suitable datasets for training and testing. These datasets must consist of many images for the same subject covering a wide range of age. Collecting such datasets necessitates images taken from the past. The currently available datasets FG-NET and MORPH that support facial aging experimentation have a limitation in that the FG-NET database consists of images showing significant non-aging related variation and the MORPH database consists of for each subject only little samples (Ahmed etal, 2017).
4. **The diversity of aging variation:** Facial wrinkles amount may differentiate between individuals related to the same age group. As a result of that for different groups of subjects, different approaches for age estimation may be required (Ahmed etal, 2017).

2.3 AGE ESTIMATION MODELS FOR FACIAL IMAGES

Facial representation has many different models to extract useful features for age estimation purposes. They are recognized in (Grd, 2013), (Sarita etal, 2016), (Guo etal, 2009), (Geng etal, 2010), and (Guo, 2012) which are:

1. Anthropometric Model Active.
2. Aging pattern subspace Model.
3. Active Appearance Model.
4. Aging Manifold.

2.3.1 Anthropometric model

This model measures the sizes and proportions of human faces. It consults research associated to the growth and development of craniofacial. Theory of Craniofacial research deals with a mathematical model to describe the head of the person from birth to adulthood (Grd, 2013). Farakas (Farakas, 1994) provided a facial anthropometry overview. Facial anthropometric is defined as measurements over 57 facial points and landmarks at different ages from infancy to adulthood. Using these measures, computational methods can be developed to make face characterization at different ages. Anthropometric model is focused on the human face rations, as shown in Figure 2.1. This model is doing well with the people classification in and minors, but it provides bad results in estimating adults' ages. It is suitable for younger people. It concerns only on geometry of the face; no information of texture is taken (Grd, 2013). A network where every node is connected to other nodes on the network through multiple or single hops and some may be connected with more than one hop. In a mesh topology, every node not only sends its own signals but also relays data from other nodes. This type of topology is very expensive as there are many redundant connections, thus it is not mostly used in computer networks. It is commonly used in wireless networks. Flooding or routing technique is used in mesh topology. Figure 2.1 ~~The following figure~~ shows a sample of full mesh topology.

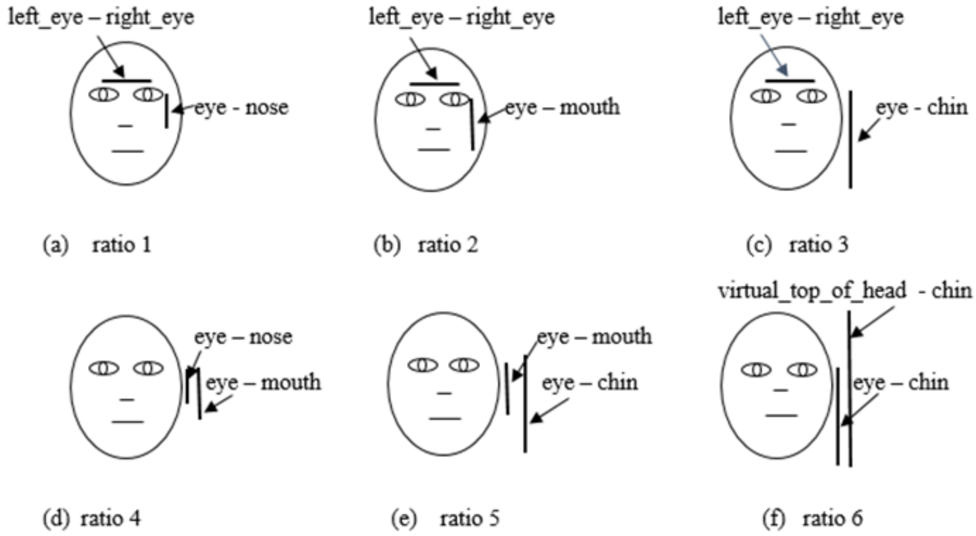

Figure 2.1: Ratios on human face.

2.3.2 Aging pattern subspace model.

This model uses a series of facial images instead of using images separately to model the process of aging. These images are sorted by time. This model is called Aging pattern Subspace it (AGES). It involves of two phases. The first phase is the learning phase, the second phase is the age estimation phase. In the first phase, the subspace representation is gained using the PCA. It differentiates from the standard PCA in that no images for each aging pattern for each year are found. For that reason, a method of iterative learning is used which is EM (Expectation-Maximization) to reduce reconstructions error. This error is the difference between the face reconstructed images and the available images of the face. In the second phase, the test image requires to discover a pattern of aging that suits the test image, and the exact estimated age of a person for the sample. The importance of the AGES is the use of images for the human at different ages to get the aging pattern (Grd, 2013) (Sarita etal, 2016).

2.3.3 Active Appearance Model.

Cootes et al. (Cootes etal, 1998) proposes the active appearance model (AAM) which is a statistical model used for facial image representation. A set of landmarks that labelled manually are used to encode the structure of the face, like facial contour, mouth, nose, eyes, and so on. Then PCA is applied to those landmarks to find the shape representation. To gain the representation of every new face the AAM models can be used for new images. The AAM can be applied to diverse tasks of face processing and it is a general method for face encoding (Sarita etal, 2016).

2.3.4 Aging Manifold

This model learns the age pattern for more than one human in diverse ages instead of learning it for each human. More than one image for each age is used to represent the age. It is flexible and better than AGES model because a number of facial images can be used for each person in single age or age range. It is also easier to gather a greater number of facial images (samples) and construct a bigger database. A manifold embedding technique is used in this model to learn a low-dimensional aging trend of the same age for several face images. The size of the sample for this model must be large enough which helps the embedded manifold to get the statistical sufficiency (Grd, 2013).

2.4 FACIAL AGE ESTIMATION DATASETS

This section describes and compares the publicly available datasets for Age estimation analysis.

There are many categories of datasets that can be used for facial age estimation such as MORPH (Ricanek etal, 2006), FG-NET (Qiu, 2016), FERET (Phillips etal, 1996), CACD (Chen etal, 2014), YGA (Fu etal, 2007), (Guo etal, 2009), PAL (Minear etal, 2004), HOIP (Fukai, 2011), (Fu etal, 2010), WIT (Ueki etal, 2006), IAD (Ni etal, 2009), AI &R (Fu etal, 2010), (Fu etal, 2006), LHI (Suo etal, 2007), and Iranian Face Database (Bastanfard etal, 2007).

Nevertheless, the vast majority of these datasets are not openly accessible. The publicly accessible datasets that give a lot of face image with precise age data are: MORPH (Ricanek etal, 2006), FERET (Phillips etal, 1996), FG-NET (Qiu, 2016), CACD (Chen etal, 2014), PAL (Minear etal, 2004) and Iranian face dataset (Bastanfard etal, 2007).

The strength and limitations of the publicly accessible datasets, i.e. MORPH, FG-NET, FERET, CACD and PAL are discussed in this section.

FG-NET (Qiu, 2016) is a well-known standard dataset. It comprises pictures for infants and a progression of pictures for people during various age movement stages. Be that as it may, FG-NET pictures were gathered by scanning photographs, therefore there were extremely huge varieties in resolution, noise, background and illumination from scanner. This may create difficulties in the identifying of texture information, for example, wrinkles. Figure 0.2 displays sample images of one person in the FG-NET database.

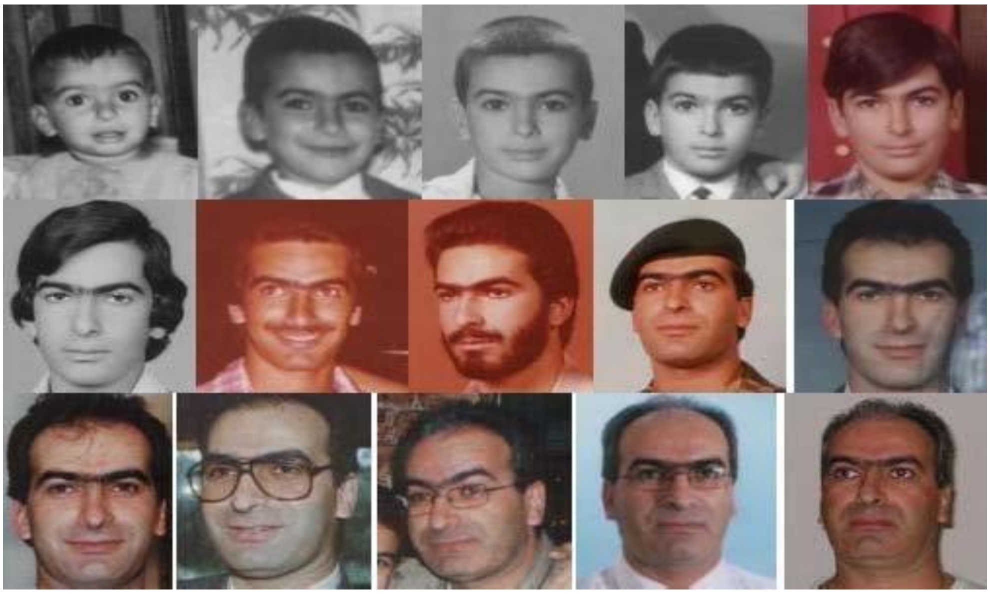

Figure 0.2: Sample images from the FG-NET and MORPH datasets (Qiu, 2016)

MORPH (Ricanek etal, 2006) is a benchmark dataset that has 55,134 images, a comparatively large-scale data-set whenever contrasted with FG-NET. It was an assortment of mugshot images, containing information about ethnicity, date of birth, gender and date of acquisition. 42,589 pictures are African, the remaining images are Asian, European and Hispanic.

FERET (Phillips etal, 1996) dataset was gathered in a controlled situation and the resolution of image is better when contrasted with MORPH and FG-NET. It has supplementary representative texture information that works better for local features extraction. Figure 0.3 displays sample images for FG-NE, FERET and MORPH datasets.

Figure 0.3: Samples of datasets a) FG-NE, b) FERET and c) MORPH.

PAL (Minear etal, 2004) dataset covers a wide age range for adulthood. It includes pictures for a subject with most extreme age of 93 years old, which is significant for research on aging. The pictures were taken with acceptable resolution.

CACD (Chen etal, 2014) is the biggest freely accessible database that was utilized recently in face recognition, age estimation and retrieval across age research, it consists of pictures for a huge number of superstars that were gathered from the web across ten years. Though, this dataset excludes face pictures with age of 10 or more youthful. Despite the fact that web is a decent stage for information assortment, it is difficult to mark the ground truth old enough or the ground truth age probably will not be exact. Furthermore, the data that gathered from the web could be noisy (Osman etal, 2018). Figure 0.4 shows sample images from the publicly available benchmark databases

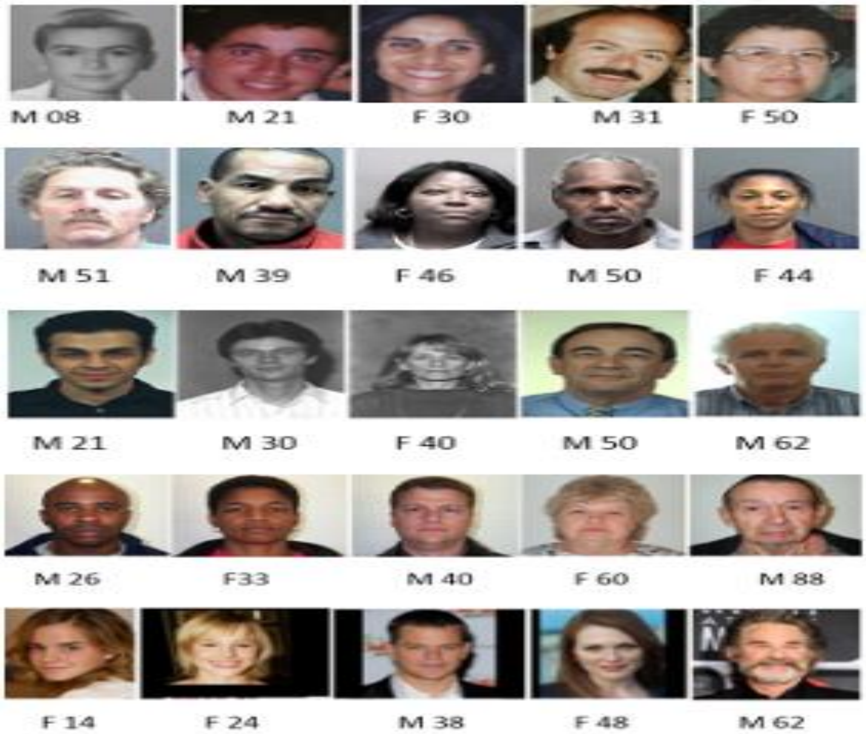

Figure 0.4: Sample face images from the publicly available benchmark databases. Row 1: FG-NET (Qiu, 2016); Row 2: MORPH (Ricanek etal, 2006); Row 3: FERET (Phillips etal, 1996); Row 4: PAL (Minear etal, 2004); and Row 5: CACD (Chen etal, 2014). These databases are diverse in pose, lighting, expression, and image resolution (Osman etal, 2018).

Despite the fact that CACD and MORPH datasets have a huge number of pictures, PAL and FERET with superior resolution, these data sets do not cover face pictures in childhood. Further constraints of the current datasets are the consideration of various ethnicities, for example, LHI YGA, and AI&R are databases comprised of Asian subjects and WIT covers just Japanese. Then again, catching face pictures at various ages for a similar subject is a major challenge. Moreover, it can be noticed that there is an absence of standardization and consistency in data collection. For example, the quantity of pictures per subject and the age range (Osman etal, 2018).

In a number of databases, for example, MORPH, YGA and LHI, just one or few pictures for each subject is caught, whereas FG-NET comprises 10 pictures for every subject, WIT catches 1 - 14 picture samples from every person, and AI&R data-set incorporates restricted numbers of face pictures. As stated by Guo et al. (Guo etal, 2012), datasets with a narrow number of face pictures of each subject probably won't be ideal for the age development algorithms. Despite the fact that there are various datasets for age estimation, a big database with balanced distribution of individuals, age range, gender and ethnicity is still required. Figure 0.4 displays the variety in diverse conditions, for example, expression, pose, makeup and lighting in openly accessible databases, and Table 2.1 gives a comparison and a summary of various face aging data-sets (Osman etal, 2018).

Table 0.1: A summary of face aging databases. It illustrates and compares the limitations of the current databases including inconsistencies on the number of images per subject, age type, age range, and ethnicity (Osman etal, 2018).

Name	Total Images	Total Subjects	Age Range	Age Type	Status
CACD (Chen, 2014),	163,446	2,000	16-62	Exact age	Public
FERET (Phillips, 1996)	14,126	1199	10-70	Exact age	Public
FG-NET (Qiu, 2016)	1,002	82	0-69	Exact age	Public
MORPH (Ricanek, 2006)	55,134	13,000	16-77	Exact age	Public
PAL (Minear, 2004)	1,142	576	18-93	Age Range	Public
YGA (Fu, 2007)	8000	1600	0-93	Exact age	Private
WIT (Ueki, 2006)	26,222	5,500	3-85	Age Range	Private
AI&R(V2.0) (Fu, 2010)	34	17	22-61	Age Range	Private
HOIP (Fukai, 2011)	306,600	300	15-64	Age Range	Private
Iranian(Bastanfard, 2007)	3,600	616	1-90	Exact age	Private
LHI (Suo, 2007)	8,000	8,000	9-89	Exact age	Private
IAD (Ni, 2009)	219,892	-	1-80	Exact age	Private

2.5 DEEP LEARNING

Deep learning is a class of machine learning techniques, in a way that several layers of information processing stages are exploited in hierarchical architectures for pattern classification and for representation or feature learning. It lies in the connections of numerous research areas, including graphical modeling, neural networks, signal processing, pattern recognition, and optimization. Deep learning’s basic concept is created from research of artificial neural network (Wan etal, 2014).

2.5.1 Methods and recent developments of Deep Learning

Deep learning method has four categories: Convolutional Neural Networks (CNNs), Autoencoder, Restricted Boltzmann Machines (RBMs), and Sparse Coding.

2.5.1.1 Convolutional Neural Networks

Convolutional Neural Network (CNN) is one of the most important deep learning approaches whose multiple layers are trained in a robust manner (LeCun etal, 1998).

