

BAŞKENT UNIVERSITY
INSTITUTE OF SCIENCE AND ENGINEERING

**GENDER RECOGNITION AND AGE ESTIMATION BASED
ON HUMAN GAIT**

MURAT BERKSAN

MASTER OF SCIENCE THESIS

2019

**GENDER RECOGNITION AND AGE ESTIMATION BASED
ON HUMAN GAIT**

YÜRÜYÜŞ BİÇİMİNDEN CİNSİYET VE YAŞ TESPİTİ

MURAT BERKSAN

Thesis Submitted
in Partial Fulfillment of the Requirements
For the Degree of Master of Science
in Department of Computer Engineering
at Başkent University

2019

This thesis, titled: "GENDER RECOGNITION AND AGE ESTIMATION BASED ON HUMAN GAIT", has been approved in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN COMPUTER ENGINEERING**, by our jury on 12/09/2019.

Chairman : Assoc. Prof. Dr. Uğur Murat LELOĞLU

Member (Supervisor) : Asst. Prof. Dr. Emre SÜMER

Member : Asst. Prof. Dr. Selda GÜNEY

APPROVAL

..../09/2019

Prof. Dr. Faruk ELALDI
Fen Bilimleri Enstitüsü Müdürü



**BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ YÜKSEK LİSANS
TEZ ÇALIŞMASI ORJİNALLİK RAPORU**

Tarih: ../09/2019

Öğrencinin Adı, Soyadı : MURAT BERKSAN

Öğrencinin Numarası : 21620204

Anabilim Dalı : BİLGİSAYAR MÜHENDİSLİĞİ

Programı : BİLGİSAYAR MÜHENDİSLİĞİ TEZLİ YÜKSEK LİSANS

Danışmanın Adı, Soyadı: DR. ÖĞR. ÜYESİ EMRE SÜMER

Tez Başlığı : Yürüyüş Biçiminden Cinsiyet ve Yaş Tespiti

Yukarıda başlığı belirtilen Yüksek Lisans/Doktora tez çalışmamın; Giriş, Ana Bölümler ve Sonuç Bölümünden oluşan, toplam 70 sayfalık kısmına ilişkin, 11/08/2019 tarihinde tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı %8'dir.

Uygulanan filtrelemeler:

1. Kaynakça hariç
2. Alıntılar hariç
3. Beş (5) kelimedenden daha az örtüşme içeren metin kısımları hariç

“Başkent Üniversitesi Enstitüleri Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Usul ve Esaslarını” inceledim ve bu uygulama esaslarında belirtilen azami benzerlik oranlarına tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Öğrenci İmzası:

Onay

../ 09/ 2019

Dr. Öğr. Üyesi Emre Sümer

ACKNOWLEDGMENT

I would like to thank my master thesis advisor, Asst. Prof. Dr. Emre Sümer, for his advices and support. His valuable guidance helped me during the whole period of my research.

I would also like to thank my family for their moral support throughout writing this thesis and my studies. This accomplishment would not have been possible without them.

ABSTRACT

GENDER RECOGNITION AND AGE ESTIMATION BASED ON HUMAN GAIT

Murat BERKSAN

Baskent University Institute of Science and Engineering

Computer Engineering Department

In this study, the feasibility of Convolutional Neural Networks (CNN) for gait based gender recognition and age estimation problems were investigated. For this purpose, different networks were evaluated and a basis was selected. Further adjustments were made on the basis network by experimenting on architectural options and hyperparameters. Two distinct yet similar architectures were proposed for each problem. The experiments were conducted by using gait silhouette average which is a feature descriptor as input. The overall accuracy was computed to be 97.45% using the proposed CNN architecture for gender recognition and 5.74 years mean absolute error for age estimation. Using CNN with gait silhouette average as an input is an understudied subject in the literature for these problem domains. While there is one study that uses this approach for gait based gender recognition, there are no studies evaluating CNN for gait based age estimation. The results show successful performance comparable to existing studies. Besides, the experimental results provide insight on how network structure and hyperparameters affect performance. Considering this, obtained outcome allows to gain insight about the problem domain of using gait feature descriptor for gender recognition and age estimation, and provides guidance about deciding on a CNN network in these problem domains.

KEYWORDS: Gender Recognition, Age Estimation, Convolutional Neural Networks, Gait, Gait Silhouette

Advisor: Asst. Prof. Dr. Emre Sümer, Baskent University, Computer Engineering Department.

ÖZ

YÜRÜYÜŞ BİÇİMİNDEN CİNSİYET VE YAŞ TESPİTİ

Murat BERKSAN

Başkent Üniversitesi Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Bu çalışmada, Evrişimsel Sinir Ağları'nın (ESA) yürüyüş biçimi tabanlı cinsiyet ve yaş tespiti alanlarında uygulanabilirliği incelenmiştir. Bu amaçla farklı ağlar değerlendirilmiş olup, kaynak bir ağ seçilmiştir. Baz alınan bu ağ üzerinde farklı mimari seçenekler ve üst değişkenler ile ilgili deneyler yapılarak değişiklikler yapılmıştır. Her bir problem için benzer yapıları ancak farklı iki mimari önerilmiştir. Deneyler yürüyüş biçimi silueti ortalamasını girdi olarak alıp gerçekleştirilmiştir. Cinsiyet tespitinde sonuç olarak %97.45 doğruluk elde edilmiş olup, yaş tespitinde 5.74 yıl ortalama mutlak hata sonucu alınmıştır. Bir özellik tanımlayıcı olan yürüyüş biçimi silueti ortalamasını girdi olarak ESA kullanımı literatürde az çalışılmış bir konudur. Bu yaklaşımı yürüyüş biçimi tabanlı cinsiyet tespiti alanında kullanan tek bir çalışma bulunmakla birlikte, yürüyüş biçimi tabanlı yaş tespiti problemini çözen bir yaklaşıma literatürde rastlanmamıştır. Sonuçlar literatürde var olan çalışmalarla karşılaştırıldığında önerilen mimarilerin başarılı bir performans sergilediği görülmektedir. Ayrıca, deneyler sırasında alınan sonuçlar mimari yapı ile üst parametrelerin performansı nasıl etkilediğine dair anlayış sağlamaktadır. Bütün bunlar göz önünde bulundurulduğunda, alınan sonuçlar yürüyüş özellik tanımlayıcısının bu problem alanlarında kullanımı hakkında anlayış sağlamakla birlikte, ESA'nın bu problem alanlarında kullanımı için yol göstermektedir.

ANAHTAR SÖZCÜKLER: Cinsiyet tespiti, Yaş tespiti, Evrişimsel Sinir Ağları, Yürüyüş biçimi, Yürüyüş Biçimi Silueti

Danışman: Dr. Öğr. Üyesi Emre SÜMER, Başkent Üniversitesi, Bilgisayar Mühendisliği Bölümü.

TABLE OF CONTENTS

ABSTRACT	i
ÖZ.....	ii
LIST OF FIGURES	iv
LIST OF TABLES	vi
LIST OF ABBREVIATIONS	vii
1 INTRODUCTION	1
1.1 Problem Statement and Motivation.....	1
1.2 Gait Energy Image	3
1.3 Dataset.....	5
1.4 Convolutional Neural Networks	8
1.5 Contribution of the Thesis	16
1.6 Outline.....	17
2 LITERATURE REVIEW.....	18
2.1 Gait Based Gender Recognition.....	18
2.2 Gait Based Age Group Classification and Estimation.....	26
3 GAIT BASED GENDER RECOGNITION.....	30
3.1 Architecture Overview	30
3.2 Network Structure.....	31
3.3 Hyperparameter Optimization	36
4 GAIT BASED AGE ESTIMATION	42
4.1 Architecture Overview	42
4.2 Network Structure.....	43
4.3 Hyperparameter Optimization	48
5 RESULTS AND DISCUSSION.....	54
5.1 Gender Recognition Results	54
5.2 Age Estimation Results	55
5.3 Discussion.....	57
6 CONCLUSION.....	60

LIST OF FIGURES

	<u>Page</u>
Figure 1.1. Example gait silhouette [61]	3
Figure 1.2. Gait Cycle [14].....	4
Figure 1.3. The process of obtaining GEI. Rightmost frame shows GEI [61].....	5
Figure 1.4. Gender distribution of the OULP-Age dataset	6
Figure 1.5. Age distribution of the OULP-Age Dataset	7
Figure 1.6. Example of GEI in OULP-Age dataset.....	7
Figure 1.7. Logistic regression structure.....	11
Figure 1.8. NN structure	11
Figure 3.1. The CNN Architecture for gender recognition	30
Figure 3.2. Performance comparison of different architectures	31
Figure 3.3. Comparison of modified GEINet with and without normalization	32
Figure 3.4. Performance comparison of modified GEINet with different FC layer combinations	33
Figure 3.5. Performance comparison of modified GEINet with different FC layer unit numbers.....	34
Figure 3.6. Comparison of modified GEINet with and without dropout	35
Figure 3.7. Comparison of modified GEINet with different pooling	36
Figure 3.8. Performance comparison of optimization algorithms.....	37
Figure 3.9. Performance comparison of different learning rates	37
Figure 3.10. Performance comparison of stride on convolutional layer	38
Figure 3.11. Performance comparison of kernel size.....	39
Figure 3.12. Performance comparison of kernel number.....	39
Figure 3.13. Performance comparison of network with different batch size values	40
Figure 3.14. Performance of the network with L2 regularization.....	40
Figure 3.15. Performance of the network with L1 regularization.....	41
Figure 4.1. Overall architecture of gait based age estimation.....	42
Figure 4.2. Performance comparison of different architectures	43
Figure 4.3. Performance comparison of network with normalization and no normalization	44
Figure 4.4. Performance comparison of network with different FC layer options .	46
Figure 4.5. Performance comparison of network with dropout and no dropout....	47

Figure 4.6. Performance comparison of network with different pooling options ...	48
Figure 4.7. Performance comparison of network with different optimization algorithms.....	49
Figure 4.8. Performance comparison of network with different learning rate values	49
Figure 4.9. Performance comparison of network with stride values in convolutional layers.....	50
Figure 4.10. Performance comparison of network with kernel size in convolutional layers.....	51
Figure 4.11. Performance comparison of network with kernel number in convolutional layers.....	51
Figure 4.12. Performance comparison of network with different batch sizes	52
Figure 4.13. Performance of network with L2 regularization.....	53
Figure 4.14. Performance of network with L1 regularization.....	53
Figure 4.15. Performance of network with L1 regularization and more epochs ...	53
Figure 5.1. Performance of proposed CNN architecture for gender recognition ..	54
Figure 5.2. Performance of proposed CNN architecture for age estimation	55
Figure 5.3. Actual age against predicted age.....	57