The ConvNet architecture as shown in Figure 2.5 is designed as a series of stages. The first stages are consist of two layers types: convolutional layers and pooling layers. The convolutional layer units are structured in feature maps, each unit is linked to the previous layer's local patches of the feature maps through a group of weights called a filter bank (LeCun etal, 2015). A non-linearity such as a ReLU is used to pass the sum result of this local weight. All feature map units share the similar filter bank. For each layer, different feature maps use different filter banks. This architecture has twofold. Firstly, for array data like images, local values' groups are frequently more correlated, creating easily detected local motifs which are distinctive. The second reason is the invariantly of local images statistics and another signals to location (LeCun etal, 2015).

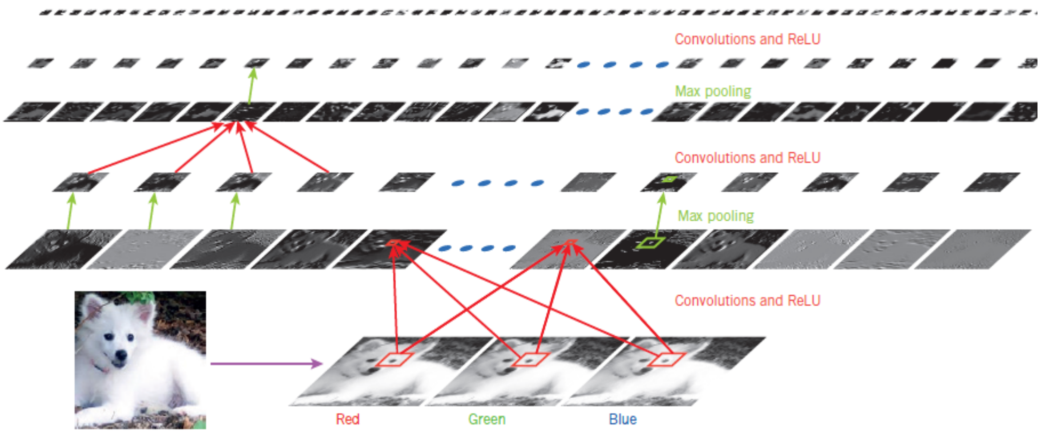

Figure 0.5: Ratios on human face (LeCun etal, 2015).

1. **Convolutional layers.**

   A CNN in convolutional layers convolve the whole image by exploiting different kernels as well as the intermediate feature maps, generating various feature maps, as shown in Figure 2.6. CNN in general is a hierarchical neural network, where its convolutional layers alternate with pooling layers. Convolution operation has three key advantages (Zeiler, 2013): 1) for the same feature map, the mechanism of weight sharing minimizes the parameters' number. 2) Correlations among neighboring are learned by local connectivity. 3) Invariance to object location. Due to this benefits, some famous research papers replace the fully connected layers by the convolutional layer to accelerate the learning process [(Szegedy etal, 2015), (Oquab etal, 2015)]. The Network in Network (NIN) method (Wan, L., 2013) is an exciting approach of the convolutional layers' handling. The main idea of it is to substitute the conventional convolutional layer with a tiny multilayer perceptron containing many fully connected layers with functions of nonlinear activation. Thus using a nonlinear neural networks as a replacement of the linear filters. This method attains good outcomes in image classification.

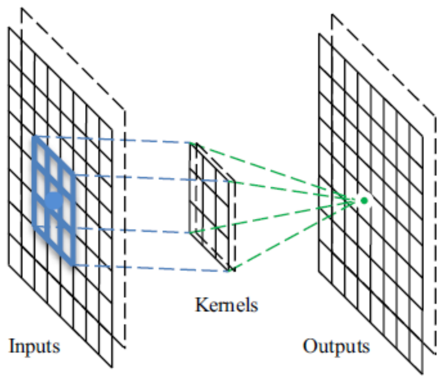

Figure 0.6: The operation of the convolutional layer (Guo etal, 2016).

2. **Pooling layers.**

   It follows the convolutional layer and it minimizes the feature maps' dimensions and network parameters. Due to taking computations of neighboring pixels into account, these layers are translation invariant. Figure 2.7 shows an example for a max pooling process.For8_8 feature maps, the output maps minimize to 4_4 dimensions, with a max pooling operator which has size 2_2 and stride2 (LeCun etal, 1998).

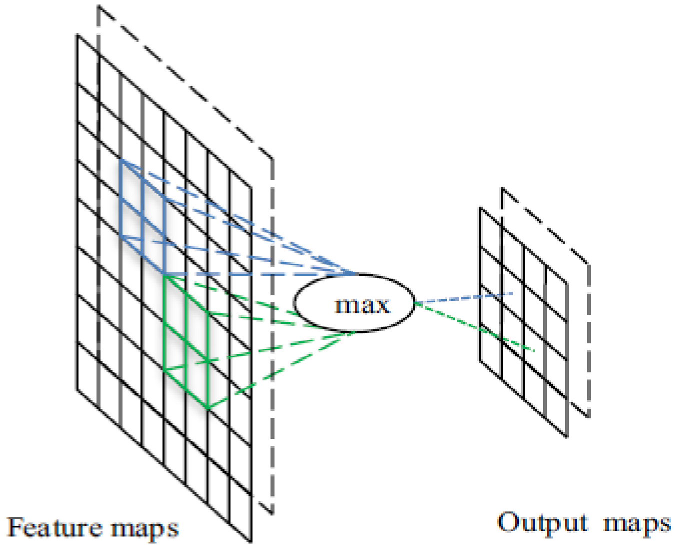

Figure 0.7: The operation of the max pooling layer (LeCun etal, 1998).

3. **Fully-connected layers.**

It follows the last pooling layer in the network. There are numerous fully connected layers which convert the 2D feature maps into a 1D feature vector, for additional feature representation, as seen in Figure 2.8. Fully-connected layers achieve like a traditional neural network and consist of about 90% of the parameters in a CNN (LeCun etal, 1998).

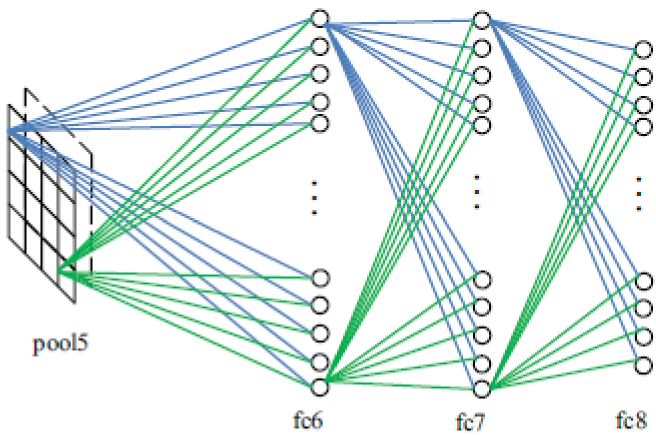

Figure 0.8: The operation of the fully-connected layer (LeCun etal, 1998).

2.5.1.2 Restricted Boltzmann Machines.

Restricted Boltzmann Machine (RBM) is a generative stochastic neural network, and was proposed .in 1986 by Hinton et al (Hinton, 1986). It is a variant of the Boltzmann Machine, with the restriction that the hidden units and visible units must form a bipartite graph. This restriction allows for more efficient training algorithms, in particular the gradient based contrastive divergence algorithm (Carreira-Perpinan, 2005). Since the model is a bipartite graph, the hidden units Hand the visible units V1 are conditionally independent. Therefore, in the equation (2.1), both Hand V1 satisfy Boltzmann distribution. Given input V1, we can get H through $\boldsymbol{P(HV_1)}$. Similarly, we can figure out V2 through $\boldsymbol{P(V_2H)}$. By adjusting the parameters, we can minimize the difference between V1 and V2, and the resulting H will act as a good feature of V1 (LeCun etal, 1998).

$$\boldsymbol{P(HV_1) = P(H_1V_1)P(H_2V_1) \ldots P(H_nV_1)} \qquad \boldsymbol{2.1}$$

Where H is the hidden unit and V_1 is the vector of the visible units

2.5.1.3 Autoencoder

It is a special kind of neural network, which used for learning efficient encodings (Guo etal, 2016). Rather than training the network given inputs X to predict some target value Y, an autoencoder is trained to reconstruct its own inputs X, thus, the output vectors and the input vector have the same dimensionality. The general an autoencoder process is shown in Figure 2.9: During the process, the autoencoder is optimized by decreasing the reconstruction error, and the learned feature is the corresponding code.

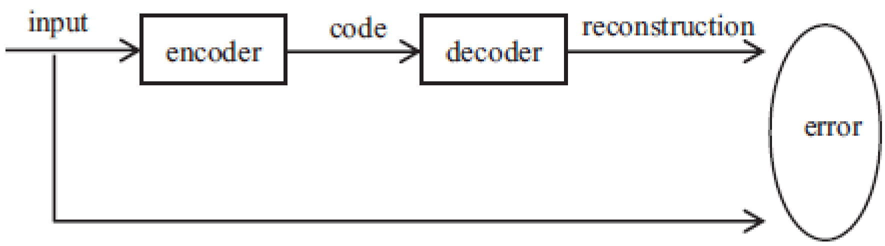

Figure 0.9: The operation of the fully-connected layer (LeCun etal, 1998).

2.5.1.4 Sparse coding.

It is used to learn an over-complete set of basic functions to describe the input data (Olshausen etal, 1997). Sparse coding has many benefits [(Yu etal, 2009)– (Yang etal, 2009)]: 1) it can reconstruct the descriptor better by using multiple bases and capturing the correlations between the same descriptors sharing bases;(2) the sparsity allows the representation to capture salient properties of images;(3) it is in line with the biological visual system, which argues that sparse features of signals are useful for learning (4) image statistics study indicates that image patches are sparse signals (5) patterns with sparse features are more linearly separable (LeCun etal, 1998).

2.5.2 A review of state-of-the-art CNN architectures

Recently, CNN has established a promising performance in face recognition. It has the skill to learn discriminative trait descriptors straight from image pixels (Liu etal, 2017a). These traits are required to correctly estimate the age of individuals. AlexNet, VGGNet, GoogLeNet, ResNet, Xception and SqueezeNet.

Generally, CNN architecture is deliberated as the greatest common architectures since of their performance on the state-of-the-art for diverse benchmarks containing age estimation task.

The following are the explanation of the architectures:

2.5.2.1 AlexNet architecture

Alexnet is one of the initial CNN architecture which has been introduced by (Levi etal, 2015) for age estimation. Alexnet, which is trained on “imageNet” database with about 1.2 million images, was logged as the first fruitful CNN architecture.

AlexNet comprises of a simple layout of eight layers with three fully-connected layers and five convolutional layers.

The CNN architecture is like that of LeNet however more profound with more filters and stacked convolutional layers. The AlexNet’s depth donated to the performance of works in (Levi etal, 2015), (Anand etal, 2017) and (Agbo-Ajala etal, 2019). It is utilized significantly to solve an extremely challenging facial analysis issue including gender recognition, age estimation and so forth. Though, more deeper models like GoogleNet and ResNet outperformed AlexNet however it is computationally costly. Figure 0.10 displays a representation of the network architecture.

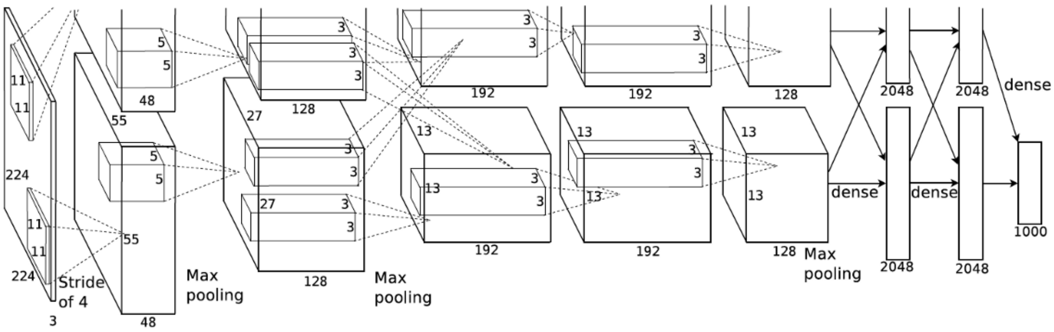

Figure 0.10: The An illustration of the AlexNet Architecture (Agbo-Ajala etal, 2019).

2.5.2.2 VGGNet architecture

An improved CNN architecture is proposed by (Simonyan etal, 2014) which is presented in Figure 0.11. VGGNet is the most well-known CNN architecture in the literature. It has a little filter trained to progressively higher profundities (16–19 layers) to achieve a state-of-the-art result on the "ImageNet" classification challenge.

The architecture is frequently utilized for transfer learning (fine-tuning), because it shows a better than expected capacity to generalize to a non-trained on datasets. VGGNet has recognized to be operative in age estimation and this is perceived in [(Rothe etal, 2015) - (Liu etal, 2017)]. Nevertheless, training VGG from scratch needs great computational power and is time-consuming; the network training is cruelly slow, and the network architecture’s weights themselves are large. The network with the Back-propagation and “fully-connected” layers, donated to its weight size. VGG architectures battle to learn on the off chance that they were arbitrarily instated and prepared from scratch; the networks were just excessively profound for essential random initialization.

With the aim of solving the above-mentioned problems, (Simonyan etal, 2014) settled a "pre-training" method that merely wanted less weight layers for training previous joining network weights as the preliminary for profounder networks.

Yet, training slighter variants of network architecture and afterward utilizing the merged weights as initializations to the more profound forms of the network, is timewasting particularly for more profound networks with numerous fully-connected layers similar to VGG; it needs training and tuning the hyperparameters to accomplish a good outcome. However, VGG architecture has verified itself to be further appropriate for simplification tasks. There is a lesser variety of VGG called "MiniVGGNet" that is very much fit for slighter datasets.

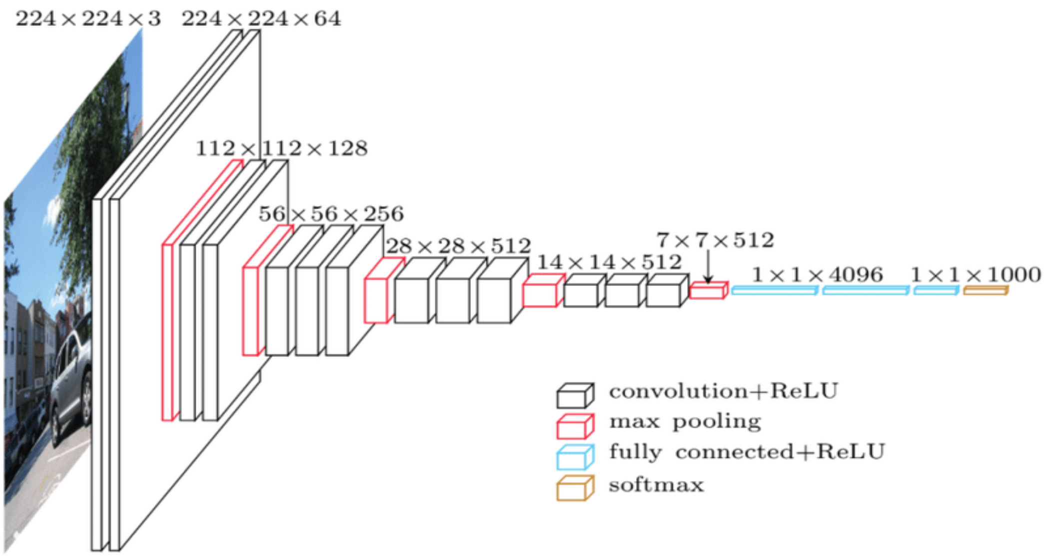

Figure 0.11: The VGGNet architecture (Agbo-Ajala etal, 2019).

2.5.2.3 GoogLeNet architecture

A GoogLeNet which is wider and deeper architecture than Alexnet has been proposed by (Szegedy etal, 2015). Its model weight is 28MB, which is lighter than VGG 19 (19-layers) and VGG-16 (16 layers). As an alternative of "fully-connected" layers established in prior architectures, global average pooling is used in GoogLeNet and this intensely decreases its weight size. Even though GoogLeNet is a small architecture when likened to VGGNet and AlexNet, it outstripped the VGG model of the

"ILSVRC" in the 2014 edition; the model creates use of a "micro-architecture" or "network in network" when building the whole "macro-architecture". An enhanced version of Inception (Inception V3) by (Szegedy etal, 2016) to supplementary increase the classification accuracy of "ImageNet" and this also influence the works' performance in (Liao etal, 2018) and (Liu etal, 2015).

Nevertheless, a basic version with lighter layers of the "inception module" can be used for a tinier dataset (with lesser image spatial dimensions) which might need a smaller number of network parameters. Figure 0.12 displays a graphic representation of the GoogLeNet network architecture.