LIST OF TABLES

	<u>Page</u>
Table 1.1. AlexNet Architecture	14
Table 1.2. GEINet Architecture.....	15
Table 3.1. Modified GEINet architecture	32
Table 3.2. Architecture after removal of LRN	33
Table 3.3. Architecture after FC layer evaluation	34
Table 3.4. Network structure after dropout experiment	35
Table 3.5. Final network structure of gait based gender recognition	39
Table 4.1. Structure of network without normalization	44
Table 4.2. Network structure after FC layer evaluation	46
Table 4.3. Network structure without dropout.....	47
Table 4.4. Final network structure	51
Table 5.1. Accuracy of different gait based gender recognition methods	55
Table 5.2. MAE of different gait based age estimation methods	56

LIST OF ABBREVIATIONS

CCA	Canonical Correlation Analysis
CNN	Convolutional Neural Network
CSLPP	Cost Sensitive Locality Preserving Projections
CSPCA	Cost Sensitive Principal Component Analysis
FC	Fully connected
FED	Frame to Exemplar
Gait Energy Image	GEI
HMM	Hidden Markov Model
KNN	K-nearest Neighbor
LDA	Linear Discriminant Analysis
LRN	Local Response Normalization
MMI	Maximization of Mutual Information
OPLDA	Ordinary Preserving Linear Discriminant Analysis
OPMFA	Ordinary Preserving Margin Fisher Analysis
OULP-Age	OU-ISIR Gait Database, Large Population Dataset with Age
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
SDG	Stochastic Gradient Descent
SVM	Support Vector Machine

1 INTRODUCTION

1.1 Problem Statement and Motivation

Biometrics is the study of human identification based on human characteristics. Biometric identifiers are categorized as physiological and behavioral characteristics [1]. Human face and fingerprint can be shown as physiological identifiers. Human gait is a behavioral identifier and it has received attention in the computer vision community. While it is not possible to identify a person who is not willing to provide their physiological identifier, capturing human gait does not require cooperation. For example, it is not possible to perform facial recognition on an individual who cover their face, who does not face the capturing device or who is far away from the capturing device so that details of their face cannot be seen. However, it is possible to capture gait from distance because the information needed can be obtained from movement of body and body parts which are more apparent than face in distance [2]. In addition to this, it is possible to make deduction from human gait that is captured from different angles therefore the subject does not need to face the capturing device and a system that uses gait information does not require the cooperation of the subject. These benefits of using human gait as identifier provides opportunities to overcome shortcomings of physiological identifier based systems.

It is possible to infer various human traits by analyzing human gait. Gender and age are examples of these. Studies show that different genders and ages have different walking patterns. In their experiment, Kozlowski et al. [3] attached point-light displays on subjects and a group of observers were asked to determine the gender of subjects. Results show that human observers were able to recognize gender with 63% accuracy without any cue on clothing or hair. This evidence suggests that it is possible to derive gender based on gait. Gait differences among different age groups were confirmed by several studies. Grabiner et al [4] showed that stride width is effected by age. In a similar way, Elbe et al. [5] proved that elderly people has decreased walking speed and stride width which causes different arm swing, rotations of joints and the maximum amount of foot height in gait. These show that people with different ages has different gait patterns which

makes it possible to infer age based on gait.

Gait based gender recognition and age estimation may be leveraged by different kinds of systems. Today CCTVs are widely available and used in criminal investigations. Although there may be existing systems that assist on these investigations by using face as a biometric identifier, these systems suffer from low resolution or the subject's face may not be apparent in the footage. In such cases, gait based gender recognition may provide the gender of the subject which may be helpful for investigations of law enforcement. In addition to this, ability to infer age from distance may contribute to identification of suspect which would speed up the process. Another possible use case can be shown in marketing. A gait based gender recognition system can be setup on a street which can infer gender of pedestrians without getting close-up image of people. This information can be then used to display advertisement on electronic billboards based on pedestrians' gender. Furthermore, estimation of subject's age would provide important material for such a system as different age groups may have different interests. Such cases show that gait based gender recognition and age estimation may provide useful information.

A gait recognition framework generally consists of detection of the subject, preprocessing of subject's footage in order to extract feature, feature extraction, feature selection and classification. There are two approaches of preprocessing: model-based and model-free [6]. While model-based approaches focus on creating models by extracting body parts and obtaining information from locomotion of those models, model-free approaches focus on spatiotemporal pattern of human in motion by forming a representation from silhouette of subjects. Both approaches have their advantages and disadvantages. Model-based approaches are view and scale independent due to ability of deduction from human joints and connected body parts as a whole [7]. However, this also introduces a complexity problem. In order to be able to model human gait in a way that would sufficiently classify the subject, several parameters must be introduced and this adds up to computational cost [8]. On the contrary, model-free approaches are computationally efficient due to not doing any body part detection and using the footage as whole [9].



Figure 1.1. Example gait silhouette [61]

However, quality of gait silhouette plays an important role in gait recognition. The quality of gait silhouette and features extracted from it are affected by illumination, clothing of the subject and carrying condition due to background subtraction scheme that is used [10]. Despite the drawbacks of this approach, it is possible to apply it in real-time environments. Thus, the recent trend in gait based recognition is utilizing a model-free approach and this study adopts it.

1.2 Gait Energy Image

Model-free approaches utilize gait silhouettes and aim to obtain a spatiotemporal gait representation. A gait silhouette is obtained by subtracting background from the footage and binarizing the image. An example of gait silhouette is shown in Figure 1.1.

There are multiple methods to obtain spatiotemporal gait representation in the literature. Liu et al. [11] constructed frieze pattern from series of gait silhouettes. Mathematically a frieze pattern means a motif that repeats indefinitely over one dimension which is suitable for the nature of human gait. By constructing frieze pattern for both column and width, it is possible to obtain a gait representation. Another method is Gait Curve [12] which defines gait representation as difference in silhouette contour over the frames in the footage. By finding the change of silhouette boundary over time it is possible to obtain both spatial and temporal change that occurs during walking which can be used for pattern recognition. Finally, Gait Energy Image (GEI) [13] is a spatiotemporal gait representation that is able to represent the subject's both movements and posture. This ability makes GEI an important gait feature descriptor.

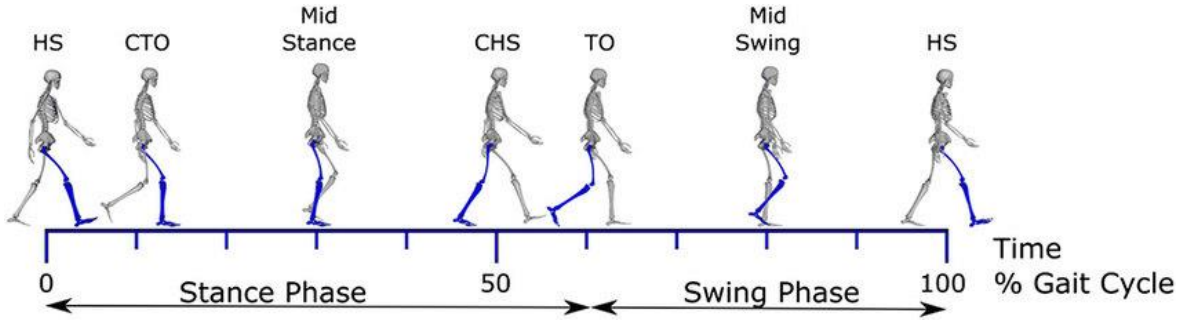


Figure 1.2. Gait Cycle [14]

The process to obtain GEI from footage of a walking subject starts by converting the footage into binary gait silhouettes which is performed by subtracting the background and binarization of the image. The background subtraction and binarization method follows the procedure described in [15] by defining a bounding box on the subject manually, calculating mean and covariance of pixel out of the box and finally computing Mahalanobis distance of pixel values within the box to the calculated out of the box statistics. Thus each pixel in the image is classified as foreground and background. After this, frames that include a complete gait cycle which is the moment passed from subject's foot contacts the ground until the same foot contacts the ground again is chosen. Figure 1.2 depicts a gait cycle where HS stands for heel strike; CTO stands for contralateral toe off the ground; CHS stands for contralateral heel strike; TO stands for toe off.

After the extraction of gait cycle frames, the frames are averaged by summing up each frame and dividing by the number of frames therefore obtaining GEI. The equation of the averaging procedure is shown in Equation 1.1 where $B_t(x,y)$ is binarized silhouette image and N is the number of frames for a gait cycle.

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N B_t(x, y) \quad (1.1)$$

The visual process of obtaining GEI can be seen in Figure 1.3 where each frame until the rightmost one shows binarized gait silhouette and the rightmost frame shows the GEI obtained by averaging the previous frames.



Figure 1.3. The process of obtaining GEI. Rightmost frame shows GEI [61]

While human posture is preserved in this process, pixel intensity depicts the frequency of movement. Posture and shape of body represents the spatial aspect and pixel intensity represents temporal aspect of gait. This means that high pixel value illustrates less spatial change over time and low pixel value shows higher one. Thus attributes such as stride frequency and stride length of gait are encoded in a single image.

1.3 Dataset

There are several available datasets for different purposes in gait based recognition. CASIA Gait Database: Dataset B [16] is one of these datasets that can be used for gender recognition. The dataset consists of gait footage of 124 subjects taken from multiple views. Although the data is labeled with gender, number of males is 93 and number of females is 31. The dataset was constructed mainly for multiple view gait recognition under normal walking conditions along different clothing and carrying condition. The gender ratio in this dataset suggests that it is biased in terms of gender distribution. Studies that use this dataset generally randomly select 31 males in order to overcome the bias. However, the size of the dataset is smaller compared to existing ones. Likewise, SOTON Large Dataset [17] consists of 115 subjects with number of male being 91 and female being 24 which shows that the dataset is not balanced in terms of gender. This dataset includes footage of subjects captured from side and oblique view, and in different environments.