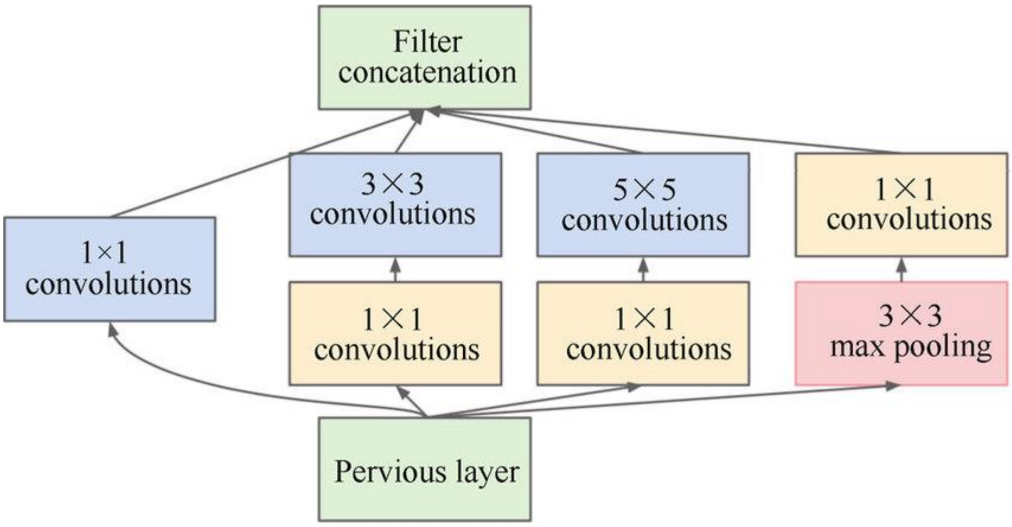

Figure 0.12: The Inception module in GoogLeNet (Agbo-Ajala etal, 2019).

2.5.2.4 ResNet architecture

To advance the performance of the current CNN architectures like AlexNet, GoogLeNet and VGGNet, (He, 2016) developed a ResNet architecture. ResNet is an exhilarating network as it presents the concept of "identity mappings" and "residual module" not in the current architecture but able to accomplish state-of-the-art results. It is profounder than VGG network but with a considerably slighter model size as a result of utilizing of "global average pooling" instead of "fully-connected" layers established in VGGNet. The ResNet's strength is the "residual module" presented by (He etal, 2016). The “residual module” contains two branches: The head is merely a shortcut that links the input to the accumulation of the second branch, and a sequence of activations and convolutions. Nevertheless, the “residual module’s extension” which is the “bottleneck”, achieves better once training a profounder network.

In addition, in the modernized publication (He etal, 2016), investigated with the convolution’s ordering, “batch normalization” and activation layers inside the “residual module” and they found that by modernizing the “residual module” to utilize “identity mappings”, higher accuracy could be gotten (Zhang etal, 2017). Though, ResNet is computationally costly with enormous profundity; it takes a long period to train. There are further shallow varieties of ResNet similar to “ResNet-10”, “ResNet-34”, and “ResNet-18” which will likewise generalize on fewer demanding task however won't get as high a correctness as the profounder ones.

It is tenderly utilized when classifying a more difficult and challenging task of facial age estimation. Figure 0.13 a, b are the architecture’s representations.

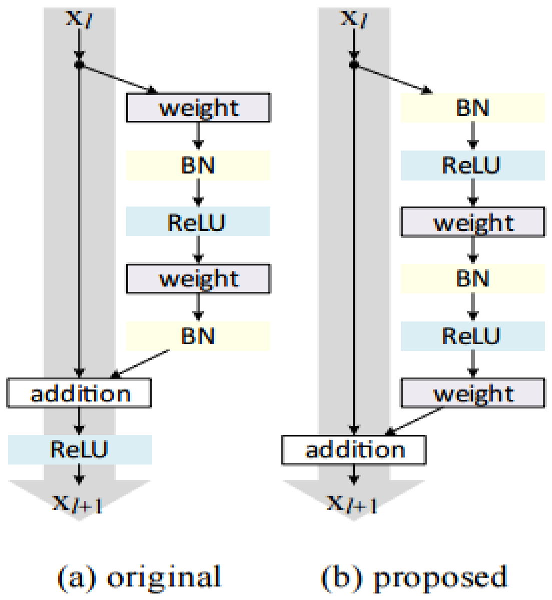

Figure 0.13: The original and updated residual module in ResNet architecture (Agbo-Ajala etal, 2019).

2.5.2.5 SqueezeNet architecture

A lighter CNN architecture has been proposed (Iandola etal, 2017) which is frequently used once we want a petite footprint. It is a little network when equated with GoogLeNet, VGGNet, AlexNet and ResNet. The model weight is 4.9M and it can more be compressed to 0.5MB.

SqueezeNet is frequently used when networks require to be trained and then deployed over a “network” or/and to “resource-constrained” devices. The network can be trained with a lessening in the parameters’ number and still achieve a high level of accuracy. It uses “fire module” that depends on a reduced and expansive phase of only 3 × 3 and 1 × 1 convolutions. The module diminishes the size of the spatial volume in the network by its comparatively small amount of filters and the attendance of “global average pooling”. “Global average pooling” operated in place of “fully-connected (FC)” layers, eliminating the FC layers, has the additional advantage of decreasing the parameters’ number needed by the network lengthily. Though, SqueezeNet excluded in the Keras core library. As presented in Figure 0.14, the petite the architecture’s nature will have emotional impact in its performance on the generalization.

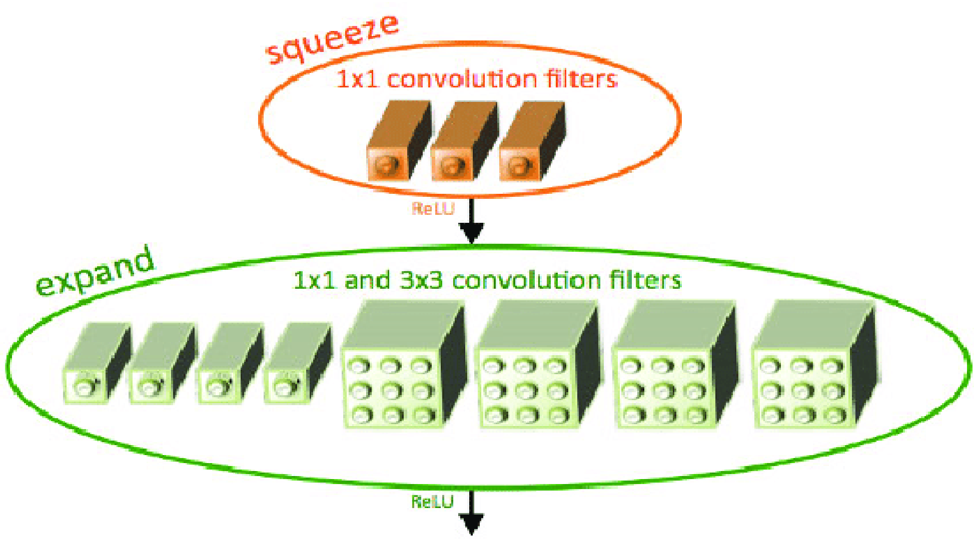

Figure 0.14: The fire module in SqueezeNet architecture (Iandola etal, 2017).

2.5.2.6 *Xception architecture*

Newly, an “Xception” (“Extreme Inception”) network has been proposed (Chollet, 2017). It is an “Inception” architecture’s extension, nevertheless the Inception module is exchanged with “depthwise separable convolutions”.

It is a CNN architecture based exclusively on “depthwise separable convolution” layers using the tiniest “weight serialization” at merely 91MB.

As displayed in Figure 0.15, the architecture of the network has 36 “convolutional layers” organized into 14 “modules”, altogether of which have direct “residual connections” around them, excluding the leading and final modules.

The architecture of “Xception” is a linear pile of layers with “residual connections” that results in making the architecture simpler to characterize and alter when contrasted with “Inception” architecture. “Xception” demonstrates a superior outcome on “ImageNet” database, not solitary on “Inception V3” but then again on “ResNet-101”, “ResNet-50” and “ResNet-152” (He etal, 2016). Nevertheless, “Xception” is slightly leisurelier than “Inception” modules.

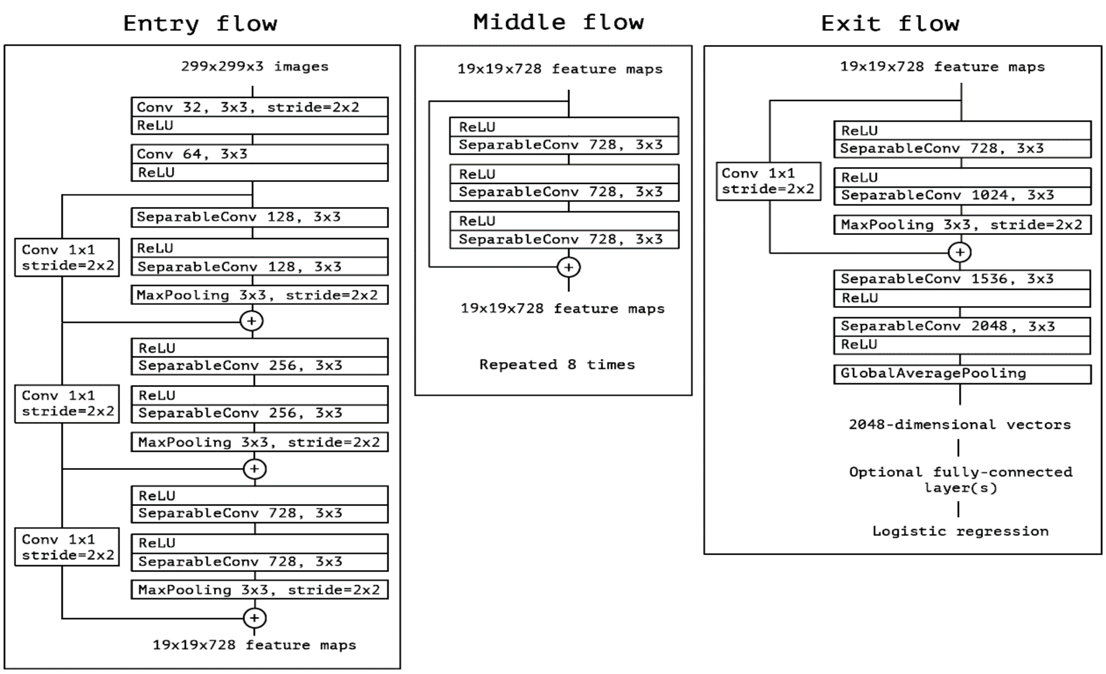

Figure 0.15: The Xception architecture (Agbo-Ajala etal, 2019).

2.5.3 TRANSFER LEARNING

Transfer learning is a significant tool to resolve the basic problem of inadequate training data in machine learning. It transfers the knowledge from the source domain to the target domain. This will result in a great positive effect on many areas that are challenging to improve due to insufficient training data (Tan etal, 2018).

The learning process of transfer learning is displayed in the Figure 2.16. In another way, transfer learning can be defined as follows: Assumed that there is a source domain DS with an equivalent source task TS and a target domain DT with an equivalent task TT . Then transfer learning is defined as the method of improving the target predictive function fT(.) relying on the associated information from TS and DS, where TS , TT or DS , DT (Weiss etal, 2016).

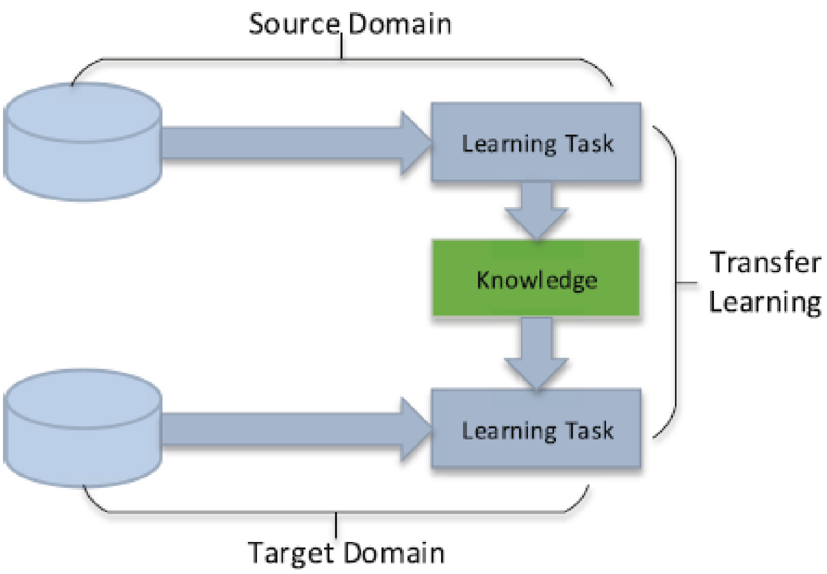

Figure 0.16: Learning process of transfer learning (Tan etal, 2018).

2.6 BAYESIAN OPTIMIZATION

Bayesian optimization is the process of discovering the minimum of a function f(x) on some bounded set X. It constructs a probabilistic model for f(x), which is exploited to provide decisions round where will be the next function evaluation in X, whereas integrating out uncertainty. The major philosophy is to utilize all of the offered information from previous evaluations of f(x) and not only depend on local gradient and Hessian Approximations. This resulted in a technique that can obtains the least possible of difficult non-convex functions through proportion to tiny evaluations, at the cost of functioning more computation to choose which the next point to try.

When evaluations of f(x) are costly to accomplish as is the state when it requests training a machine learning algorithm as a result it is easy to warrant some further computation to make better decisions. For extra overview of the Bayesian optimization and a previous works review, refer to (Brochu etal, 2010).

Algorithm: Bayesian Optimization

1: **for** n= 1,2,..., **do**
2: select new x_{n+1} by optimizing acquition function α
$x_{n+1} = arg_{max}\alpha(x; D^n)$
3: query objective function to obtain y_{n+1}
4: augment data $D_{n+1} = D_n(x_{n+1}, y_{n+1})$
5: update statistical model
6: **end for**

Mathematically, taking into account the problem of determining a global minimizer (or maximiser) of an unrecognized objective function f (Brochu etal, 2010)

$$X^* = \mathrm{argma}x_{x \in X}\ \mathrm{f}(x) \qquad 2.2$$

Where X is carefully chosen design area of interest; surrounded by global optimization, X is frequently a compact subset of Rd however the Bayesian optimizations framework can be stratified to additional unusual search spaces that include categorical or conditional inputs, or even combinatorial search areas with many categorical inputs (Shahriari etal, 2015).

There are two main selections that must be made when performing Bayesian optimization. A prior over functions must be chosen firstly that prompt assumptions around the function wanted to be optimized. Gaussian process prior, is tractable and flexible. Then an acquisition function must be selected. This function is used to construct a utility function from the model posterior, allowing us to decide which the next point to be evaluated is (Snoek etal, 2012).

2.6.1 Gaussian Processes

The Gaussian Process (GP) is strength-full and suitable on the functions earlier distribution, the functions will be used here to the form $f: X \rightarrow R$. GP is defined by the quality or characteristics which is restricted in size of N points $\{x_n \in X\}_{n=1}^{N}$ Persuade an assorted Gaussian dispersed onR_n. The nth of these points is employed to be operated value $f(x_n)$, and the elegant marginalization's properties for the Gaussian distribution authorize us to compute marginal and conditionals in closed form. The support and properties of the distribution that lead to functions are indicated by a mean function and a positive definite covariance function. The support and properties of the resulting distribution on functions are quantified by a positive denite covariance function $K: X \times X \rightarrow R$ and a mean function $m: X \rightarrow R$ (Ahmed etal, 2019).

2.6.2 Acquisition Functions for Bayesian Optimization:

Supposing that this function $f(x)$ is drawn from a Gaussian process prior moreover that the observations are shaped with $\{x_n, y_n\}_{n=1}^{N}$ where v and $y_{n \sim N(f(x_{n,v})}$ is the noises variance proclaimed into the functions observations. These prior and data yield a posterior over functions; the acquisition function, which is denoted by $a: X \rightarrow R^{+}$, describes which point in X advised to be assessed next over a proxy optimization $x_{next} = argmax_x a(x)$, where several varied functions have been proposed. In general, these acquisition functions depend on the previous observations as well as the GP hyper parameters; this dependency is signified as $\sigma(x; \{x_n, y_n\}, \theta)$. Acquisition functions have a numerous popular options. Under the Gaussian process prior, these functions depend on the model exclusively above predictive variance function $\mu\ (x; \{x_n, y_n\}, \theta)$ and its predictive mean function $\sigma^2(x; \{x_n, y_n\}, \theta))$. In the proceeding, the best current value is signified as $x_{best} = argmin_{xn} f(x_n)$ in addition to the cumulative distribution function of the standard normal as $\emptyset(.)$ (Ahmed etal, 2019).

2.6.3 Probability of Enhancement

One instinctive strategy is to increase the improving possibility above the best current value. This can be calculated systematically under the GP as (Ahmed etal, 2019):

$$\sigma_{P1}(x, \{x_n, y_n\}, \theta) = \emptyset\big(y(x)\big), \quad y(x) = \frac{f(x_{best}) - \mu(x, \{x_n, y_n\}, \theta)}{(\sigma, \{x_n, y_n\}, \theta)} \qquad 2.3$$

2.6.4 Expected Improvement

Else, one could decide on to increase the expected improvement (EI) over the current best. This correspondingly has closed form under the Gaussian process (Ahmed etal, 2019):

$$\sigma_{E1}(x, \{x_n, y_n\}, \theta) = \sigma\Big(x, \{x_n, y_n\}, \theta)(\gamma(x)\emptyset\big(\gamma(x)\big) + N(\gamma(x); 0,1)\Big) \qquad 2.4$$

2.6.5 GP Upper Condence Bound

An extra recent development is to exploit lower condence bounds (upper, when dealing with maximization) to build acquisition functions that reduction regret over the course of their optimization (Srinivas etal, 2009). These acquisition functions have the form (Ahmed etal, 2019)

$$\sigma_{LBC}(x, \{x_n, y_n\}, \theta) = \mu(x, \{x_n, y_n\}, \theta) - k\sigma\,(x, \{x_n, y_n\}, \theta) \qquad 2.5$$

2.7 GENDER CLASSIFICATION

Automatic gender classification is a significant for a lot of applications like targeted advertisements, surveillance etc. It is a task of differentiating between females and males based on the human’s features (Trivedi etal, 2009). Females and Males have a variable type of appearance aging pattern; this leads to age differently (Fu etal, 2010) (Guo etal, 2008). This is caused by the difference in beard and moustache in males and makeup, hairstyle, accessories females.

To solve the gender classification problem (Narvekar etal, 2020) numerous techniques use physical appearance as the classification's input. Physical appearance incorporates facial features such as the cheeks, eyes, lips, ears, nose, forehead, hair and the lower and mid body parts, for example, hands, stomach area, legs and so forth. Many study papers have facial features as the classification issue's input. Recently, PC has become well known and picking up significant consideration in distinguishing proof of ethnicity of human faces, age and gender, for that reason image processing have a major function in computer learning fields (Mohammed etal, 2017). While diagnosing gender there are some recognizable features that exist among male and female which are utilized by computerized strategies to categorize gender (Mohammed etal, 2017).

2.8 SUMMARY

This chapter presented a background about age estimation. Then different age estimation models for facial images have been displayed. Also all datasets that used by researchers in age estimation field are shown. Finally two concepts that used by this research are presented which are: Deep Learning and Bayesian Optimization.

CHAPTER THREE

RELATED WORK

3.1 INTRODUCTION

This chapter presents some related works that have been done in the same research area. Also a comparison of some of the published methods for age estimation have been displayed.

3.2 FACIAL AGE ESTIMATION

In recent years, many efforts have been devoted to age estimation study for the human images (Grd, 2013). A good starting point to explore estimating ages from facial images is the survey paper on Human Age Estimation using Face Images by Petra GRD (Grd, 2013). Facial aging is divided into two stages in age estimation, the first one is from birth to adulthood and the second is the time from the end of growth to old age, in which most changes are in the appearance and the shape of the face in the first phase as shown in Figure 3.1 while the main changes in the second one are changes in skin texture because the skin becomes darker, thinner more leathery and less elastic as displayed in Figure 3.2.

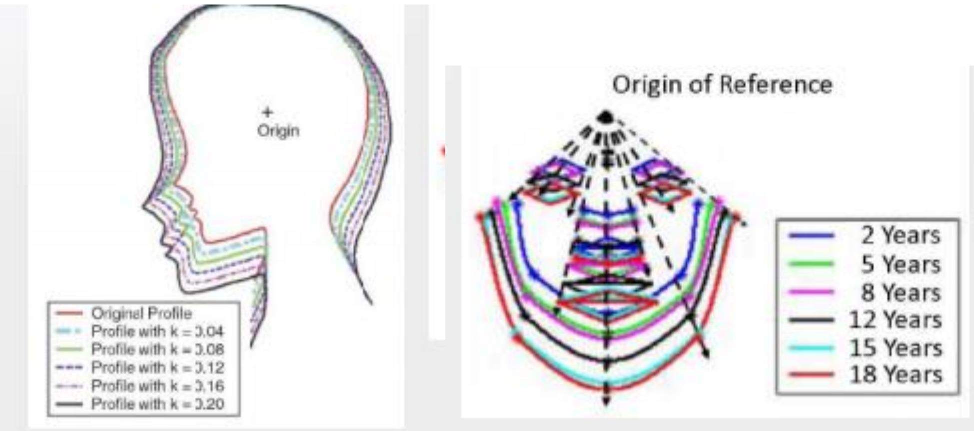

Figure 3.1: Birth to adulthood stage (Fu etal, 2010)

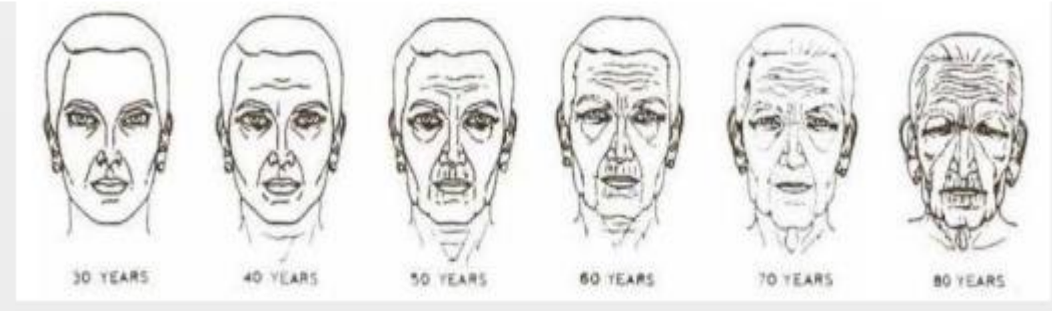

Figure 3.2: From the end of growth to old age (Fu etal, 2010)

During craniofacial growth the size of the face becomes larger, and during the adult aging the changes of shape still continue, but with fewer dramatically due to typical patterns in tissue and skin (Sarita etal, 2016). Facial representation has many models, which are: Anthropometric Model, Active Appearance Model (AAM), Aging Pattern Subspace and Age Manifold. Anthropometric Model consult studies in craniofacial theory, which uses a mathematical model to define the persons head growth from infancy to adulthood. While the active appearance model is a statistical model for facial image based on the principal component analysis. An aging pattern is declared as a series of personal facial images for the same individual. At Age Manifold common pattern or aging, the trend can be studied for many persons at different ages. The major two phases in the age estimation process are feature extraction and classification. Intensity encoding features, Bio-inspired features, Local Binary Pattern (LBP), Linear Discriminant Analysis (LDA) etc. are the common methods for feature extraction (Kamra etal, 2016). Feature extraction is the most significant part of the age estimation process. A great number of features extracted lead to having lower Mean Absolute Error (MAE) for researchers. The classification process contains two phases, training and testing to predict ages. There are three categories of facial features which are local features, global features and hybrid features. Global features consist of the whole characteristics of individual like gender, identity, ethnicity and expression along with characteristic of aging. Global features give well information about person characteristics such as the shape and the appearance of the face compared by the wrinkles and skin.

Local features contain features like hair, wrinkles and skin. Local features do better with age groups than the individuals (Ahmed etal, 2017). For that reason they classify facial images into age groups. Hybrid feature is a combination of local and global features. They were proposed using a hierarchical face model through constructing a number of features containing AAM, hair, wrinkles, skin and face configuration features.

There are three methods that can be used to learn the aging function which are: regression, classification, and hybrid (Guo etal, 2012). Classification is used when each age label is considered as one class. Many classifiers can be used such as SVMs, KNNs and artificial neural networks (ANN). Age estimation also can be viewed as a regression problem because the age labels are continuous numbers. To learn the aging function several regressors can be used such as a multiple linear regressor and the quadratic function for regression. Also to know which is better using regression or classification a comparison can be done on multiple datasets empirically and take advantages of both (Grd etal, 2016).

The formulation of suitable metrics for evaluating the performance of age estimators is a significant aspects of the age estimation problem (Lanitis etal, 2010). The Mean Absolute Error (MAE) is the most commonly used error metric between actual and estimated ages of faces in a test set. The use of the cumulative score (CS) also is proposed by (Geng etal, 2007). CS presents the percentage of cases in the test set that its age estimation error obtained is less than a threshold. It is considered as a more representative measure relative to the age estimator performance. When the age-group estimation is considered, for performance evaluation purpose a percentage of correct age-group classifications can be used (Lanitis etal, 2010).

Recently available datasets that can be used for classification or age estimation are: MORPH Database, Fg-net Aging Database, AI-R Asian Face Database, LHI Face Database, YGA Database, HOIP Face Database, PAL database, WITDB Database, Burt’s Caucasian Face Database, Iranian Face Database, Gallagher’s Web-Collected Database and Ni’s Web-Collected Database (Guo, 2012).

An overview of the method of a new database creation for classification and age estimation is provided (Grd etal, 2016).

3.2.1 Hand-crafted feature extraction methods:

The new proposed database ageCFBP can be also used for gender classification or face recognition. When comparing the Fg-net database with ageCFBP, the results show that the ageCFBP database has more subjects while the Fg-net database has a wider age range. An approach is suggested for age estimation which uses biometric ratio and wrinkles analysis for facial images. The methodology is independent of race and color by using the ratio instead of distance. This shows that it is easier to predict the lower age group than the older ones. This approach has limitation such as its failure when the forehead or other wrinkle areas are covered with hair (Ali etal, 2015).

A fast age and gender estimation system for facial images is developed (Tin, 2011). Principal Component Analysis (PCA) method and geometric feature based method are used to extract Features. Then a KNN classifier is used for classification phase. The results illustrate that better age estimation and gender classification.

Various methods for age estimation are investigated. And a new model is proposed for age estimation using a group of features that extracted from facial images through many algorithms which are active shape model, local binary patterns (LBP) and the histogram of oriented gradients (HOG) (Reade, 2015). Three age groups are used to estimate ages: child, adult, senior. In classification face K-nearest neighbor (KNN), gradient boosting tree (GBT) and support vector machine (SVM) are used. The FG-NET aging database is used to evaluate the proposed model. The results obtained show that 82\% success rate is gained when using the GBT classifier.

Principal Component Analysis (PCA) is used to predict age (Tin, 2012). 7 different groups for ages have been divided for facial images from 10 to 60 years old. The feature extraction is done through principal component analysis (PCA) method and the geometric feature-based method. The results indicate that the systems performance in age prediction is 92.5\%. There is a limitation in predicting ages for people wearing glasses, have disappearance of wrinkles and make modification of eyebrows by makeup.

Biologically inspired features (BIF) has been investigated to estimate human age from facial images. Gabor filters pyramid have been used as previously in bio inspired models at all image points for the S1 units (Guo etal, 2009). They found that using prelearned prototypes of S2 layer which is then developed to C2 did not do well for estimating ages. Gabor filters are used with smaller sizes. A new operator named STD has been proposed to encode the faces ages. The evaluation is done based on two databases YGA which has 8,000 facial images and FG-NET database which is public and available. The results indicate that their method improved the accuracy of age estimation over the state-of-the-art methods.

A system for estimating age in groups is proposed (Horng etal, 2001). The system is designed for the gray-scale images. They categorized all images into four groups: babies, young adults, middle aged adults, and old adults. The system process consists of three phases which are location then extracting features and finally classifying ages. They determine the position of noses, mouth, and eyes using region labeling and the operator of Sobel edge based on the gray levels variation and the human faces symmetry. Then they obtained three wrinkle features and two geometric features from the faces. Finally classification is done by neural networks. Two classifiers are used, the first one used the geometric features to decide if the image belonged to a baby. If the facial image did not belong to a baby, then the another classifier uses the wrinkle features to categorize the facial image to one of the other three groups. The experiments are made on a database containing 230 facial images. 115 of the images used for training and the other images used for testing. Time used for classifying image was 0.235 second. The experiment results show that the accuracy of the system is 81.58%.

A novel framework for estimating ages is developed based on the manifold analysis of facial images (Fu etal, 2007). They apply the learning methods of manifold to get a sufficient embedding space. Also they use a function of multiple linear regression to solve the low-dimensional manifold data. The results indicate that the framework was very effective. The experiments were on a large dataset of images.

A system based on the method of the bio-inspired features (BIF) for age estimation is presented (El Dib etal, 2010). BIF has been combined to the Active Shape Model (ASM). Then fine features from facial images are extracted to make the experiments. The MORPH and FG-NET benchmark databases are used for evaluation purpose. Their algorithm gives higher accuracy than other methods. Their future work will be in using the Active Appearance Model and Gabor functions.

A framework is presented for facial age evaluation centered on joint of single age evaluators (Choobeh, 2012). Experimental and mathematical evidence shows, if the individual age evaluators are varied in error, and so to advance the outcomes the ensemble age estimator can be made using the best selected individual age estimators. It is accentuated that despite the neural networks done to show the experiments, the suggested context is readily relevant to any other regressor.

A classified method is suggested for estimating age automatically and an investigation of how aging affects single facial mechanisms is offered (Han etal, 2013). Investigational outcomes on the FGNET, MORPH Albums, and PCSO databases display that nose and eyes are more revealing than other facial mechanisms in estimating age automatically. Human observation capacity for estimating age is assessed using gathering obtained data gained via the Amazon Mechanical Turk service and paralleled with the presentation of the suggested automatic age estimation. The presentation of the planned age estimation approach is shown to be better than or equaled to the age estimates delivered by individuals on FG-NET and a less subsection of PCSO database record. However a paralleled presentation has been reported to the finest recognized outcome on MORPH Album2 and organize it so without looking at the benefits of the real demographic information delivered with the database, yet the combination of per-component age estimator showed no benefits on the MORPH Album2 and PCSO databases. The future work will be done through investigation approaches to develop our accuracy by automatically detecting key point.

Two novel approaches have been presented in this paper (Huerta etal, 2015). The first one is a simple effective descriptors fusion based on local appearance and texture. And the second one is a scheme of deep learning to estimate accurate age. The two approaches under multiple settings are evaluated and the experiments are done on two big databases, which are FRGC and MORPH. Experiments show better results compared with the previous works.

A system has been implemented to prevent children from accessing the materials or contents of adults on the internet and stop young children from buying cigarettes, alcohol, etc. (Sarita etal, 2016). All the images are preprocessed and filtered to get good results. Viola-Jones algorithm is used to extract features. And the KNN classifier is trained and tested through the extracted features in the final stage. The results show that the K-NN classifier works better than other classifiers for the age-group prediction. Their future work will be in creating Indian Aging Database and using it.

The methodologies in each phase is analyzed which produce better results (Deepa, 2016). Discrete Cosine Transform method works better in extracting local and global features for facial images as founded. Extracting global features help in recognizing the age group classification. HOG, LBP and DCT Mod2 methods are used in feature extraction stage. Then a comparison is done against each other. Results indicate that DCT method provides better accuracy. The key challenge was finding the best features combination (angle, distance and ratio).

Age Estimation method named Global and Local feature based Age estiMation (GLAAM) is introduced which rely on local and global features (Günay etal, 2015). Active Appearance Models (AAM) is used to obtain the global features while regional 2DDCT (2- dimensional Discrete Cosine Transform) is used to extract the local features. After that global and local features of the images are combined together. The FG-NET aging database is used in the experiments. Results display that GLAAM works better than existing methods. GLAAM can be improved by using Principal Component Regression, as it combines Regression and PCA in the same stage.

For age feature extraction a new framework is built which is based on the deep learning model. They use the convolutional neural network (CNN) to estimate ages (Wang etal, 2015). As a comparison between their model and existing models, feature maps gained from different layers are used for age estimation as an alternative of using a feature gained from the top layer.

Furthermore, for their proposed scheme they incorporate the algorithm of manifold learning. This significantly increases the performance. Also the deep learned aging pattern (DLA) is used to evaluate different regression and classification schemes for age estimation. Their results based on two datasets indicate that their approach is significantly better than the state-of the-art.