The TUM Gait Audio, Image and Depth Database [18] is the only multimodal database available in the literature. The subjects were recorded using depth cameras and their audio was recorded during walking. The database has both gender and age labels. Gait footage of 186 males and 119 females were recorded and their age range from 18 to 55 years with the average age being 24.8 years.

Although gender ratio is more balanced than previously mentioned datasets, this dataset is not suitable for age group classification because of missing elderly and children data. In addition, capturing took place in different seasons which led to have clothing covariate and this affects the quality of obtained silhouette.

Another frequently used dataset USF Human ID Gait Challenge Dataset [15] consists of gait footage of 122 subjects captured under different conditions such as different views, carrying conditions and different environments. The data is labeled with both gender and age but the 75% of subjects are male and age ranges between 19 and 59 years with age of subjects predominantly being between 21 and 33. These statistics about the dataset indicate an imbalance in both gender and age.

In this work, the OU-ISIR Gait Database, Large Population Dataset with Age (OULP-Age) [19] was used for the experiments. OULP-Age is currently the largest gait data set that is published and comprises gait of 63,846 subjects. 31,093 of the subjects are male and the remaining 32,753 are female. This indicates that the data set is balanced in terms of gender and is suitable for gender recognition. The age range in the dataset is between 2 and 90 years old. This means that the dataset comprises of entire generation which makes it suitable for age group classification.

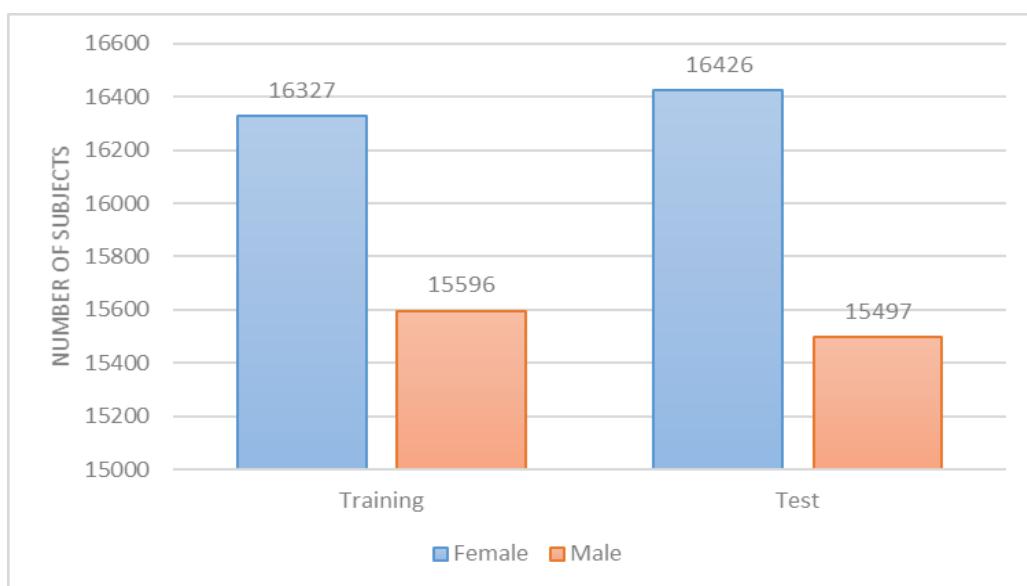


Figure 1.4. Gender distribution of the OULP-Age dataset

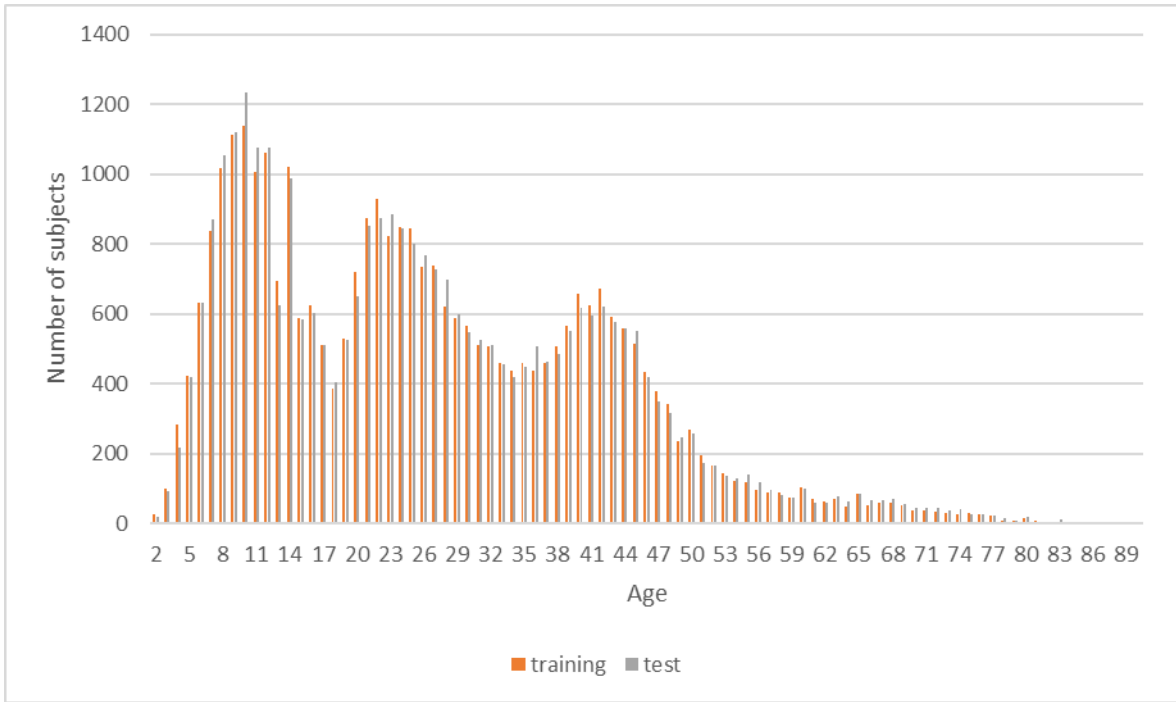


Figure 1.5. Age distribution of the OULP-Age Dataset

Furthermore, although the number of subjects is predominantly between ages 6 and 50, it is able to sufficiently depict change from childhood to elder age which is something that no existing dataset encompasses. The data set was split into training and test sets considering gender and age balance. Both training and test sets consist of 31,923 subjects therefore they are equal sizes and the gender and age distribution in these sets can be seen in Figure 1.4. and 1.5.

The data set was constructed by capturing gait of subjects in a controlled environment from side view from approximately 4 meters away.



Figure 1.6. Example of GEI in OULP-Age dataset

Footage used to construct OULP-Age were taken in front of a green screen therefore gait silhouette construction was automatized and thus the errors that could occur at this stage were minimized. Background subtraction used to obtain gait silhouettes follows the procedure defined in [20] and applies graph-cut segmentation.

After obtaining gait silhouettes, GEI of each subject was constructed and thus dataset was created. The quality of the dataset was checked manually to ensure that image quality is sufficient and there is no covariance that would affect the walking of the subject. Furthermore, gender and age information provided by participants were checked by asking participants validity of this information in order to ensure correct information was given. An example of constructed GEI in the dataset can be seen in Figure 1.6.

OULP-Age dataset is suitable for the purposes of this work because a) the dataset is balanced which is an important aspect to have in machine learning approaches, b) the dataset is large compared to the existing datasets and the size of the dataset is one of the keys to success in deep learning approaches, c) the dataset consists of GEI which is constructed in controlled environment.

1.4 Convolutional Neural Networks

The next step in gait based recognition is to do feature selection and use an appropriate method for classification and estimation. For this purpose, Convolutional Neural Network (CNN) [21] was employed in this study. A CNN is a hierarchical multilayered neural network which can learn visual patterns directly from the image. CNN has recently received great deal of attention due to increased amounts of data which is a key for success in deep learning and increased computational power.

Understanding how Neural Networks (NN) and CNNs work can be achieved by understanding how logistic regression works. Logistic regression is a statistical model that is used to predict the probability of a given binary variable which is depicted in Equation 1.2 where x is the input and \hat{y} is the output.

$$\hat{y} = P(y = 1 | x) \quad (1.2)$$

Given input vector $x \in \mathfrak{R}^{n_x}$, weight vector $w \in \mathfrak{R}^{n_x}$ and bias $b \in \mathfrak{R}$ where n_x is the size of the vector, we want to find output \hat{y} . The first step to achieve this is to form a linear function of the input x which can be seen in Equation 1.3.

$$\hat{y} = w^T x + b \quad (1.3)$$

However, this function can output a value greater than one which is something that is not desired therefore sigmoid function is applied to restrict the output to be between 0 and 1. In NNs, different activation functions can be applied for different problems. The final state of the equation can be seen in Equation 1.4

$$\hat{y} = \sigma(w^T x + b) \quad (1.4)$$

The aim is to learn parameters w and b in order to obtain prediction \hat{y} that is closest to y which can be described as the label in the training set. In order to do this, the relation between \hat{y} and y needs to be defined in a way that measures how close they are. This is done by defining the loss function that can be seen in Equation 1.5 which represents how good the prediction \hat{y} is given the label y . This loss function is used for binary classification and different loss function needs to be defined for different problems.

$$\ell(\hat{y}, y) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y})) \quad (1.5)$$

The desired outcome is to have lowest possible loss value which describes the closest prediction value against the label. The loss function is applied for a single training data. In order to find the performance in the whole dataset, the cost function needs to be defined which can be seen in Equation 1.6 where m is the number of examples in the set.

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m \ell(\hat{y}^{(i)}, y^{(i)}) \quad (1.6)$$

The objective is to find such values for parameters w and b so that the cost minimized. The problem becomes an optimization problem which has the aim of finding the minimum cost. There are several optimization algorithms defined in the literature to tackle this problem and gradient descent is one of these. Basically, gradient descent takes steps on steepest downhill direction on the surface of the cost function and finds the minimum of the function. This is achieved through repeatedly updating parameters w and b by taking derivative of cost function with respect to each parameter and subtracting from them. The formula for this process can be seen in Equation 1.7 and the process continues until it converges.

An important part of this formula to point out is the learning rate α which basically defines how small or big steps can the algorithm takes and effectively change the learning capacity. In NN context, calculating cost function is called forward propagation and calculating gradient of cost function is called backward propagation.

Different optimization algorithms can be used update parameters. Combination of one forward and backward propagation is called an epoch. The learning takes place by applying this procedure for multiple epochs and therefore updating parameters to have such values that the predicted value is closest to the label.

$$w = w - \frac{\partial J(w, b)}{\partial w}$$

$$b = b - \frac{\partial J(w, b)}{\partial b} \tag{1.7}$$

The learning process of NNs is nearly indistinctive from logistic regression as logistic regression can be arguably called a very small NN. Figure 1.7 illustrates the logistic regression and Figure 1.8 shows a NN. As can be seen from these figures, NNs have multiple computational units called neurons whereas logistic regression has one. It can also be seen that NNs have neurons that stacked horizontally which are called layers.

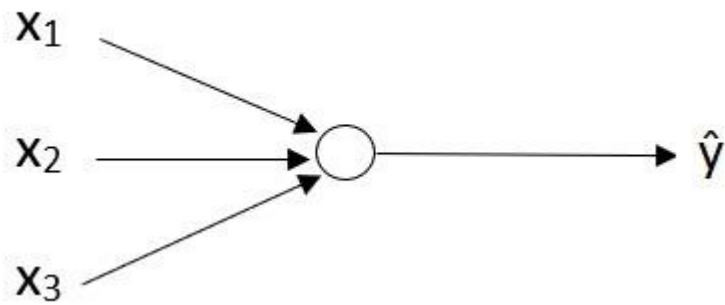


Figure 1.7. Logistic regression structure

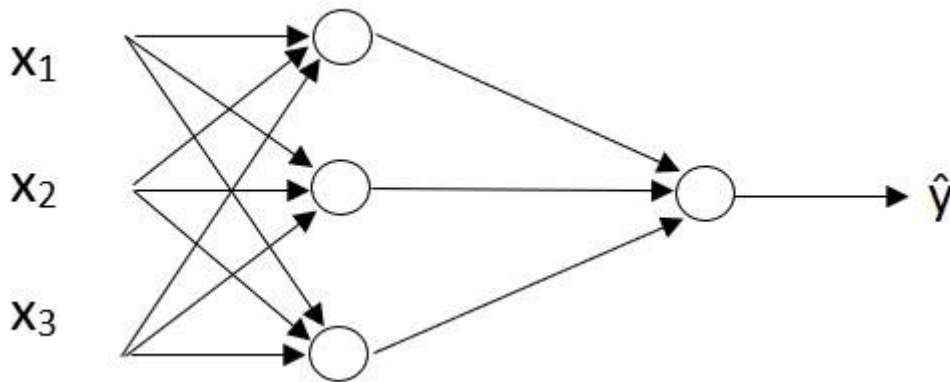


Figure 1.8. NN structure

These neurons can feed following neurons with their output and therefore learn more complex features. In principle, NNs work in similar manner to logistic regression but their ability to stack neurons and work on more parameters enables them to tackle harder problems.

CNN can be considered a subset of NN. By design, it is used to solve problems in visual domain and therefore CNNs have special layers that are different than NN layers. CNNs typically consist of multiple layers which are convolutional layer, pooling layer and fully connected (FC) layer, usually in that order.

Although it has special layers, the basic principles are the same as NN. One of the reason it has these special layers is that images contain too much information and therefore computational cost increases. These special layers extract the important

information which are called features and do learning based on them. This is done in convolutional layers by convolving image inputs with a set of filters and outputting images which are smaller but contain vital information. This operation not only helps to reduce the dimensions but also helps to preserve localization. If we were to use NN for learning, we would need to flatten image pixels into single vector and do the learning which would eliminate information that would be obtained from relations of neighbor pixels and reduce the learning capacity. The size, stride and padding of the filters can be adjusted to affect localization of the network.

The feature extraction process in convolutional layers are automatized and do not require filters to be defined. In fact, each value in filters is a weight that is learned in this process just as the weights in NN are learned. This allows both automatized process to obtain more relevant information and to stack convolutional layers in order to extract more complex features.

Another special layer is called pooling layer and no learning is done in this layer. The purpose of this layer is to reduce the dimensionality further and to pass the important information to the next layer. There are multiple choices of pooling layer which are maximum, minimum and average pooling. These resemble the filters in convolutional layers and have adjustable size, stride and padding. However, these are not convolved with the image in the mathematical sense, but are just slid over the input to find maximum, minimum or average value in the input.

Finally, FC layers are supplied to the network in order to learn from the extracted features and do classification or estimation. Before these layers, extracted feature images are flattened and fed to first FC layer. These layers are exactly the same as NN and all the learning procedure is identical.

Different combinations of layers can be introduced to a network for different problems. A deep network is generally capable of capturing high level and more complex features [22]. The first layer generally outputs low level features such as edges. The depth of network may depend on the complexity of the problem. Although this is the general view that is accepted, currently deciding on a CNN

architecture does not pass through a single universally accepted path but rather is based on rules of thumb and previous works. In order to construct a CNN, architectural options such as number of layers and their order should be regarded along with the decision of hyperparameter choices. However, instead of designing a CNN from scratch, it is suggested to work on existing ones and modify them accordingly.

In this work, GEINet [23] which is based on AlexNet [24] was chosen as the basis architecture. AlexNet is one of the most popular CNN architectures published and was designed to classify 1.2 million images into 1000 different classes. The architecture consists of five convolutional layers and three FC layers. Table 1.1 shows AlexNet architecture.

There are some important points in this work worth mentioning. AlexNet utilizes rectified linear unit (ReLU) [25] instead of tanh neurons which was the standard way to model a neuron's output and it is reported that the network reaches to same training error rate six times faster compared to network using tanh neurons. In NNs, nonlinear activation functions are used in the hidden layers because a) linear function does not confine the boundaries which may result function to have negative or infinite value that are not desired in some cases b) nonlinear functions enable the model to generalize better. ReLU is one of nonlinear functions and the formula for the function can be seen in Equation 1.8 which changes negative values to 0 and is equal to z when above or equal to zero

$$R(z) = \max(0, z) \tag{1.8}$$

In their work, they introduced Local Response Normalization (LRN) which is a normalization method that helps generalization by performing a “brightness normalization” as they put on output of some layers. They also used overlapping pooling by setting stride amount less than pool size which as they reported reduced error rate.

Finally, they utilized Dropout [26] to cope with overfitting which cancels neurons randomly in order to prevent co-dependency amongst each other which restrains

a neurons ability to learn potent features. Overfitting occurs when the model is able to predict well in the training, however when a new data is fed into model it is not able to perform as good. Except dropout there are some other methods to tackle this problem. L1 and L2 regularization methods are examples of these. These methods help to tackle overfitting problem by penalizing large weights and therefore increasing generalization [27]. This is done by adding an extra term to the cost function. Equation 1.9 depicts the term for L1 regularization and Equation 1.10 depicts the term for L2 regularization. The variable λ is called regularization parameters which can be tuned to define the penalty rate in the network and n is size of training set.

$$\lambda \sum_{i=1}^n |w_i| \quad (1.9)$$

$$\lambda \sum_{i=1}^n w_i^2 \quad (1.10)$$

Table 1.1. AlexNet Architecture

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	11 * 11	96	4	
Max Pooling	3 * 3		2	
LRN				
Conv2 + ReLu	5 * 5	256	1	2
Max Pooling	3 * 3		2	
LRN				
Conv3 + ReLu	3 * 3	384	1	1
Conv4 + ReLu	3 * 3	384	1	1
Conv5 + ReLu	3 * 3	256	1	1
Max Pooling	3 * 3		2	
Dropout (rate = 0.5)				
FC6 + ReLu (unit number = 4096)				
Dropout (rate = 0.5)				
FC7 + ReLu (unit number = 4096)				
FC8 + softmax (unit number = 1000)				

GEINet on the other hand is a CNN architecture designed for person re-identification based on gait. Although the architecture is based on AlexNet, GEINet is smaller than AlexNet due to the fact that AlexNet was designed to tackle problems in a wider area. Table 1.2 shows GEINet architecture.

The input image for GEINet is a GEI with size as 88 x 128 pixels. This is shown to be the reason for a smaller architecture because AlexNet is capable of classification under different circumstances and for different object whereas GEINet is supposed to function on GEI which is a preprocessed image and includes only humans. GEINet utilizes important factors such as Dropout and LRN that made AlexNet successful.

The purpose of networks is not only to recognize the individual but to do this in a view invariant manner. Therefore, the dataset that was used for this studies' experiments has GEIs that are constructed from footage taken from different angles.

It is also important to note some of the hyperparameter settings of this study. The loss function was selected as cross-entropy loss because this is a non-binary

Table 1.2. GEINet Architecture

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	7 * 7	18	1	
Max Pooling	2 * 2		2	
LRN				
Conv2 + ReLu	5 * 5	45	1	
Max Pooling	3 * 3		2	
LRN				
FC3 + ReLu (unit number = 1024)				
Dropout				
FC4 + Softmax (unit number = # of individuals in the dataset)				

classification and the optimization algorithm was chosen as stochastic gradient descent (SDG) [28] with mini batches of 239 and initial learning rate of 0.2.

1.5 Contribution of the Thesis

The contribution of the thesis can be explained in twofold: gait based gender recognition and gait based age estimation. However, the main contribution of both parts is to investigate the feasibility of CNN for gait based gender recognition and age estimation problem with GEI being input, which is an understudied topic in the literature, and proposal of high performance CNN architectures. During performance evaluations, computational cost was regarded as well.

Although CNN has recently received a great deal of attention, there are few studies that employ CNN for gait based gender recognition. Marin-Jiménez et al. [29] proposed a multi-task CNN architecture which use optical flow channels as input. The main task in this work is person re-identification based on gait and gender recognition is an auxiliary task. Liu et al. [30] explored combination of CNN and Support Vector Machine (SVM) to recognize gender based on gait feature. In our study, application of CNN for gait based gender recognition is explored with the aim of high accuracy and low computational cost. In this context, different CNN architectures were analyzed and further adjustments were made in terms of architecture and hyperparameters. The proposed architecture outputs results that are comparable with the state-of-the-art methods and important insight about the problem.

On the other hand, there are no studies in the literature for gait based age estimation with GEI being the input. While another auxiliary task in the work [29] is age estimation, the input is optical flow channels instead of GEI. Thus, the contribution of this part of the work is mainly exploration of utilizing CNN in this context. Similar to gait based gender recognition with CNN, several different architectural options were evaluated and results showed that although the performance falls behind some methods available in the literature, it is still a viable option for gait based age estimation.