In human age estimation, a novel kernel method to create statistics of local binary pattern is proposed for facial images (Ylioinas etal, 2013). Their main work contributions consist of evaluating a pose correction method by using a simple image flipping, and using facial representations to compare two local binary patterns. The two local binary patterns are called a novel kernel density estimate and a spatially enhanced histogram. The experiments result in cross- and single-database shows that that the kernel density estimate produces better accuracy in age estimation than the spatially enhanced histogram. The constructed system for age estimation gives better performance beside the state-of-the-art methods.

A method for subjective age estimation is proposed (Miyamoto etal, 2005). They use facial images of people and real age. A rating scale for the faces is experimented. Subjects are simulated through it. The image is evaluated as seeing older than the real with a range of response. The experiments results indicate that a subjective age turns to be in negative direction (estimating facial image age as younger than the real age). In the future they plan to test objective age which is the age defined by others as it appears to them.

A new method based on divide-and-conquer is proposed for facial images which named fusion of multiple binary age grouping estimation systems (Liu etal, 2016). Firstly a several binary grouping systems are employed to classify age groups. All images have been classified to the two predefined groups. For age estimation purposes they trained two models. The influence of diverse age grouping systems performance is investigated for age estimation error and age grouping accuracy. A sequentially selection algorithm is proposed to find the last result of age estimation. They use MORPH2 database in their experiments and results indicate that the proposed framework achieves satisfying results.

It is shown that such aged shapes can be successfully removed from a discriminant subspace knowledge procedure and pictured as separate assorted constructions over the various process of inquiry on face pictures, the dimensionality repetition of the new picture space can apparently be less with subspace learning (Fu etal, 2008). A manifold lined reversion method, exactly with a quadratic ideal purpose can be simplified by the little dimensionality to display the diverse space symbolizing the discriminative assets. A treating like this has been assessed by wide imitations and matched with the state-of-the-art procedures. Investigational outcomes on big scope aged database prove the value and strength of our future framework.

Accuracy of binary classification has been improved by (Olsson, C.) when using the Kwon-Lobo papers features and two novel features. Combining wrinkle analysis enhanced results of 96.

A preprocessing stage for all used images is made to strengthen their system (Tao, 2014). This stage includes detection of nose and eyes. Gabor filter and local binary patterns (LBP) are used to extract features for age estimation. Histograms from
LBP analysis is used for feature classification stage. FG-NET aging database is trained and results of the images trained give an accuracy of 39.8%.
An effective system of age estimation is presented for facial images (Dabi etal, 2014). The system uses 2D-Gabor filter for feature extraction. Then Multi linear Principle Component Analysis (MPCA) is used to compress the extracted features. For classification phase the K-NN classifier assigned the input image into one of the age groups. All ages are categorized into 10 groups. A large Indian database for facial images has been developed to find the exact age. The database consists of 750 Indian images, their ages between 0 and 74. They tried to gain maximum Indian images at different ages.

A Multi-scale Wrinkle Patterns (MWP) representation has been introduced by (Ng etal, 2018) to explore the impact of wrinkles on facial age estimation and a Hybrid Aging Patterns (HAP) has been proposed for face age estimation. To characterize the wrinkle areas all the more definitely, a format comprising of 10 locales developed moderately to a lot of located facial landmarks automatically is utilized. The multi-scale wrinkles in every area have been extricated and encode them into MWP. Support Vector Regression is used to assess age from the blend of such patterns. The algorithms' performance is evaluated via utilizing Mean Absolute Error (MAE) on three datasets – FERET, MORPH, and FG-NET. MWP is observed that it yields a close MAE of 4.16 on FERET to the state of the art. As a final point HAP is proposed, which consolidates the Facial Appearance Model (FAM) and the highlights from MWP. It shows improved performance on MORPH and FERET with MAE of 3.68 (±2.98) and 3.02 (±2.92), respectively.

A Multi-layer Age Regression (MAR) is proposed where the age of facial image is anticipated dependent on a coarse-to-fine estimation utilizing local and global features (Ng etal, 2017). Support Vector Regression (SVR) in the first layer makes a between bunch expectation by the Facial Appearance Model's parameters (FAM). In the subsequent layer, an inside gathering assessment is performed utilizing FAM, Kernel-based Local Binary Patterns (KLBP), Bio-Inspired Features (BIF), and Multi-scale Wrinkle Patterns (MWP). The MAR's performance is evaluated on four benchmark datasets: PAL, FERET, MORPH and FG-NET. Results demonstrated that MAR achieves a Mean Absolute Error (MAE) of 3.00 (_4.14) on FERET, which outperforms the state of the art.

3.2.2 Deep Learning based methods:

Recently, deep learning algorithms have been applied by a number of researchers to face related tasks like face verification, gender identification and age estimation. A Deep ID structure is proposed to extract discriminative features from the face for face verification process (Sun etal, 2014). To improve the Deep ID algorithm, a verification constraint is added in loss function to obtain better performance (Sun etal, 2014). For detecting landmark points of the face, a cascaded Deep ConvNets structure is proposed (Sun etal, 2013). Also (Zhang etal, 2014) proposed a new algorithm to detect landmark points, which named deep multi-task learning algorithm.

A framework for age estimation based on deep learning is proposed. Transfer learning is used due to the lack of labeled images. Due to the ordered labels in age estimation, a new loss function for age classification is defined through distance loss addition to cross-entropy loss for relationships description between labels. Results obtained prove the excellent algorithm performance against the state-of- the-art methods (Dong etal, 2016).

A robust and a fast age modeling algorithm are used with the deep learning to propose age estimation system. They indicate that the local regressors performance for most groups are better than the global regressor. Samples are firstly classified into overlapping age groups. Local regressors estimates the apparent age for each group. The outputs are used for the final estimate. The system is evaluated on the ChaLearn Looking at People 2016 – Apparent Age Estimation challenge dataset, and results in 0.3740 normal score on the test set (Gurpinar etal, 2016).

The largest public IMDB-WIKI dataset with gender and age labels is introduced by (Rothe etal, 2018). VGG-16 architecture for convolutional neural networks is used which are pre-trained on ImageNet dataset. A robust face alignment is done. They study the perceived age by other humans and the apparent age estimation. The methods are evaluated on standard benchmarks and results achieve state-of-the-art for both apparent and real age estimation.

An End-to-End learning method has been proposed (Niu etal, 2016) to solve the problems of ordinal regression utilizing deep Convolutional Neural Network, which could at the same time conduct regression modeling and feature learning. Particularly, an ordinal regression issue is changed into a sequence of binary classification sub-issues. Also, a multiple output CNN learning algorithm is proposed to jointly address these classification sub-issues, so the relationship between these chores could be investigated.

Likewise, an Asian Face Age Dataset (AFAD) is published holding more than 160K facial pictures with exact age ground- truths, which is the biggest free age dataset to date. Apparently, this is the main work to solve ordinal regression issues by utilizing CNN, and accomplishes the state-of-the-art performance on both the AFAD and MORPH datasets.

Both final label encoding and the structure innovation are explored (Qiu, 2016) for the performance evaluation, two tasks of computer vision are conducted. For the first task, a novel hierarchical aggregation is proposed based on deep network to study features of aging and their encoding method is applied to transmit the discrete aging labels to a possibility label, this allows the CNN to conduct a classification task rather than regression task. Their deep aging feature can capture both global and local cues in aging. Experimental results of age prediction on the FG-NET and the MORPH-II databases indicate that the proposed deep aging feature outperforms state-of-the-art aging features.

To overcome the issue of solid invariance of the model brought about by datasets with different sale of information, a multi-path CNN model was projected by (Liu etal, 2017) offered a "Gathering Aware Deep Feature Learning" (GA-DFL) strategy for facial age approximation. The required features for face description are extracted by learning a “discriminative feature descriptor” straightforwardly from the raw pixels.

In order to smoothen the neighboring age groups, they presented a covered coupled learning technique. They additionally utilized a "multi-way" profound CNN design to coordinate various scale data into the educated face presentation which enhanced the performance of the strategy. They surveyed the adequacy of the proposed strategy on three freely accessible datasets on facial age approximation that were acquired in both controlled and uncontrolled conditions and it accomplished a superior performance when contrasted to most state-of-the-art facial age approximation techniques (Liu etal, 2017).

An ordinal deep learning approach (Liu etal, 2017) has been proposed for facial age estimation to study descriptors directly for face representation from raw pixels. ODFL imposes two descriptors’ standards which have been learned at the deep networks’ top: the first one is engaged to feat the order information in the erudite feature space, and the second one is leveraged to vigorously measure face twosomes with diverse age value gaps. Nevertheless, the two techniques of age estimation and feature extraction are found out autonomously in ODFL, which may prompt imperfect issue. To solve this issue, an end-to-end ordinal deep learning (ODL) framework has been proposed, where the corresponding data of the two methods is abused to strengthen their model. Broad experimental outcomes on five facial age databases indicate that both ODL and ODFL accomplish greater performance in contracts with most state-of-the-art techniques.

A novel CNN approach using Bayesian Optimization (DLBO) has been proposed for facial age estimation (Ahmed etal, 2019). It is the first time that Bayesian optimization is utilized with Deep Learning in the facial age estimation field. Bayesian Optimization minimized the classification error on the validation set for CNN model. The experiments’ result on three datasets: MORPH, FG-NET and FERET, show that DLBO outperforms the state of the arts on FG-NET and FERET datasets with a Mean Absolute Error (MAE) of 2.88 and 1.3, and achieves comparable results compared to the most of the state-of-the-art methods on MORPH dataset with a 3.01 MAE.

Recently, (Agbo-Ajala etal, 2019) proposed a CNN-based model to categorize unconstrained realistic face images into gender and age. The methodology included an algorithm for image preprocessing that readies the input images and furthermore a CNN construction that makes the extraction of features and image classification to gender and age group. The experimental outcomes were assessed on the OIU-Adience dataset, and it affirms the efficiency of their methodology, outperformed different studies on the identical dataset.

(Nam etal, 2020) solved the age estimation’s issue of low- resolution facial pictures with a profound CNN-based model that recreate faces with low-goal as faces with high-goal. The CNN-based arrangement comprises of a provisional generative adversarial network (GAN) that pre-handled low-resolution facial pictures before utilized as input. The model at that point utilized the state of the art CNN network architecture like DEX, ResNet, and VGG, for the age estimation of the recreated facial pictures. The test results on MORPH, FG-NET and PAL databases, exhibit the proposed strategy technique efficiency in high resolution recreation. It accomplishes state-of-the-art results about age estimation of low- resolution pictures.

Further, a lightweight CNN network (ShuffleNetV2) has been developed by (Liu etal, 2020) in view of the mixed attention mechanism (MA-SFV2). The model; MA-SFV2-ShuffleNetV2, changes the output layer, that model age estimation as a regression issue (that rank the age of the human face having a specific request), a classification issue (that characterize age as an isolated label), and distribution learning (that study the age connection between neighboring ages). The model incorporates picture pre-processing that decreases the impact of a data augmentation strategy and noise vectors like histogram enhancement, sharpening, filtering and so on that upsurge the size of image and reduce the network’s over-fitting. For the age estimation task, the model consolidates regression, classification and distributed learnings. The exploratory outcomes on FG-NET and MORPH-II datasets, demonstrate the model’s applicability in real-life circumstances, particularly in mobile terminals.

3.2.3 Comparison of Some of the Published Methods for Face Age Estimation

A comparison is presented in Table 3.1 for some of the existing and recent algorithms for age estimation problem. The table displays that the existing methods either use local features like wrinkles or represent faces with a holistic representation. The table discusses the techniques and database used in different papers. Classifier, results, Procs and Concs are also shown. As mentioned, using wrinkles to estimate age increases the performance of the systems, but sometimes these systems fail because it gets higher edge density when there is hair on the forehead, which the system perceives as wrinkles (Ali etal, 2015). And this provides wrong results for ages that have been estimated. The FG-Net dataset which is public was used by most of the researchers for the experiment results purposes [(Han etal, 2013), (Deepa etal, 2016), (Günay etal, 2015), (Wang etal, 2015), and (Olsson, C.). Another database; which is one of the largest Indian facial image datasets is developed; gives 80 percentage as determined as high result, but ages are estimated in groups only and also this dataset did not have large numbers of facial images of people at different ages (Dabi etal, 2014). Deep learning technique have been introduced and used for the first time for age estimation purpose (Wang etal, 2015) and this increases the performance as compared by the state of the arts. The two age estimation algorithms used are classification and regression. For classification, K-nearest neighborhood and SVM are the common methods. Still there are research gaps to improve accuracy in age estimation.

Table 0.1: A Comparison of Some of the Published Methods that used traditional hand-crafted methods for Age Estimation.

Ref.	**Techniques**	**Database used**	**Classifier**	**Results**	**Pros**	**Cons**
(Ali etal, 2015)	Biometric ratios, Wrinkle analysis	N/A	K-NN, Decision Tree, Nave Bayesian	Better recall And precision for middle ages.	wrinkle increases the performance	Gets higher edge density for hair in forehead.
(Han etal, 2013)	Holistic BIF features, eye region BIF Fusing	The FG-NET, MORPH Album2, and PCSO.	SVM classifiers	On FG-NET the MAE is 4.7 while on the PCSO data the MAE is 7.2	different features for different age groups improve the performance	The accuracy varies when using a big subset of PCSO databases.
(Huerta etal, 2015)	LBP, Speeded-Up Robust, Features(SURF), and HOG	MORPH dataset, FRGC dataset	(Canonical Correlation Analysis(CCA)	Early fusion of HOG, LBP and SURF improves the MAE score 4.25 years	Deep analysis of the different parameters for feature detectors	Need larger images for their experiments which is not available
(Deepa etal, 2016)	PCA, LBP, HOG and Discrete Cosine Transform (DCT)	The FG-NET database	Neural network, KNN, AGES, SVM	DCT method provides better accuracy	Feature normalization enhances the images contract	Lies in identifying the best combination of the features
(Günay etal, 2015)	Active Appearance Models (AAM), 2-dimensional Discrete Cosine Transform	The FG-NET database	Multiple linear regression	Achieves better results than earlier methods on the FG-NET	Use global and local features of facial	Did not use methods that do not require normalization such as SIFT and ASIFT.
(Fu etal, 2008)	Manifold representation, Conformal Embedding Analysis	the UIUC-IFP database	Quadratic regression	CEA shows superiority for most of dimensionality reduction cases.	CEA increase the performance	The results depend only on regression
(Olsson, C.)	Wrinkle analysis, Feature based on face shape.	The FG-NET aging database	SVM regression	binary classification accuracy improved (97.84\% vs. 92.94\%). 96 \% total accuracy	Wrinkle analysis pushed classification accuracy higher	Resolution of images are not high.

(Dabi etal, 2014)	2D-Gabor filter, Multi Linear Principle Component Analysis (MPCA)	Indian Facial Image, Georgia Tech Databases	K-NN Classifier	Near about 80\% result	Develop one of the largest Indian facial image database	Estimate age in groups, facial image of people at different age are few.
(Wang etal, 2015)	Deep learning techniques, manifold learning	MORPH,FG-NET databases	SVMs, SVR, PLS and CCA	MAE obtained on MORPH is 4:77.	first time that deep learning is applied for age estimation	---
(Ahmed etal, 2019)	Deep Learning, Bayesian Optimization	MORPH,FG-NET, FERET databases	Deep Learning	MAE on MORPH is 3.01, on FERET is 1.3 and FG-NET is 2.88	The first time that deep learning is applied with Bayesian Optimization.	---
(Niu etal, 2016)	CNN, ordinal regression	MORPH , AFAD database	Deep Convolutional Neural Network	Achieves the state-of-the-art performance on both the MORPH and AFAD datasets.	address the non-stationary property of aging patterns - publish an Asian Face Age Dataset (AFAD)	ignored to take advantages of the quadruplet-based ordinal relation during batch-wise training procedure, which makes the learned features less efficiency for age prediction

3.2.4 Summary

This chapter presented some related works that have been done in same research area. Also a comparison of some of the published methods for age estimation have been displayed. The comparison showed that there are some drawbacks in the previous works. Also all the previous works used a determined values for all the training options. Furthermore for all previous works a determined network layers are used in each time. The depth of the network can improve the performance or the opposite. This research select some variables to optimize that used in the training options. Also the network depth has been optimized to improve the age estimator performance.

CHAPTER FOUR
METHODOLOGY

4.1 INTRODUCTION

This chapter presents the methodology for two proposed models. The first model uses Deep learning with Bayesian Optimization. And the second model uses Gender Information to enhance the age estimator performance.