1.6 Outline

Outline of the thesis is as follows: Chapter 2 presents literature review with short summaries of work done. Chapter 3 shows evaluated architectural options and hyperparameters to achieve the final architecture for gait based gender recognition. In a similar manner to that, chapter 4 demonstrates the feasibility of CNN for gait based age estimation by assessing several architectures and fine-tuning hyperparameters. In chapter 5, performances of final architectures for both problems are compared with the existing methods in the literature and the results are discussed. Finally, the work is concluded in chapter 6.

2 LITERATURE REVIEW

2.1 Gait Based Gender Recognition

There have been various studies on gait based gender recognition. While some of the studies focus on developing new methods for gender recognition, others focus on addressing existing challenges in gender recognition and approach the case from another perspective. One of the well-known challenges is recognition when clothing and carrying conditions change. Hu et al. [31] addressed this issue and applied Gabor wavelets for feature extraction. In order to construct the feature space, features are mapped to a low dimensional subspace by Maximization of Mutual Information (MMI). These Gabor-MMI features are then classified with Gaussian Mixture Model - Hidden Markov Model for each class. The experiments were done using CASIA B Dataset on 31 females and 31 males and results show 96.77% correct classification rate.

Another challenge is recognition from multiple views and several studies addressed this. Makihara et al. [32] introduced a new multi-view gait database which includes gait of 168 people with different genders. They performed gender recognition on this database by extracting frequency-domain features from individuals' gait cycles and classified these features with K-nearest neighbor (KNN). The classification results indicate correct classification rate of 80% among genders. The results also indicate that the view used for classification affects the outputs for different classes. Chen et al. [33] segmented gait silhouette images which are captured from different angles into eight regions. Weights are calculated for each region and a template is created for both genders. Euclidian distance is employed as the metric and the distance between the probe and the constructed gender templates that does not include the probe is calculated for each subject. Smaller distance to either gender template indicates the subject's gender. Experiments done on IRIP dataset displays 93.3% correct classification rate for gender recognition. Lu and Tan [34] proposed uncorrelated discriminant simplex analysis for gait recognition and gender recognition from different views. They used GEI as the feature and mapped features into low dimensional subspace by decreasing distance between intraclass samples and increasing distance between

interclass samples. KNN simplex metric was employed to characterize the distance and uncorrelated constraint was imposed. Experiments done on CASIA B Dataset shows that the best performance of the method was 79.8% accuracy for gender recognition. Chang and Wu [35] employed two-dimensional discrete cosine transform to extract features and train them with embedded hidden Markov model for gait based gender recognition. The experiments done on CASIA B Dataset show 94% accuracy. In Troje's [36] study retroreflective markers were attached onto subjects and they were recorded while walking. Using the markers, major joints of the body were marked virtually and thus posture of the subject could be represented. Troje applied a two-staged principal component analysis (PCA) and used a linear classifier for gender recognition. The approach achieved 90% correct classification. The study showed that frontal views are better for gender recognition. Huang and Wang [37] studied the gender recognition on multi-view gait sequences. Specifically, they chose 0° , 90° and 180° views with the reason that these views are capable of representing gender attributes on human gait better. 0° and 180° viewed silhouettes were separated into five regions and ellipses were fitted to represent the body parts, 90° viewed silhouettes were separated into seven regions and ellipses were fitted the same way. They extracted features from each region and calculated the Mean Euclidian Distance for each view and gender. Obtained similarity measures were concatenated for each gender and thus view invariant feature representation for each gender were created. Finally, the sum rule was employed for the fusion scheme. The fused feature set was classified using SVM and experiments done on CASIA B Dataset showed 89.5% recognition rate. Zhang and Wang [38] conducted their study in order to find out effects of different views on gender recognition. In order to do this, they introduced the IRIP Dataset which consists of gait data captured from different angles and the dataset is balanced in terms of gender. For the analysis of their data, they utilized the method of Liu et al. [11] with the exception of using gait cycle for feature extraction instead of the whole gait sequence. They used SVM for classification and the results indicate that frontal part of the body provides better cues for gender recognition. Zhang and Wang [39] fused multi-view gait sequences in order to classify human gait according to gender. The fusion was performed at feature level and GEI is used as feature. After the fusion multiple PCA were administered in order to cope with the high dimensionality. Experiments

were done on 31 subjects chosen from CASIA B Dataset and SVM was used as classifier. The results show 98.1% correct classification rate.

Studies on multiple views consider that the walking direction of the observee does not change. However, in real life scenarios that may not be true and the users direction may change before an assesment can be made. Hence, to address this issue, some studies were conducted on gait based gender recognition in uncontrolled environements. Unlike other works in which experiments are done on people walking in controlled environments, the work of Lu et al. [40] focuses on classification of gender based on uncontrolled gait. The method they proposed start by extracting silhouette of a person walking and then the extracted silhouettes are clustered by their view in order to construct averaged gait image as it should be done from single view. After having the set, averaged gait images are clustered with point-to-set distance metric. The results show that the correct classification rate using this approach is 91.3% and they showed that proposed approach outperforms other metrics such as neighborhood component analysis, large margin nearest neighbor and information-theoretic metric learning. Lu et al. [41] developed a new metric learning method called sparse reconstruction based metric learning. The basis of the method is to minimize the intra-class sparse reconstruction errors and maximize the inter-class sparse reconstruction errors. The method employs point-to-set distance to learn distance metric. Experiments were done on USF Dataset, CASIA B Dataset and a dataset they constructed in this work called ADSC-AWD. Results show commensurate performance with the existing approaches while the gender recognition done on the introduced dataset displays 93.1% correct classification rate. Zaki and Sayed [42] proposed a method to classify pedestrians by gender. They extracted step length, step frequency along with gait variations and stability indicators which decompose into several metrics. KNN and Semi-supervised learning are employed as classifiers. For experiment purposes pedestrian data was collected and experiments show 94% correct classification rate for gender classification.

It is also possible to approach gait based gender recognition as a real-time problem and some studies provided efficient methods for this problem. Chang et al. [43] approached the gender recognition problem as a real-time problem. They

proposed a method to achieve good results in real-time view angle invariant gender recognition. In order to make the method invariant to angle, they classified GEI of CASIA B Dataset according to viewing angle by utilizing Linear Discriminant Analysis (LDA). Then they used Fisher-Boosting algorithm to classify for gender for each angle. Gender recognition performance of this algorithm is 97.32% when the dataset is classified into 11 angles manually. However, the performance drops when the proposed automated angle classification scheme and gender recognition approach is combined due to lower accuracy of the angle classification. In this situation, best results are achieved with dataset being classified into five angles which is 96.79% accuracy. KalaiSelvan and Raja [44] approached the gait based gender recognition problem as a real-time problem therefore proposed a fast feature extraction scheme. The approach includes a human detection and tracking method robust for crowded environments. Once a subject is detected and tracked, GEI is constructed and partitioned into five regions. PCA is applied for dimensionality reduction and features are classified by using KNN. Experiments were conducted on CASIA B Dataset and TUM-IITKGP and 94.5% accuracy was achieved.

Experiments are generally carried out on single dataset and the performance of the proposed methods are not evaluated by training algorithms with one dataset and testing with another. Guan and Wei [45] carried out experiments to observe this. They set out an experiment to find out performance of well-studied algorithms for gender recognition based on face and gait between different datasets in cross-dataset manner. As authors describe, most of the experiments are done on single dataset, however real-world scenarios may contain different inputs. On the gender recognition by gait part of the study, they tested PCA + LDA and SVM approaches between two datasets which are CASIA-B and USF dataset. Results indicate that performance of these approaches drop significantly and this shows that although an approach may provide promising results in an intra-set test, it may not in a real-world scenario.

In addition, the state of dataset in terms of balance is an important case to consider and this may have direct effect on results. Some studies pointed to this problem and proposed methods to deal with class imbalance problem in gait

based gender recognition domain. Félez et al. [46] studied high dimensionality and class imbalance problems which can occur during gait based gender recognition. Gait representations such as GEI provide high dimensional data. In order to address this issue feature selection approach is proposed and thus RELIEF method and threshold-based feature selection were chosen. Some of the available datasets are imbalanced in terms of gender which may cause ambiguous outputs. Resampling the dataset is a frequently referenced approach to overcome this issue and in order to study it in this problem domain, Synthetic Minority Over-sampling Technique and Random Under-Sampling methods were chosen. Combination of resampling, feature selection and original data were tested out and results show that first resampling then feature selection provides the most stable approach.

Some of the studies approached the case from a new dimension and used depth information in order to infer gender of subjects. Borrás et al. [47] studied gender recognition on the dataset they introduced which includes depth information. They extracted 2D features by calculating Euclidian distance between randomly selected points on silhouette border, composing histogram of these results and finally building up a piecewise linear function. 3D features are extracted using the same scheme with the exception of selecting points on silhouette surface. After these, they performed PCA+LDA for dimensionality reduction. Experiments were done on the dataset they introduced which is called DGait using SVM classifier and the results show that utilizing depth information makes the classification more robust to view changes and the combination of 2D and 3D features provides better performance. Igual et al. [48] proposed a new approach for gait based gender recognition using depth information. For 2D feature extraction, they divided the GEI into five parts representing body parts and applied PCA+LDA. For 3D feature extraction, they formed 3D cloud points, aligned these points according to the centroid, applied PCA and projected the points to PCA plane, constructed 3D histogram, divided the histogram into five parts and finally applied PCA+LDA. They utilized SVM for classification and conducted experiments on DGait Dataset. Results show 93.90% accuracy for 2D features and 98.57% for 3D features when experiments are conducted on labeled cycles which does not include different views. Experiments were also done on real-conditions and 93% accuracy was

achieved. Hoffman et al. [18] introduced the TUM GAID Database which includes captured RGB images, depth images and audio stream of subjects. They applied various experiments on this database. In one experiment to recognize gender, they constructed Depth Gradient Histogram Energy Image which is constructed by calculating depth gradients and accumulating into direction histograms. SVM was applied to classify into male or female and the results show 95.8% accuracy.

Another way to approach gender recognition is to use multiple cues from subjects. In their work, Shan et al. [49] showed a gender classification method based on fusion of human gait and face. GEI is used as the gait feature and authors indicate that details of facial gender classification of this study is left out to future research. They proposed combining gait and face features by utilizing Canonical Correlation Analysis (CCA). Experiments are done on CASIA-B Gait Database and SVM is used as the classifier. The results indicate that fusion of human gait and face has higher performance than studies that do not use fusion of these features. On top of this, it is shown that using CCA instead of simple fusion techniques such as concatenation provide more promising results. Zhang and Wang [50] approached the gender recognition problem by fusing both gait and face information. They captured both face image and side-view gait image sequence of subjects. For the face based gender classification, PCA + SVM approach was employed. For the gait based gender classification, they applied the procedure in work [51] and fitted seven ellipses on the human silhouette which represent the human body parts. Features were extracted from these ellipses and classified by using SVM. Results of both classifications were fused by the sum rule. Result of experiments show 90% accuracy for both face and gait based classifications when done separately, and 93.33% accuracy when two classification results are fused.

Humans may sometimes perceive gender of people wrongly under some circumstances. Davis and Gao [52] proposed an adaptive three-mode PCA in which modes describe posture, time and gender. The method is capable of recognizing both real and perceived gender of the subject therefore it is adaptive. Experiments done on gait of 40 subjects display recognition rate of %92.5 for real gender.

Considerable amounts of studies that are based on recognizing from gait rely on extracted features of gait matchers. DeCann et al. [53] explored clustering of human gait based on three gait matchers namely GEI, Gait Curve Matching and Frieze Pattern Matching. K-means algorithm is used for clustering and the distance metrics used for each are as follows: Euclidean for GEI, Procrustes for Gait Curves and Dynamic Time Warping for Frieze Pattern. CASIA B Dataset is used in this study and clustering is performed for human identification as well as other physical attributes such as gender, stride, cadence, height and area. In experimental results, it is shown that body area and gender are best clustered and the gender clustering is best done with GEI.

Other works proposed new methods to recognize gender or explored feasibility of combination of well known feature extraction methods and machine learning algorithms on gait based gender recognition. Lee and Grimson [51] divided silhouette of subjects into seven regions by fitting ellipses and computed the difference of shapes caused by person's movements for each region. They applied ANOVA to select best features among these and applied SVM to classify the subject with gender. The best result being 84.5% accuracy was obtained by SVM with linear and polynomial kernels. Yu et al. [54] approached the subject of gait-based gender classification problem by combining human knowledge with their method. They surveyed human observers on an experiment to find out the most influential factors to determine a person's gender. According to results they divided GEI into five parts and assigned different weights to each of them according to their influence on gender recognition. They applied linear SVM for gender classification and obtained 95.97% correct classification rate. Li et al. [55] proposed a method to divide the average gait image into seven parts which represent human body parts. For each part median of Euclidian distance is calculated between the probe and the gallery. The similarity is classified for gender using SVM. The experiments are done on 12 probes from USF Gait Database and results display effect of various body parts on gender recognition. According to the results, while head, back-leg, and feet has no positive effect on gender recognition, body and front-leg usually have positive effect. These results show that some parts of human body can be ignored for gender recognition. Yoo et al. [56] constructed human stick figure from captured gait images. By using this, they

extracted temporal and spatial gait features and classified them for gender by using SVM and Neural Networks. Experiments were conducted on 100 subjects and the results show that SVM performed 96% accuracy while NN performed 92% accuracy. Felez et al. [57] set out to provide a method to mark human body parts in the human silhouette for better representation and their work is based on work [51]. Thus they separated silhouettes into eight regions. Experiments were done on CASIA Dataset by utilizing SVM for gender recognition. The results show 94.7% accuracy with the proposed method while method proposed in [51] displays 93.33% accuracy for the whole dataset.

Hu et al. [58] employ conditional random field for gender recognition but they mix spatial and temporal features of human gait in order to do so. Thus they proposed the mixed conditional random field method for gait based gender recognition. Experiments were done on CASIA and IRIP Dataset. The results show 98.39% correct classification rate for CASIA dataset and 98.33% correct classification rate for IRIP dataset. Handri et al. [59] used 2D discrete wavelet transformation and 2D fast Fourier transformation to extract gait features. They utilized an AdaBoost algorithm to classify these according to gender. Modest AdaBoost algorithm displayed 93.4% accuracy for gender classification. Oskuie et al. [60] introduced a new method based on Radon Transform of Mean GEI which is a spatio-temporal gait representation. They extracted Zernike moments from this and trained features using SVM. Experiments done on CASIA B dataset show 98.94% correct classification rate. Hu et al. [61] introduced a new pattern for representing the human gait. This new representation, called Gait Principal Component Image, is similar to GEI but instead of just averaging gait silhouettes, the method utilizes PCA and therefore bring forth the variations of human gait in terms of body parts. The method was tested on the IRIP Gait Database and KNN was employed as the classifier for gender recognition. Results show 92.33% correct classification rate. Hassan et al. [62] proposed using lifting 5/3 wavelet filters for gender classification. They obtained outer contour of GEI, extracted features using lifting 5/3 wavelet filters, applied PCA for dimensionality reduction and classified using C4.5 Decision Tree Algorithm. Experiments done on OU-ISIR Gait Dataset show 97.5% accuracy and on CASIA-B Dataset show 97.98% accuracy. Livne et al. [63] studied gender recognition by performing 3D pose tracking. They achieved tracking by applying

Annealed Particle Filter. They extracted motion features by tracking 3D joint positions. Logistic regression was applied for learning and results show that success rate of gender recognition using 3D model is 0.90. Lawson et al. [64] calculated amount of translation in human gait frames and applied Spatiotemporal Independent Component Analysis in order to extract features. Using these, subjects are classified into male and female by utilizing nearest neighbor classification. The accuracy of the proposed method is 86%. Arai and Andrie [65] approached the gait based gender classification problem by generating skeleton model of human gait. They constructed human silhouettes and obtained skeleton model of silhouettes using morphological operations. Major body points and angles were extracted from the skeleton model. These were used to classify the subject according to the gender and SVM was utilized to achieve this. Experiments done on CASIA B dataset shows 85.33% accuracy. Handri and Nakamura [66] proposed extraction of silhouette width in the frequency domain. They performed feature selection and classified the subjects by constructing a Choquet Integral Agent Network. Experiments were done on 53 subjects which consist of different genders. The results 78.5% accuracy for gender classification.

2.2 Gait Based Age Group Classification and Estimation

Age is another trait that can be inferred from human gait. Several studies were conducted on this subject. However, unlike the works conducted for gender recognition, these studies generally do not aim to develop methods for different conditions such as multiple-view or clothing variance but rather aim to propose methods for the problem itself. Additionally, there are fewer studies done on age trait compared to gender in the gait biometric domain. The work conducted on to discover age trait by human gait can be categorized under age group classification and estimation.

Gait based age group classification studies focus on classifying subjects into their age group such as children, adult or elderly. Makihara et al. [32] performed age group classification in their work along with gender recognition. They used frequency-domain features and classified subjects into children and adult using KNN with accuracy 74%. Mansouri et al. [67] proposed a new descriptor for age classification in their work. The combined silhouette longitudinal projection and

silhouette transverse projection in order to be able to represent important parameters for age classification in one model. On top of this, they combined this proposed descriptor with two other existing descriptors namely GEI and feature to exemplar distance. They conducted their experiments on OUISIR: large population dataset and used SVM for classification. The results indicate 76.76% correct classification rate between young and elder. The combination of descriptors outperforms the three descriptors when used individually. Beggs et al. [68] described a method to recognize the gait of young and elder people in their work. They extracted Minimum Foot Clearance feature from gait of 30 young and 28 elderly people and classified using SVM and Neural Network. The results show that classifications with SVM and Neural Network yield accuracies of 83.3% and 75%, respectively. When a hill-climbing algorithm is used to select features, accuracy with SVM increases to 90%. Zhang et al. [69] proposed a framework to classify human gait as young and elderly. The framework consists of extracting Frame to Exemplar (FED) distance from gait contour and classifying it with Hidden Markov Model (HMM). The experiments were done on seven people and for comparison purposes they applied FED to both gait contour and silhouette, and classified with HMM and Naïve Bayes. The results show that best performance was obtained with gait contour and HMM combination. The correct classification rate of this experiment was 83.33% and this indicates that using gait contour is more suitable for age classification as the redundant information is omitted. Handri et al. [59] classified age group along with gender and utilized the same gait features. Modest AdaBoost algorithm displayed 94.3% accuracy for young and elderly classification. Davis [70] provided a method for age classification based on human gait. In this work, reflective markers were attached onto subjects' head and ankle and they were recorded with infrared video camcorder. The images were thresholded in order to obtain point-lights and thus relative stride length and stride frequency features were calculated. Perceptron neural network was trained for classification. The results show that the provided method performs 93-95% correct classification rate for classification between adult and children. Chuen et al. [71] extracted stride length, stride frequency, head length, body length, head-to-body ratio, leg length and stature from GEI by identifying body parts and joints. They classified the feature set into elderly and young by utilizing SVM. Experiments done on 20 subjects picked from Large Population OUISIR gait database show

100% accuracy. Although authors state that this performance may have been a result of doing tests on a small set, it also shows feasibility of the approach. Zaki and Sayed [42] performed young-elderly classification in addition to gender recognition. They used step length and step frequency features, and classified by KNN. They obtained 90% correct classification rate for young-elderly classification.