4.2 BACKGROUND

Machine learning algorithms are rarely parameter-free: parameters controlling the rate of learning or the capacity of the underlying model must often be specified. These parameters are frequently considered annoyances, making it interesting to develop machine learning algorithms with fewer of them.

Another good solution is to optimize such parameters as a procedure to be automated. Specifically, viewing such tuning as optimizing an unknown black-box function and appeal algorithms developed for such problems. Bayesian optimization (Fu etal, 2010) is a good choice, which outperforms other state of the art global optimization algorithms on many challenging optimization benchmark functions (Sarita etal, 2016). The parameters of the machine learning algorithms that control the capacity and learning rate of the underlying model must be identified. These parameters are frequently considered annoyances, making it interesting to develop machine learning algorithms with fewer of them.

In this research, Bayesian Optimization is used to improve the performance of age estimation. Bayesian Optimization minimizes the classification error on the validation set. Then the impact of gender information on age estimation is investigated.

4.3 BAYESIAN OPTIMIZATION WITH DEEP LEARNING

In order to train CNN, the CNN architecture and options of the training algorithm must be specified firstly. The process of choosing and tuning these parameters is difficult and takes long time. The algorithm of Bayesian Optimization (BO) is compatible to optimize internal parameters of regression and classification models. BO algorithm can be used to optimize functions that are discontinuous, non-differentiable, and time-consuming to evaluate. It maintains a Gaussian process model internally of the objective function. This objective function is used to evaluate training in this model. Bayesian optimization is applied to deep learning to find optimal training options and network parameters for CNNs. Figure 4.1 shows the stages of the proposed model. The first stage is the Data acquisition which is done by getting the publicly available datasets from the internet. The second stage is the data pre-processing which contains three steps: face alignment, face detection and cropping, and finally the data augmentation. The third stage is applying Bayesian Optimization with Deep Learning which will be illustrated in details.

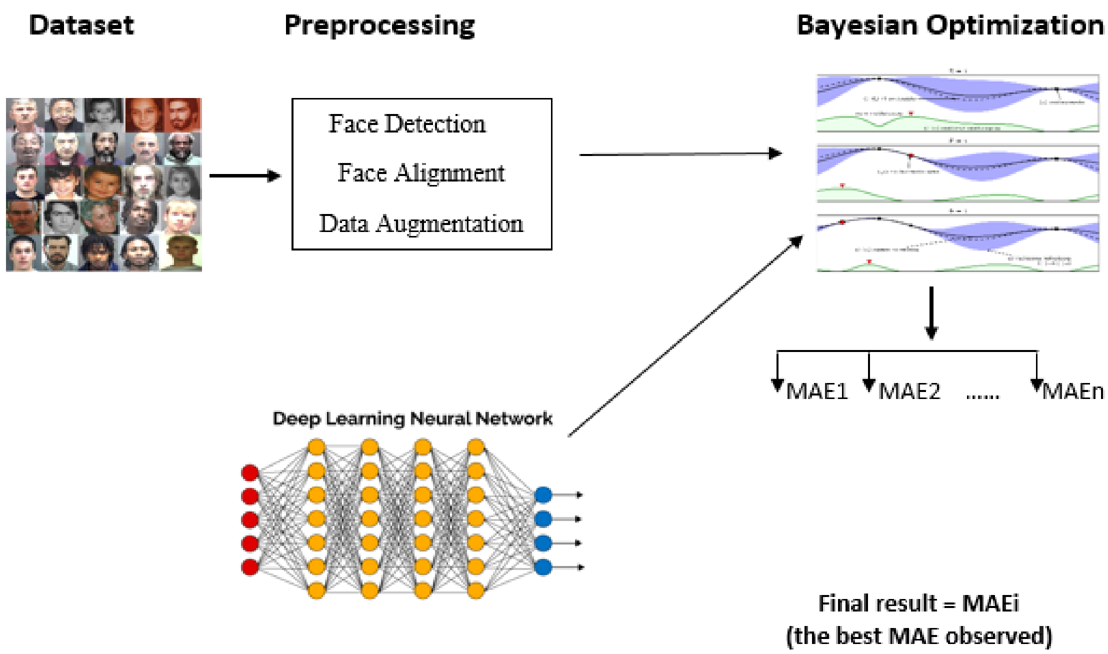

Figure 0.1: The Proposed Model

To apply Bayesian Optimization with deep learning there are some steps to follow:

1. Dataset Pre-processing for Network Training.
2. Specify the variables to be optimized using Bayesian optimization.
3. Define the objective function.
4. Perform Bayesian optimization
5. Load the best network.

4.3.1 Dataset Pre-processing

At the beginning, all the images are aligned. For the alignment stage, the input coordinates which are the facial landmarks are used to warp and transform the image to an output coordinate space. The alignment depends on the position of the two eyes for all images. The alignment is done for MORPH only and failed for FG-NET and FERET datasets. Figure 4.2 shows an input sample that is used to get an output aligned image depending on the coordinates of the two eyes marked with red color.

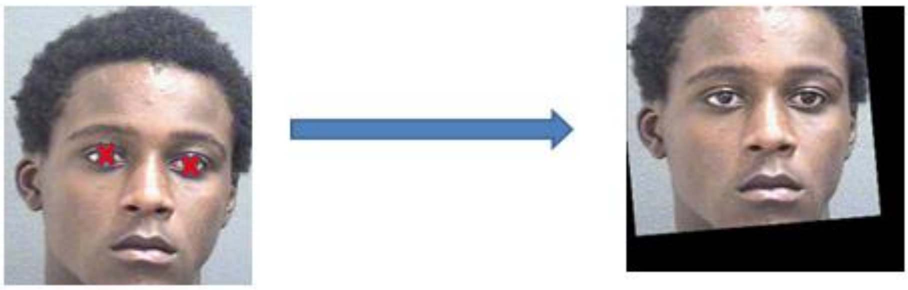

Figure 0.2: Sample Images before and after Image Alignment.

After the face alignment, a face detection process is done using face++ Detector (Tao, 2014) to extract the face only without hair and other features that could be a noise. Figure 4.3 displays how the detection and cropping process is done.

Figure 0.3: Sample Images before and after Detection and cropping.

Figure 4.4 shows the output of the pre-process stage on a sample image for the three datasets. The MORPH dataset has three images the original image, aligned image, and detected image respectively. While the FG-NET and FERET datasets have two images the original image and the detected one. Finally all of the detected faces are resized to 32x32 to fit the network for training. Except for MORPH dataset, the data have been resized to 110x110.

Figure 0.4: Sample Images before and after Preprocessing.

In this research, Bayesian Optimization is used to improve the performance of age estimation. The brief illustration is shown in Figure 4.5. The variables that required to be optimized are specified firstly. These selected variables are the parameters of the network architecture, as well as options of the training algorithm. Then the objective function is defined, which receipts the values of these specified variables as inputs. This function specifies the training options and network architecture, training the network on training set and validating it on the test set. Then the Bayesian Optimization is performed with several objectives by minimizing the classification error on the validation set. Finally, the best network is loaded from the disk and evaluated on the test set.

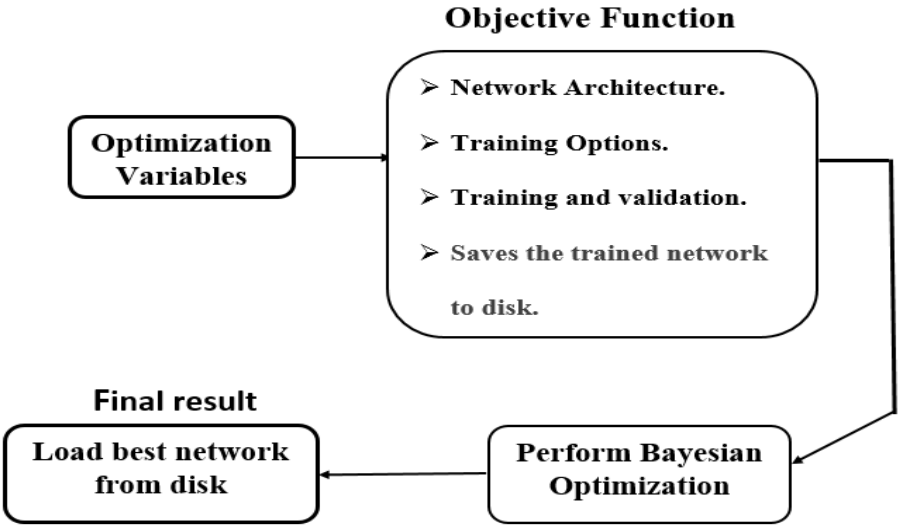

Figure 0.5: Deep Learning with Bayesian Optimization.

Every stage in the Deep learning with Bayesian Optimization are illustrated in details latter.

4.3.2 Select Variables for Optimization

The variables that selected to be optimized are: The Network Depth, Momentum, Initial Learn Rate, and L2Regularization as shown in Table 4.1. The Network depth controls the depth of the network. L2 regularization strength is used to prevent over-fitting. Batch normalization and data augmentation and batch normalization also help regularize the network. The best initial learning rate can depend on the data and the network used. These variables are the parameters of the network architecture as well as options of the training algorithm.

Table 0.1: The selected variables for optimization.

Network Depth	Momentum	Initial Learn Rate	L2Regularization
[1 – 3]	[0.8 – 0.95]	[1e-3 1e-1]	[1e-10 1e-2]

4.3.3 Objective Function for Optimization

The objective function that includes the Bayesian optimization's variables and the CNN architecture is defined. Its inputs are the values of the optimization variables that are illustrated in Table 4.1 as well as the training and testing data. This function states the network architecture that will be used for training and specifies the training options. Our Network consists of six convolution layers, which depends on Network variable of Bayesian. Then training and validating this network is involved in this objective function. For every time that the training is done, the objective function saves the trained network to disk.

4.3.4 Bayesian Optimization

Bayesian optimization is performed, which minimizes the classification error on the validation set. The objective function trains a convolutional neural network and returns the classification error on the validation set. It uses the error rate to choose the best model. The final selected model is tested on the test set to estimate the Mean Absolute Error (MAE). The maximum number of objective function evaluations is specified by 25, and the maximum total optimization time is set to eight hours for FG-NET and FERET datasets. All networks are trained on a single Graphical Processing Unit (GPU) (NIVIDIA GeForce 840M) with 2GB RAM.

This research uses MAE to evaluate the performance of the age estimation model. MAE is used to measure the error between the ground truth and predicted age, which is computed as (Ahmed etal, 2019):

$$\epsilon = \frac{1}{N} \sum_{i=1}^{N} |y_i' - y_i| \qquad 4.1$$

Where y_i and y_i' are predicted and ground-truth age value, respectively, and N indicates the number of the testing samples.

4.4 USING GENDER INFORMATION FOR FACE AGE ESTIMATION (ENHANCED DLBO)

Automatic age estimation for facial images has a series problem since the determined age of humans vary based on a lot of aspects which may be internal factors such as gender, genetic, etc. (Fu etal, 2010).

Males and Females have varying types of face aging patterns. This results in age differently (Lin etal, 2012), (Sarita etal, 2016). This is caused by the difference in beard and moustache in males and hair style, makeup, accessories in case of female. These above facts lead to assume that using gender information may improve the age estimator performance (Kamra etal, 2016).

In this research, Gender Information is used to improve the age estimation performance. A CNN is trained for Gender Classification. After training, the resulted CNN is used as an input for age estimation model which uses Bayesian Optimization to obtain good results through minimizing classification error on validation set.

This research uses Gender Information to increase the performance of age estimation. After getting this information relying on Convolution Neural Network (CNN), Bayesian Optimization is used to select the best result. The variables that require optimization are defined firstly. These specified variables are the options of the training algorithm. At that point the objective function is well-defined, which takes the values of these specified variables as inputs. This function determines the training options. Training the outputted network from

Gender Information on training set and validating it on the test set. Then the Bayesian Optimization is executed with numerous objectives by minimizing the classification error on the validation set. As a final point, the best network is loaded from the disk and evaluated on the test set.

In order to train a CNN, the CNN architecture has been identified firstly. This CNN is trained for Gender Classification on the datasets. Options of the training algorithm must be specified firstly for age estimation process. The procedure of selecting and tuning these training parameters is challenging and time-consuming. Bayesian Optimization (BO) algorithm is suitable to optimize internal parameters of classification and regression models. It can be used to optimize functions that are non-differentiable, discontinuous, and time-consuming to evaluate. A Gaussian process model of the objective function is maintained internally, which is used to evaluate this model. Bayesian optimization is applied to the pre-trained network to find optimal training options. Figure 0.6 shows the framework of the proposed system, beginning by pre-processing images, which includes Face detection, Face cropping and face resizing. The pre-processed data is used by the CNN for gender classification purpose. BO uses the pre-processed data and the resulted pre-trained network with Gender Information. This pre-trained network is fine-tuned for age estimation task. Training and testing this model gives a number of results (n) which is equal to the objectives' number that stated in the BO. As a final point, the best result that has been achieved from these n results is chosen as the final accepted result.

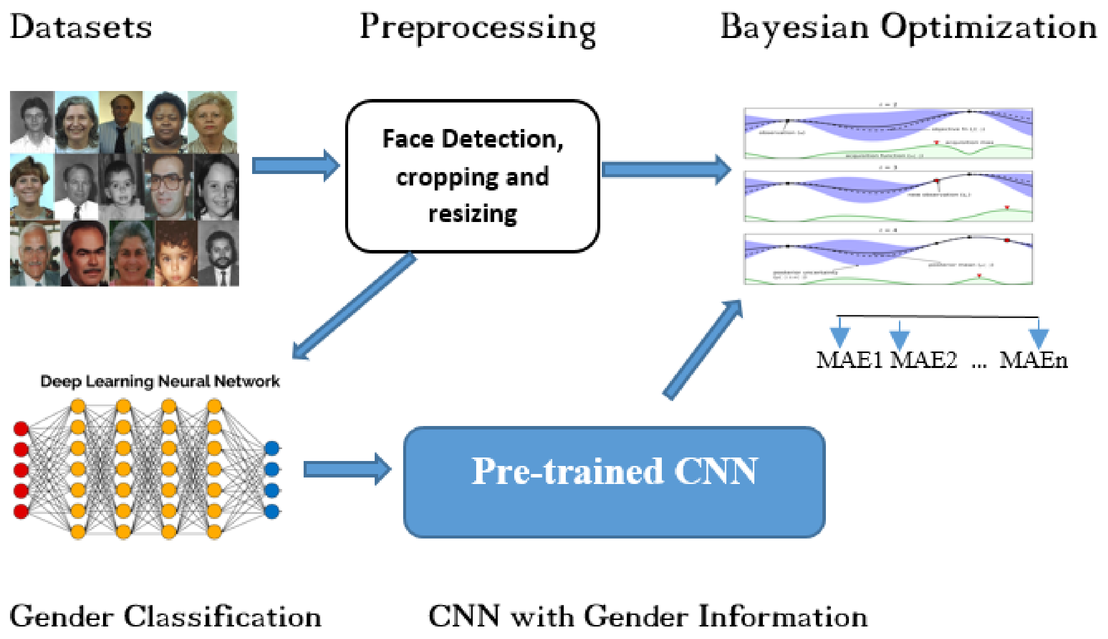

Figure 0.6: Proposed Model with Gender Information

The following steps explain the details of the proposed model.

1) Dataset Pre-processing for Network Training.
2) Training CNN to get Gender Information.
3) Specify the variables to be optimized using Bayesian optimization.
4) Define the objective function and the network architecture.
5) Perform Bayesian optimization.
6) Load the best network.

4.4.1 Dataset Pre-processing:

At the beginning, for all the images a face detection process is done using face++ Detector (Ylioinas etal, 2013) to extract the face only without hair and other features that could be a noise. Figure 4.7 shows the pre-process stage on a sample image for the two datasets. For each dataset any subject has two images the original image and the detected one. As a final point all the detected faces are resized to 80x80 to fit the network for training.

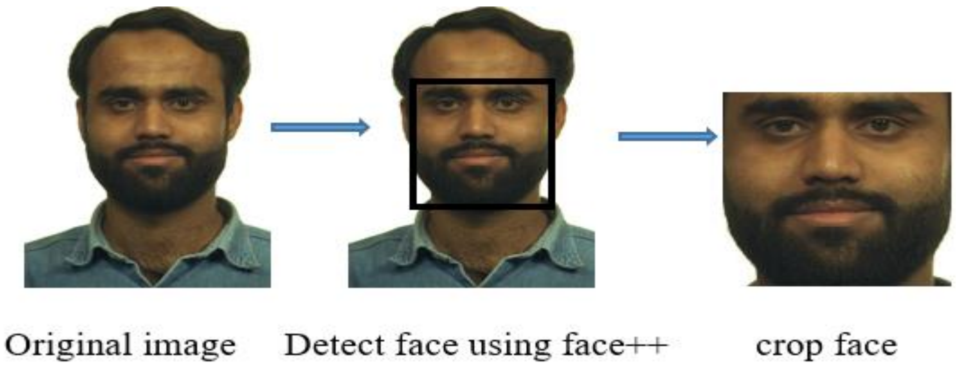

Figure 0.7: Sample Image before and after Pre-processing II.