Gait based age estimation on the other hand focuses on predicting the exact age and this is considered a more challenging problem. Lu and Tan [72] proposed an age estimation framework in which Gabor magnitude and Gabor phase features are extracted from GEI and fused. Features are labeled with both age and gender in order to utilize age feature differences between genders. In order to reduce dimensions, multilabel-guided subspace is learned. Features are then used in multilabel KNN for age estimation. The experiments are done on USF gait database and the results show that proposed framework has 5.42 years MAE. The results also indicate that using gender labels increases performance significantly. Punyani et al. [73] proposed a human age estimation system which combines features from face and gait. The gait features are obtained using GEI and face features are extracted using biologically inspired features. Dimensionality of both face and gait features are reduced by Principal Component Analysis and these features are fused by concatenation. The experiments are done on OU-ISIR Gait Database Large Population Dataset and USF Gait database. The results indicate that MAE is 6.11 years which is lower compared to experiments done using only face or gait. Hence fusion of gait and face for human age estimation provides better results. Lu and Tan [74] propose ordinary preserving linear discriminant analysis (OPLDA) and ordinary preserving margin fisher analysis (OPMFA) in order to overcome weaknesses of LDA and MFA. Experiments were conducted on USF Gait Database and thus GEI was used as gait feature. The results indicate that proposed methods outperform existing manifold analysis methods. OPMFA displays 4.28 years MAE while OPLDA shows 4.65 years MAE. Makihara et al. [75] employed Gaussian process regression for gait-based age estimation. The method is experimented on three gait based features namely GEI, frequency-domain features and gait periods. Results show that best performance is achieved with frequency-domain features with MAE being 8.2 years. Li et al. [76] proposed a method for age estimation based on human gait. The method consists of three

steps which are training SVMs for the age groups, age-group dependent manifold learning and training non-linear Support Vector Regression for age estimation. The benefit of this method is to perform age estimation based on age group, which helps to overcome errors caused by variation among age groups. Experiments done on OULP-Age dataset displays 6.78 years MAE. Lu and Tan [77] introduced cost sensitive versions of principal component analysis (CSPCA) and locality preserving projections (CSLPP) by fusing a cost matrix with the methods. They applied regression with quadratic model on USF Gait Database for age estimation and results show 4.68 years MAE for CSPCA and 4.35 years MAE for CSLPP. Livne et al. [63] obtained motion features by tracking 3D joint positions. Along with gender recognition, they performed age estimation using only motion features and logistic regression. The root mean square error of this approach was 6.63 years. Hoffman et al. [18] used depth information and calculated depth gradients which were collected into direction histograms. Besides performing gender recognition, they also estimated age of subjects by using unpruned REPTrees and results show 3.27 years MAE.

3 GAIT BASED GENDER RECOGNITION

In this part of the study, effects of different architectures and hyperparameters on gender recognition were explored by taking GEINet as the basis. During the experiments, while aiming for high accuracy, the computational time is also considered to be an important factor. The experiments were conducted using Keras [78] on NVIDIA GeForce GTX 1070.

3.1 Architecture Overview

The CNN architecture that is decided for gait based gender recognition task consists of convolutional layer, pooling layer and FC layers as shown in Figure 3.1. GEI is provided as input to the network and the input is convolved with 20 filters which are size of 11 by 11. The convolution takes place with stride of 2 and ReLU is used as the activation function. Convolution layer is followed by max pooling operation with pool size of 2 by 2 and stride of 2. After the pooling layer, FC layer with 512 units is followed. Finally, FC layer with 1 unit is placed and it is followed with sigmoid function. The learning is performed by minimizing the binary cross entropy loss function of which the equation can be found in 3.1 where \hat{y} is the predicted value and y is the label, and Adam [79] is used as optimization algorithm with learning rate 0.00025.

$$\ell(\hat{y}, y) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y})) \quad (3.1)$$

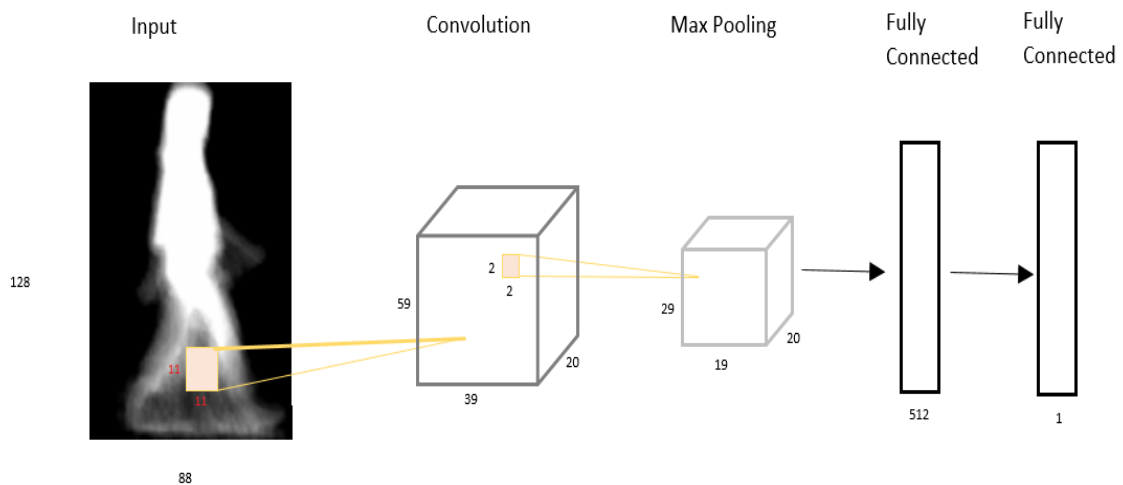


Figure 3.1. The CNN Architecture for gender recognition

3.2 Network Structure

In order to select an architecture, performance of GEINet was evaluated on OULP-Age dataset. The last FC layer of GEINet was modified for binary classification. Initial experiments indicated that the architecture is a suitable choice therefore additional experiments were conducted on this architecture to observe accuracy change. Under some circumstances, it is considered that adding more layers in an architecture assists to tackle harder problems. On the other hand, a network may be larger than the minimum required size and adding layers after some point may not contribute to the objective. In such cases computational resources are wasted and more time is consumed to train the network. Considering these facts, the GEINet was modified to have less layers and this architecture was evaluated. Contrary to this, effect of a deeper architecture was also tested. As authors of GEINet indicate, the network has a similar structure to AlexNet and authors showed that having more layers does not contribute to gait recognition. In order to discover results of a deeper network in gender recognition, AlexNet was evaluated on OULP-Age by adapting it to the dataset. Figure 3.2 shows comparison of GEINet, AlexNet and GEINet without the second convolutional, pooling and normalization layers. Results indicate that these three architectures perform similar in terms of accuracy. However, computation time is another important aspect to consider and an epoch of AlexNet took 91 seconds and GEINet took 50 seconds while the architecture with lesser layers took 36 seconds to complete. Considering these findings, specified layers were omitted from GEINet and further adjustment were made on this architecture which can be seen in Table 3.1.

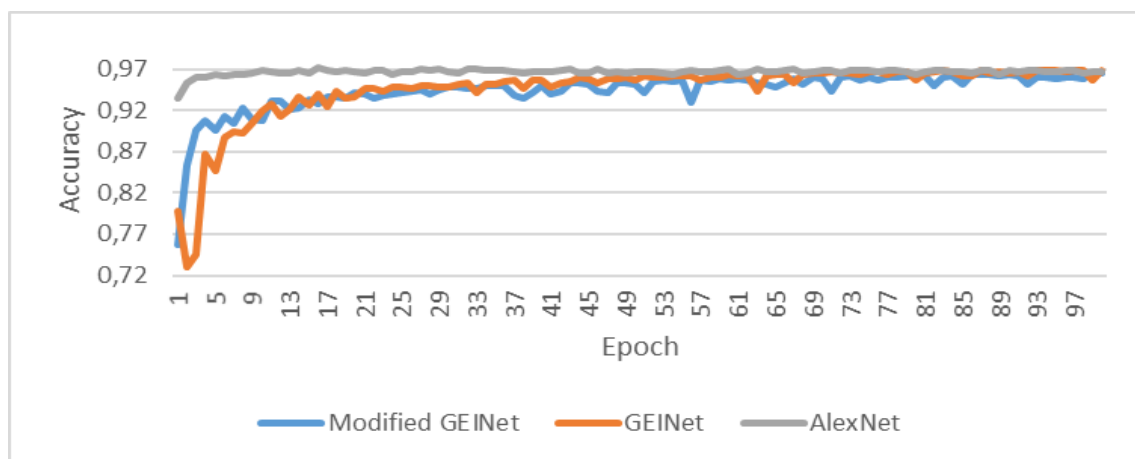


Figure 3.2. Performance comparison of different architectures

Table 3.1. Modified GEINet architecture

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	7 * 7	18	1	
Max Pooling	2 * 2		2	
LRN				
FC2 + ReLu (unit number = 1024)				
Dropout				
FC3 + Sigmoid (unit number = 1)				

The effect of normalization was also evaluated in this context. The aim of normalization is to standardize the training data and therefore to enhance the learning. Normalization aids generalization. One of the normalization methods tested in this study is LRN which was introduced with AlexNet and it was utilized in GEINet as well. Normalization was applied after every activation function in the convolutional layer. In contrast, the architecture without normalization has been tested. Comparison of LRN and no normalization can be seen in Figure 3.3. These results suggest that using LRN and no normalization contributes similar in terms of accuracy. However, using LRN introduces additional computational cost and this makes using no normalization more feasible. Furthermore, LRN causes the training performance to reach closer to the test performance.

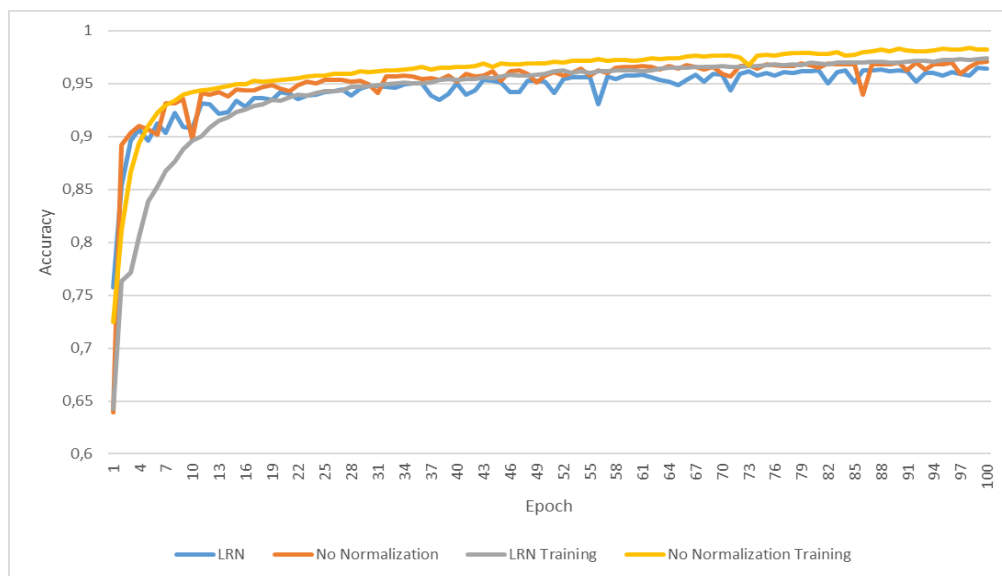


Figure 3.3. Comparison of modified GEINet with and without normalization

Table 3.2. Architecture after removal of LRN

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	7 * 7	18	1	
Max Pooling	2 * 2		2	
FC2 + ReLu (unit number = 1024)				
Dropout				
FC3 + Sigmoid (unit number = 1)				

However, this does not increase the test performance and on the contrary causes the performance to drop slightly therefore LRN is not utilized and the architecture after this change can be seen in Table 3.2

The effect of number of units in the FC layer was also tested out. The last layer is always set as a FC layer with one unit and effects of previous possible FC layers are evaluated.