4.4.2 Training CNN to get Gender Information:

Our Network consists of seven convolution layers. The options of training has been specified firstly. The pre-processed facial images have been used for training to classify Males and Females. Finally, the resulted network has been saved to disk after the training completed. It has been saved to be fine-tuned for age estimation task.

4.4.3 Select Variables for Optimization:

The selected variables for optimization are: Momentum and Initial Learning Rate as shown in Table 4.2. The best initial learning rate relies on the network used and the data. These variables are the options of the training algorithm.

Table 0.2: The selected variables for optimization 2.

Momentum	Initial Learn Rate
[0.8 – 0.95]	[1e-3 1e-1]

4.4.4 Objective Function for Optimization:

The objective function is defined. It takes the pre-trained network, values of the optimization variables, and the data as inputs. This function states the training options. Also, it involves training and validating this pre-trained network. For each time that the training is completed, the trained network is saved to disk.

4.4.5 Bayesian Optimization:

Bayesian optimization is accomplished, which reduces the classification error on the validation set. The objective function trains the pre-trained network and yields the classification error on the validation set. The error rate is used to select the best model. To estimate the MAE, the last selected model is tested on the test set. The maximum total optimization time is set to eight hours for the two datasets and the maximum number of objective function evaluations is identified by 40. All the training is done on a single GPU (NIVIDIA GeForce 840M) with 2GB RAM.

4.4.6 Final Network Evaluation

The best network has been loaded and evaluated on the test set to get the MAE result for evaluating the final network purpose.

4.5 SUMMARY

This chapter has presented the methodology for two proposed models. The first model used Deep learning with Bayesian Optimization. Bayesian optimization minimizes the classification error on the validation set for CNN model. And the second model used Gender Information to enhance the age estimator performance.

CHAPTER FIVE
RESULTS AND DISCUSSION

5.1 INTRODUCTION

This chapter presents the datasets used for the experiments. Then the evaluation metric that used for the results is shown. And finally the results and discussion are illustrated. In this chapter, the performance of Bayesian optimization with deep learning was evaluated by testing its ability to estimate ages.

5.2 DATASETS

Three facial benchmark datasets are used is this experiment as shown in Table 5.1. MORPH (Ricanek etal, 2006) is the most commonly used benchmark dataset for deep learning in age estimation. . It is a group of mugshot images, containing metadata for race, date of birth, gender, and date of acquisition. It consists of 55K facial images which range from 16 to 76 for 13,000 subjects. FG-NET (Ahmed etal, 2019) is the second benchmark dataset for facial age estimation. It contains 1002 images with ages between 0-69. Nevertheless over 50% of the subjects are aged between 0 and 13. FERET (Phillips etal, 1996) dataset is used by many researchers for age estimation. It consists of 2366 facial images of 994 subjects. The age range is between 10 and 70.

Table 0.1: The Datasets used for evaluation.

Name Age	Total Images	Total Subjects	Age Range
MORPH II	55,134	13,000	[16 - 77]
FG-NET	1,002	82	[0 - 69]
FERET	2366	994	[10 - 70]

5.2.1 MORPH II Dataset

MORPH dataset is the most commonly used dataset for age estimation using deep learning. It is used in the experiments to evaluate the DLBO model comparing to the previous works' results. Figure 5.1 displays sample images from the MORPH dataset before applying any pre-processing techniques in it.

Figure 5.1: Sample Images for MORPH dataset before Pre-processing.

Then Figure 5.2 shows the same images after applying the pre-processing techniques to display the differences and the effects of the pre-processing stage in the proposed model.

Figure 5.2: Sample Images for MORPH dataset after Pre-processing.

5.2.2 FG-NET Dataset

FG-NET dataset is a benchmark dataset that is used for evaluating age estimation models. Figure 5.3 displays some randomly selected images from the dataset without applying any pre-processing tasks.

Figure 5.3: Sample Images for FG-NET dataset before Pre-processing.

While Figure 5.4 shows the same images that are displayed in the previous Figure after applying the pre-processing step.

Figure 5.4: Sample Images for FG-NET dataset after Pre-processing.

5.2.3 FERET Dataset

FERET is the third dataset used to evaluate the proposed model in this research. Figure 5.5 presents sample images from this dataset.

Figure 5.5: Sample Images for FERET dataset before Pre-processing.

Figure 5.6 presents the same images after making pre-processing techniques to show the differences before and after applying pre-processing step.

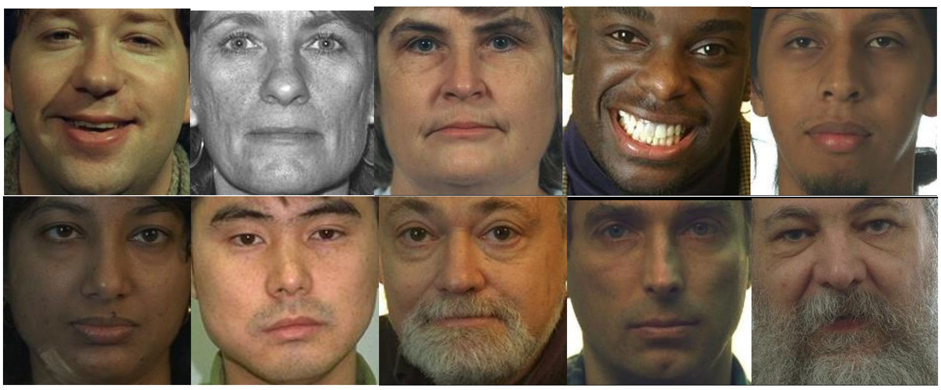

Figure 5.6: Sample Images for FERET dataset after Pre-processing.

5.3 EVALUATION METRICS MEAN ABSOLUTE ERROR

MAE has been used to evaluate the performance of the age estimation model. It is used to measure the error between the ground truth and predicted age, which is computed as (Ahmed etal, 2019):

$$\epsilon = \frac{1}{N} \sum_{i=1}^{N} |y_i' - y_i| \qquad 5.1$$

Where y_i and y_i' are predicted and ground-truth age value, respectively, and N indicates the number of the testing samples.

5.4 RESULTS COMPARISONS FOR DEEP LEARNING USING BAYESIAN OPTIMIZATION (DLBO)

For evaluating (DLBO) on the MORPH dataset, according to the settings in selected previous works on age estimation (Liu etal, 2016), (Fu etal, 2008), and (Olsson, C.), this paper randomly takes 54,362 samples of ages from 16 to 66. Then these selected samples are split into two groups: 80% of the samples used for training the network and 20% for testing. For these two sets, there is no overlapping. This paper repeats five runs independently, and the performance of the five runs has been averaged to obtain the final performance evaluation during experiments. The quantitative results are summarized in Table 5.2. DLBO has been compared with other deep learning models for age estimation. Since the experiments done on the MORPH dataset and the same setting is followed in this paper for data partition, a direct comparison of the MAE of DLBO with the ones resulting by these deep learner can be done. As displayed in this table, DLBO achieves good results compared to most of the current state-of-the-art results with 3.01 MAE.

Table 0.2: COMPARISON OF MAES WITH DIFFERENT STATE-OF-THE-ART APPROACHES ON THE **MORPH DATASET**.

Deep Learning-Based Methods	**MAE**
OR-CNN (Niu etal, 2016)	3.34
MR-CNN (Niu etal, 2016)	3.27
DEX (Rothe etal, 2018)	3.25
GoogLeNet (Hu etal, 2016)	3.13
Ranking-CNN (Chen etal, 2017)	3.03
DLBO (Ahmed, M., 2019)	**3.01**

For FG-NET Leave One Person Out (LOPO) is employed for evaluation protocol as in previous works. This paper selected facial images of one person randomly for testing purpose, and the facial images of the remaining subjects for training. This procedure is repeated for 82 folds to evaluate BO. Eventually the average of the 82 folds results is approved as the final age estimations result. The quantitative results are summarized in Table 5.3. DLBO achieves superior results compared to all the current state-of-the-arts. And this is the first time that a MAE error below 3.0 is obtained on FG-NET dataset.

Table 0.3: COMPARISON OF MAES WITH DIFFERENT STATE-OF-THE-ART APPROACHES ON THE **FG-NET DATASET.**

Deep Learning-Based Methods	**MAE**
DEX (Rothe etal, 2018)	4.63
Ranking-CNN (Chen etal, 2017)	4.13
GA-DFL (Liu etal, 2017)	3.93
DRFs (Liu etal, 2017)	3.85
ODL + OHRanker (Liu etal, 2017)	3.89
ODL (Liu etal., 2017)	3.71
DLBO (Ahmed etal, 2019)	**2.88**

For FERET dataset a 10-Fold cross-validation is performed as in (. Liu etal, 2017). Specially, the whole dataset is divided into ten folds with equal size for each fold. Nine folds are used for training and the remaining fold for testing. This process is repeated ten times and the final age estimation result is the average of the ten results. The quantitative results are summarized in Table 5.4. As presented in the table, DLBO achieves superior results compared to all the previous state-of-the-art results on FERET. Note that, it is the only paper using deep learning with FERET and has a MAE error below 2.0 on FERET dataset.

Table 0.4: COMPARISON OF MAES WITH DIFFERENT STATE-OF-THE-ART APPROACHES ON THE **FERET DATASET.**

Hand-crafted Methods	**MAE**
MAP (Ng etal, 2015)	4.87
HAP (Ng etal, 2018)	3.02
MAR (Ng etal, 2017)	3.0
Deep Learning-Based Methods	**MAE**
DLBO (Ahmed etal, 2019)	**1.3**

5.5 RESULTS COMPARISONS FOR ENHANCED DLBO USIND GENDER INFORMATION IN DLBO

For investigation the impact of using gender information in DLBO, as shown in Table 5.5 the proposed model achieves superior results on the FG-NET dataset compared to all the current state-of-the-arts. As mentioned in DLBO (Ahmed etal, 2019), the Deep learning is applied with Bayesian Optimization and this results in improving the performance compared to the previous works. In this proposed model, (Ahmed etal, 2019) is extended through using DLBO with Gender Information to test if using Gender Classification can improve age estimation. The result is displayed in Table 5.5 for the enhanced model (Proposed-2 (Ahmed etal, 2020)) indicates that using Gender

Information has improved the age estimation model by increasing the performance from 2.88 to 2.67.

Table 0.5: COMPARISON OF MAES WITH DLBO ON THE FG-NET DATASET.

Deep Learning-Based Methods	**MAE**
Proposed-1 (DLBO (Ahmed etal, 2019))	**2.88**
Proposed-2 (Enhanced DLBO Ahmed etal, 2020)	**2.67**

For the FERET dataset, the quantitative results for using gender information to improve the performance of the age estimator are summarized in Table 5.6. As shown in Table 5.6, the enhanced model (Proposed-2 (Ahmed etal, 2020)) achieves superior results compared to all the prior state-of-the-art results on FERET. Compared to DLBO (Ahmed etal, 2019), using Gender Information increases the performance by decreasing the MAE from 1.3 to 1.2.

Table 0.6: COMPARISON OF MAES WITH DLBO ON THE FERET DATASET.

Deep Learning-Based Methods	**MAE**
Proposed-1 (DLBO (Ahmed etal, 2019))	**1.3**
Proposed-2 (Ahmed etal, 2020)	**1.2**

5.6 SUMMARY

This chapter has presented the datasets used in the experiments and the evaluation metrics. Then the results of all the experiments are displayed with the comparison to the previous works. Results show that the two proposed models outperform the stat-of-the-arts and enhanced the age estimator performance when using Bayesian optimization and gender information.

CHAPTER SIX

CONCLUSION, RECOMMENDATIONS AND FUTURE WORK

6.1 CONCLUSION

Age estimation is an important task in facial image classification. It is the process of determining the exact age or age group of a person using his biometric features. Many known world applications deal with facial age estimation directly or indirectly such as, security control, multimedia communication, human computer interaction, and Surveillance.

Currently, Aging affects the human face appearance. Skin related deformations and bone movement and growth affect the facial aging related to reduction of muscle strength and the introduction of wrinkles. Humans can estimate the age of other people when looking at their faces. However, age estimation's researchers conclude that the ages estimated by humans are not accurate .Hence the need of improving methods for automatic facial age estimation takes an attractive direction.

In this research existent and recent researches in the age estimation field have been discussed. A lot of researchers made a contribution and are still working to optimize this field. All the state-of-the-art techniques for age estimation are presented. Along with the state-of-the-art techniques, age estimation models for facial representation in order to extract useful features are described: anthropometric model, active appearance model, aging pattern subspace and age manifold. There are several challenges in order to implement a robust system for age estimation. Main challenges which may motivate novel investigations in the future have been identified.

The main aim of this research is to specify the parameters' values that controlling the network architecture and the training options to develop the deep learning methods that used in age estimation. Also, using gender information to enhance the proposed model.

Bayesian optimization is applied to deep learning for the first time in age estimation field. It is used to select the optimal training options and network parameters for CNNs. The experiments result show that Bayesian Optimization with deep learning obtains good results compared with the most state-of-the-arts: outperforms the state-of-the-arts on FG-NET and FERET datasets with a MAE of 2.88 and 1.3 respectively. And achieves good results to the state-of-the-art methods on MORPH dataset with MAE of 3.01.

Also, the impact of Gender Classification in age estimation has been investigated. Gender Classification is applied on deep learning to get benefits from the Gender Information on age estimation field. Then Bayesian Optimization is used for the resulted pre-trained network to select the optimal training options for Age estimation. The experiments' results show that using gender information with Deep learning and Bayesian Optimization achieves good results associated with the state of the arts: outperforms the state of the arts on FERET and FG-NET datasets with a MAE of 1.2 and 2.67 respectively. Also compared to DLBO (Ahmed etal, 2020), the experiments illustrate that using gender information improve the age estimation task.

6.2 RECOMMONDATIONS AND FUTURE WORK

Future work on the proposed model based on Deep Learning with Bayesian Optimization (DLBO) involves using images with larger size and pre-trained the network in WIKI-datasets before applying the DLBO on the benchmarks datasets. Also to ensemble this classifier with other CNNs classifiers that obtains good results to improve the performance of age estimation system. Also further works can include evaluating the proposed model on the datasets focusing on the implication of race and color on the age estimation. This model can also be used with dead people for criminal investigations.

For the impact of Gender Classification investigation, future works involve evaluating this model on MORH dataset which needs better Graphical Processing Unit (GPU), using images with larger size, images alignment and data augmentation.

REFERENCES

Ahmed, M. and Viriri, S., 2017, November. Age estimation using facial images: A survey of the state-of-the-art. In *2017 Sudan Conference on Computer Science and Information Technology (SCCSIT)* (pp. 1-8). IEEE.

Ahmed, M. and Viriri, S., 2019, August. Deep Learning Using Bayesian Optimization for Facial Age Estimation. In *International Conference on Image Analysis and Recognition* (pp. 243-254). Springer, Cham.

Ahmed, M. and Viriri, S., 2020, December. Facial Age Estimation using Transfer Learning and Bayesian Optimization based on Gender Information. *Signal and Image Processing: An International Journal (SIPIJ) Vol.11, No.6.*

Ali, S.M., Darbar, Z.A. and Junejo, K.N., 2015, February. Age estimation from facial images using biometric ratios and wrinkle analysis. In *2015 5th National Symposium on Information Technology: Towards New Smart World (NSITNSW)* (pp. 1-5). IEEE.

Agbo-Ajala, O. and Viriri, S., 2019, November. Face-Based Age and Gender Classification Using Deep Learning Model. In *Pacific-Rim Symposium on Image and Video Technology* (pp. 125-137). Springer, Cham.

Agbo-Ajala, O. and Viriri, S., 2020. Deep learning approach for facial age classification: a survey of the state-of-the-art. *Artificial Intelligence Review*, pp.1-35.

Agustsson, E., Timofte, R., Escalera, S., Baro, X., Guyon, I. and Rothe, R., 2017, May. Apparent and real age estimation in still images with deep residual regressors on appa-real database. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 87-94). IEEE.

Anand, A., Labati, R.D., Genovese, A., Muñoz, E., Piuri, V. and Scotti, F., 2017, November. Age estimation based on face images and pre-trained convolutional neural networks. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1-7). IEEE.

Anand, A., Labati, R.D., Genovese, A., Muñoz, E., Piuri, V. and Scotti, F., 2017, November. Age estimation based on face images and pre-trained convolutional neural networks. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1-7). IEEE.

Antipov, G., Baccouche, M., Berrani, S.A. and Dugelay, J.L., 2016. Apparent age estimation from face images combining general and children-specialized deep learning models. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 96-104).

Bastanfard, A., Nik, M.A. and Dehshibi, M.M., 2007, December. Iranian face database with age, pose and expression. In *2007 International Conference on Machine Vision* (pp. 50-55). IEEE

Brochu, E., Cora, V.M. and De Freitas, N., 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*.

Can Malli, R., Aygun, M. and Kemal Ekenel, H., 2016. Apparent age estimation using ensemble of deep learning models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 9-16).

Carreira-Perpinan, M.A. and Hinton, G.E., 2005, January. On contrastive divergence learning. In *Aistats* (Vol. 10, pp. 33-40).

Chen, B.C., Chen, C.S. and Hsu, W.H., 2014, September. Cross-age reference coding for age-invariant face recognition and retrieval. In *European conference on computer vision* (pp. 768-783). Springer, Cham.

Chen, S., Zhang, C. and Dong, M., 2017. Deep age estimation: From classification to ranking. *IEEE Transactions on Multimedia, 20*(8), pp.2209-2222.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

Choobeh, A.K., 2012. Improving automatic age estimation algorithms using an efficient ensemble technique. *International Journal of Machine Learning and Computing*, *2*(2), p.118.

Cootes, T.F., Edwards, G.J. and Taylor, C.J., 1998, June. Active appearance models. In *European conference on computer vision* (pp. 484-498). Springer, Berlin, Heidelberg.

Dabi, D.S. and Patil, S.B., 2014. A Robust Age Estimation System For Indian Facial Image Using 2D Gabor Filter and Multilinear Principle Component Analysis. *International Journal of Advance Engineering and Global Technology*, *2*(10), pp.67-71.

Deepa, A. and Sasipraba, T., 2016. Challenging aspects for facial feature extraction and age estimation. *Indian Journal of Science and Technology*, *9*(4), pp.1-6.

Dong, Y., Liu, Y. and Lian, S., 2016. Automatic age estimation based on deep learning algorithm. *Neurocomputing*, *187*, pp.4-10.

El Dib, M.Y. and El-Saban, M., 2010, September. Human age estimation using enhanced bio-inspired features (ebif). In *2010 IEEE International Conference on Image Processing* (pp. 1589-1592). IEEE.

Farkas, L.G. ed., 1994. *Anthropometry of the Head and Face*.

FGNET aging dataset, http://www-prima.inrialpes.fr/FGnet/, accessed on September 2012.

Fu, Y. and Huang, T.S., 2008. Human age estimation with regression on discriminative aging manifold. *IEEE Transactions on Multimedia*, *10*(4), pp.578-584.

Fu, Y. and Zheng, N., 2006. M-face: An appearance-based photorealistic model for multiple facial attributes rendering. *IEEE Transactions on Circuits and Systems for Video technology*, *16*(7), pp.830-842.

Fu, Y., Guo, G. and Huang, T.S., 2010. Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *32*(11), pp.1955-1976.

Fu, Y., Xu, Y. and Huang, T.S., 2007, July. Estimating human age by manifold analysis of face pictures and regression on aging features. In *2007 IEEE International Conference on Multimedia and Expo* (pp. 1383-1386). IEEE.

Fukai, H., Takimoto, H., Mitsukura, Y. and Fukumi, M., 2011. Apparent age estimation system based on age perception. *Self Organizing Maps: Applications and Novel Algorithm Design*, p.441.

Geng, X., Zhou, Z.H. and Smith-Miles, K., 2007. Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence*, *29*(12), pp.2234-2240.

Geng, X., Fu, Y. and Smith-Miles, K., 2010, August. Automatic facial age estimation. In *11th Pacific rim international conference on artificial intelligence* (pp. 1-130).

Grd, P. and Bača, M., 2016, May. Creating a face database for age estimation and classification. In *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1371-1374). IEEE.

Grd, P., 2013. Introduction to human age estimation using face images. *Research Papers Faculty of Materials Science and Technology Slovak University of Technology*, *21*(Special-Issue), pp.24-30.

Günay, A. and Nabiyev, V.V., 2015. Age estimation based on AAM and 2D-DCT features of facial images. *International Journal of Computer Science and Applications*, *6*(2).

Guo, G., Mu, G., Fu, Y. and Huang, T.S., 2009, June. Human age estimation using bio-inspired features. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 112-119). IEEE.

Guo, G., 2012. Human age estimation and sex classification. In *Video Analytics for Business Intelligence* (pp. 101-131). Springer, Berlin, Heidelberg.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S. and Lew, M.S., 2016. Deep learning for visual understanding: A review. *Neurocomputing*, *187*, pp.27-48.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S. and Lew, M.S., 2016. Deep learning for visual understanding: A review. *Neurocomputing*, *187*, pp.27-48.

Guo, G., Mu, G., Fu, Y., Dyer, C. and Huang, T., 2009, September. A study on automatic age estimation using a large database. In *2009 IEEE 12th International Conference on Computer Vision* (pp. 1986-1991). IEEE.

Guo, G., Fu, Y., Huang, T.S. and Dyer, C.R., 2008, January. Locally adjusted robust regression for human age estimation. In *2008 IEEE Workshop on Applications of Computer Vision* (pp. 1-6). IEEE.

Gurpinar, F., Kaya, H., Dibeklioglu, H. and Salah, A., 2016. Kernel ELM and CNN based facial age estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 80-86).

Han, H., Otto, C. and Jain, A.K., 2013, June. Age estimation from face images: Human vs. machine performance. In *2013 International Conference on Biometrics (ICB)* (pp. 1-8). IEEE.

Hinton, G.E. and Sejnowski, T.J., 1986. Learning and relearning in Boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition*, *1*(282-317), p.2.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Horng, W.B., Lee, C.P. and Chen, C.W., 2001. Classification of age groups based on facial features. *淡江理工學刊*, *4*(3), pp.183-192.

Huerta, I., Fernández, C., Segura, C., Hernando, J. and Prati, A., 2015. A deep analysis on age estimation. *Pattern Recognition Letters*, *68*, pp.239-249.

Hu, Z., Wen, Y., Wang, J., Wang, M., Hong, R. and Yan, S., 2016. Facial age estimation with age difference. *IEEE Transactions on Image Processing*, *26*(7), pp.3087-3097.

Qawaqneh, Z., Mallouh, A.A. and Barkana, B.D., 2017. Deep convolutional neural network for age estimation based on VGG-face model. *arXiv preprint arXiv:1709.01664*.

Qiu, J., 2016. Convolutional neural network based age estimation from facial image and depth prediction from single image.

Kamra, A. and Gupta, S., 2016. Automatic Age Estimation Through Human Face: A Survey. In *International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences (EEECOS)*.

Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J. and Keutzer, K., 2017. Medical image computing and computer-assisted Intervention–MICCAI. *ICLR*, *8150*, pp.1-13.

Lanitis, A., 2010. Facial age estimation. *Scholarpedia*, *5*(1), p.9701.

LeCun, Y., Bengio, Y. and Hinton, G., Deep learning.” nature 521.7553 (2015): 436. *DOI: https://doi. org/10.1038/nature14539 https://doi. org/10.1038/nature14539*.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), pp.2278-2324.

Levi, G. and Hassner, T., 2015. Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 34-42).

Li, W., Lu, J., Feng, J., Xu, C., Zhou, J. and Tian, Q., 2019. Bridgenet: A continuity-aware probabilistic network for age estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1145-1154).

Liao, H., Yan, Y., Dai, W. and Fan, P., 2018. Age estimation of face images based on CNN and divide-and-rule strategy. *Mathematical Problems in Engineering*, *2018*.

Lin, C.T., Li, D.L., Lai, J.H., Han, M.F. and Chang, J.Y., 2012. Automatic age estimation system for face images. *International Journal of Advanced Robotic Systems*, *9*(5), p.216.

Liu, T.J., Liu, K.H., Liu, H.H. and Pei, S.C., 2016, September. Age estimation via fusion of multiple binary age grouping systems. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 609-613). IEEE.

Liu, H., Lu, J., Feng, J. and Zhou, J., 2017. Ordinal deep learning for facial age estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, *29*(2), pp.486-501.

Liu, H., Lu, J., Feng, J. and Zhou, J., 2017. Group-aware deep feature learning for facial age estimation. *Pattern Recognition*, *66*, pp.82-94.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. and Alsaadi, F.E., 2017. A survey of deep neural network architectures and their applications. *Neurocomputing*, *234*, pp.11-26.

Liu, X., Li, S., Kan, M., Zhang, J., Wu, S., Liu, W., Han, H., Shan, S. and Chen, X., 2015. Agenet: Deeply learned regressor and classifier for robust apparent age estimation. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 16-24).

Liu, X., Zou, Y., Kuang, H. and Ma, X., 2020. Face Image Age Estimation Based on Data Augmentation and Lightweight Convolutional Neural Network. *Symmetry*, *12*(1), p.146.

Minear, M. and Park, D.C., 2004. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), pp.630-633.

Miyamoto, N., Jinnouchi, Y., Nagata, N. and Inokuchi, S., 2005, September. Subjective age estimation system using facial images. In *International Conference on Entertainment Computing* (pp. 223-229). Springer, Berlin, Heidelberg.

Mohammed, S.M.O. and Viriri, S., 2017, November. Gender identification from facial images: Survey of the state-of-the-art. In *2017 Sudan Conference on Computer Science and Information Technology (SCCSIT)* (pp. 1-5). IEEE.

Nam, S.H., Kim, Y.H., Truong, N.Q., Choi, J. and Park, K.R., 2020. Age Estimation by Super-Resolution Reconstruction Based on Adversarial Networks. *IEEE Access, 8,* pp.17103-17120.

Narvekar, S. and Laxminarayana, J.A., 2020. Approaches to Gender Classification: A Survey.

Ng, C.C., Cheng, Y.T., Hsu, G.S. and Yap, M.H., 2017, May. Multi-layer age regression for face age estimation. In *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)* (pp. 294-297). IEEE.

Ng, C.C., Yap, M.H., Costen, N. and Li, B., 2015, October. Will wrinkle estimate the face age?. In *2015 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 2418-2423). IEEE.

Ng, C.C., Yap, M.H., Cheng, Y.T. and Hsu, G.S., 2018. Hybrid ageing patterns for face age estimation. *Image and Vision Computing*, *69*, pp.92-102.

Ni, B., Song, Z. and Yan, S., 2009, October. Web image mining towards universal age estimator. In *Proceedings of the 17th ACM international conference on Multimedia* (pp. 85-94).

Osman, O.F. and Yap, M.H., 2018. Computational intelligence in automatic face age estimation: A survey. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *3*(3), pp.271-285.

Olshausen, B.A. and Field, D.J., 1997. Sparse coding with an overcomplete basis set: A strategy employed by V1?. *Vision research*, *37*(23), pp.3311-3325.

Olsson, C., Age estimation From Facial Images.

Oquab, M., Bottou, L., Laptev, I. and Sivic, J., 2015. Is object localization for free?-weakly-supervised learning with convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 685-694).

Phillips, P.J., Der, S.Z., Rauss, P.J. and Der, O.Z., 1996. *FERET (face recognition technology) recognition algorithm development and test results*. Adelphi, MD: Army Research Laboratory.

Raina, R., Battle, A., Lee, H., Packer, B. and Ng, A.Y., 2007, June. Self-taught learning: transfer learning from unlabeled data. In *Proceedings of the 24th international conference on Machine learning* (pp. 759-766).

Reade, S. and Viriri, S., 2015, July. Hybrid Age Estimation Using Facial Images. In *International Conference Image Analysis and Recognition* (pp. 239-246). Springer, Cham.

Ricanek, K. and Tesafaye, T., 2006, April. Morph: A longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)* (pp. 341-345). IEEE.

Rothe, R., Timofte, R. and Van Gool, L., 2015. Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 10-15).

Rothe, R., Timofte, R. and Van Gool, L., 2018. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, *126*(2-4), pp.144-157.

Niu, Z., Zhou, M., Wang, L., Gao, X., 2016, Ordinal regression with multiple output cnn for age estimation, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4920–4928.

Sarita, J. and Patil, A.J., 2016. Human Age Prediction from Facial Images. *International Journal of Innovative Research in Computer and Communication Engineering*, *4*(7).

Shahriari, B., Swersky, K., Wang, Z., Adams, R.P. and De Freitas, N., 2015. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, *104*(1), pp.148-175.

Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Snoek, J., Larochelle, H. and Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, *25*, pp.2951-2959.

Srinivas, N., Krause, A., Kakade, S.M. and Seeger, M., 2009. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*.

Sun, Y., Chen, Y., Wang, X. and Tang, X., 2014. Deep learning face representation by joint identification-verification. *Advances in neural information processing systems*, *27*, pp.1988-1996.

Sun, Y., Wang, X. and Tang, X., 2013. Deep convolutional network cascade for facial point detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3476-3483).

Sun, Y., Wang, X. and Tang, X., 2014. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1891-1898).

Suo, J., Min, F., Zhu, S., Shan, S. and Chen, X., 2007, June. A multi-resolution dynamic model for face aging simulation. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C. and Liu, C., 2018, October. A survey on deep transfer learning. In *International conference on artificial neural networks* (pp. 270-279). Springer, Cham.

Tao, Y., 2014. Automated estimation of human age, gender and expression. *EE368 Digital Image Processing Final Project Report*, p.4.

Tin, H.H.K., 2011. Gender and Age Estimation Based on Facial Images. *International Journal: ACTA TECJNICA NAPOCENSIS Electronics and Telecommunications*, *52*(3).

Tin, H.H.K., 2012. Subjective age prediction of face images using PCA. *International Journal of Information and Electronics Engineering*, *2*(3), p.296.

Trivedi, G. and Pise, N.N., Gender Classification and Age Estimation using Neural Networks: A Survey. *International Journal of Computer Applications*, *975*, p.8887.

Ueki, K., Hayashida, T. and Kobayashi, T., 2006, April. Subspace-based age-group classification using facial images under various lighting conditions. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)* (pp. 6-pp). IEEE.

Wan, J., Wang, D., Hoi, S.C.H., Wu, P., Zhu, J., Zhang, Y. and Li, J., 2014, November. Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 157-166).

Wan, L., Zeiler, M., Zhang, S., Le Cun, Y. and Fergus, R., 2013, February. Regularization of neural networks using dropconnect. In *International conference on machine learning (pp. 1058-1066).*

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T. and Gong, Y., 2010, June. Locality-constrained linear coding for image classification. In *2010 IEEE computer society conference on computer vision and pattern recognition* (pp. 3360-3367). IEEE.

Wang, X., Guo, R. and Kambhamettu, C., 2015, January. Deeply-learned feature for age estimation. In *2015 IEEE Winter Conference on Applications of Computer Vision* (pp. 534-541). IEEE.

Weiss, K., Khoshgoftaar, T.M. and Wang, D., 2016. A survey of transfer learning. *Journal of Big data*, *3*(1), p.9.

Yang, J., Yu, K., Gong, Y. and Huang, T., 2009, June. Linear spatial pyramid matching using sparse coding for image classification. In *2009 IEEE Conference on computer vision and pattern recognition* (pp. 1794-1801). IEEE.

Ylioinas, J., Hadid, A., Hong, X. and Pietikäinen, M., 2013, September. Age estimation using local binary pattern kernel density estimate. In *International Conference on Image Analysis and Processing* (pp. 141-150). Springer, Berlin, Heidelberg.

Yu, K., Zhang, T. and Gong, Y., 2009. Nonlinear learning using local coordinate coding. *Advances in neural information processing systems*, *22*, pp.2223-2231.

Zeiler, M.D., 2013. *Hierarchical convolutional deep learning in computer vision* (Doctoral dissertation, New York University).

Zhang, K., Gao, C., Guo, L., Sun, M., Yuan, X., Han, T.X., Zhao, Z. and Li, B., 2017. Age group and gender estimation in the wild with deep RoR architecture. *IEEE Access*, *5*, pp.22492-22503.

Zhang, Z., Luo, P., Loy, C.C. and Tang, X., 2014, September. Facial landmark detection by deep multi-task learning. In *European conference on computer vision* (pp. 94-108). Springer, Cham.