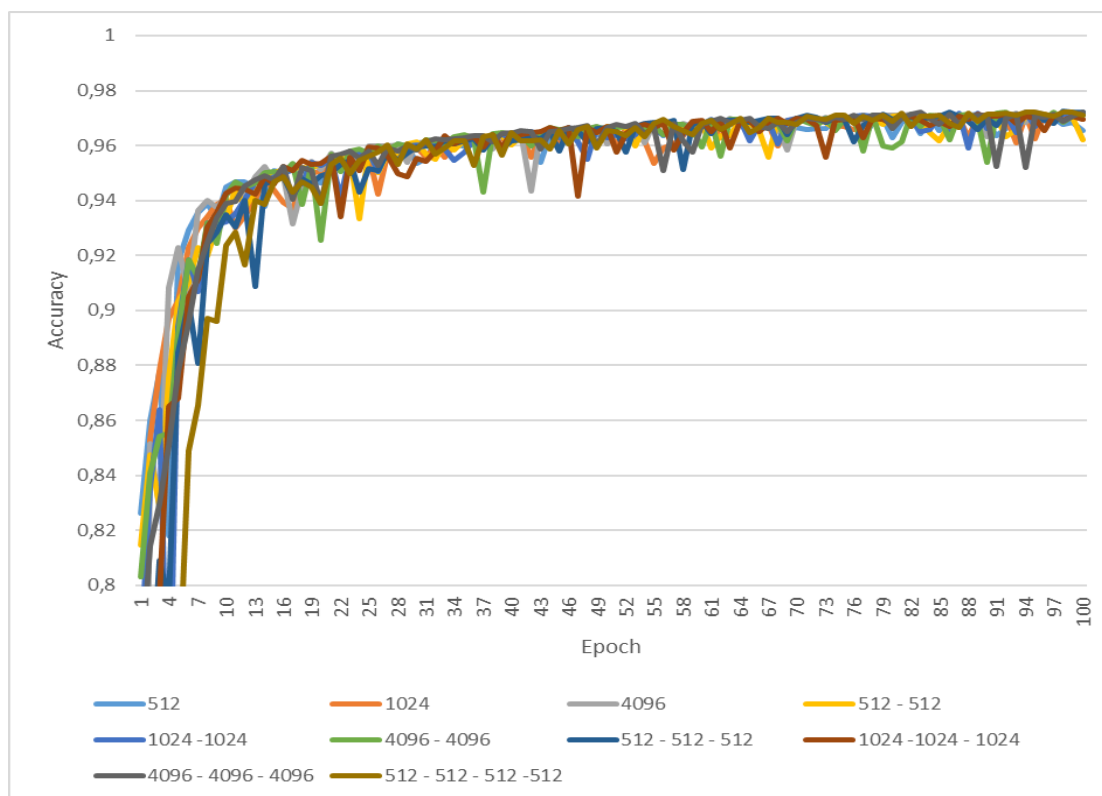


Figure 3.4. Performance comparison of modified GEINet with different FC layer combinations

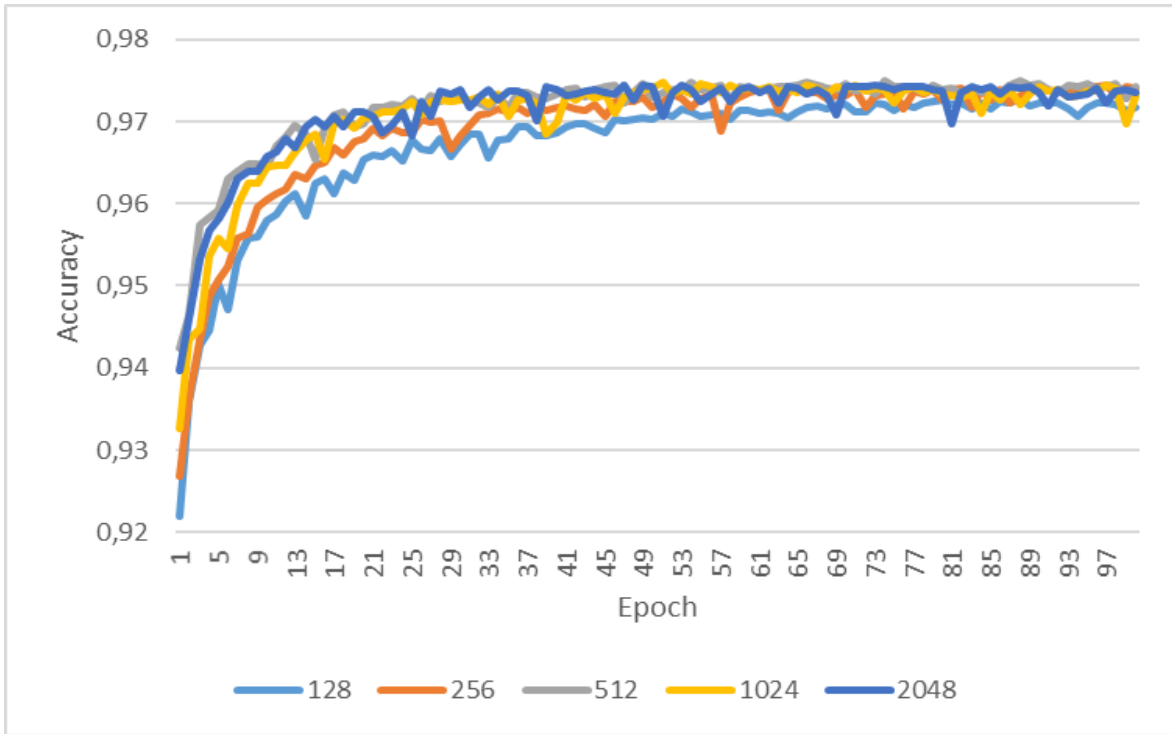


Figure 3.5. Performance comparison of modified GEINet with different FC layer unit numbers

Figure 3.4 displays the result of different FC layer combinations. Legend in this figure shows number of units and number of layers. For example, the label with “1024 – 1024” means two FC with 1024 units. The results suggest that different combinations have similar outputs and do not affect the overall accuracy drastically. Although this is the case, having a larger FC layer increases the total learnable parameter number therefore increases the computational cost. It is found that additional FC layers do not add to the accuracy goal therefore it is decided to have only one FC layer.

Table 3.3. Architecture after FC layer evaluation

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	7 * 7	18	1	
Max Pooling	2 * 2		2	
FC2 + ReLu (unit number = 512)				
Dropout				
FC3 + Sigmoid (unit number = 1)				

Figure 3.5 shows the effect of number of units for single FC layer. Even though 2048 number of units may contribute similar with the rest in terms of accuracy, this introduces unnecessary cost. Considering this, unit size of FC layer was chosen as 512 since it performs best among other sizes and it has close computational cost to sizes that are lower than it. The network structure after this can be found in Table 3.3.

Dropout is a regularization method that helps to reduce overfitting. Performances of the architecture with / without dropout layers were evaluated. The results can be seen in Figure 3.6. The results also include the training performance in order to display the effect of dropout on overfitting. The results indicate that dropout does not have any positive effect on overfitting problem therefore dropout is not utilized in FC layer and the network structure can be seen in Table 3.4.

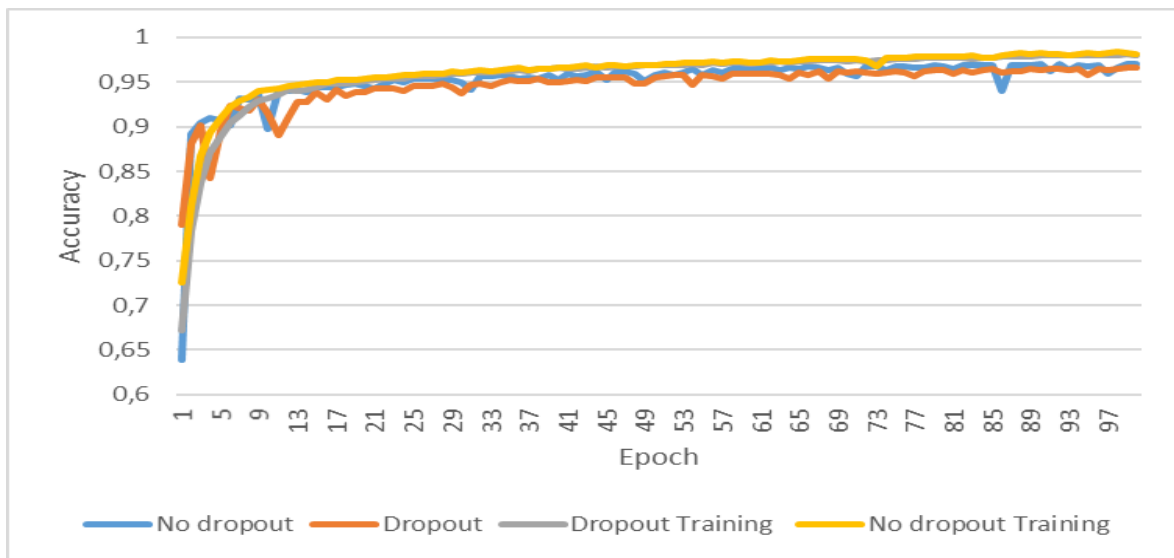


Figure 3.6. Comparison of modified GEINet with and without dropout

Table 3.4. Network structure after dropout experiment

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	7 * 7	18	1	
Max Pooling	2 * 2		2	
FC2 + ReLu (unit number = 512)				
FC3 + Sigmoid (unit number = 1)				

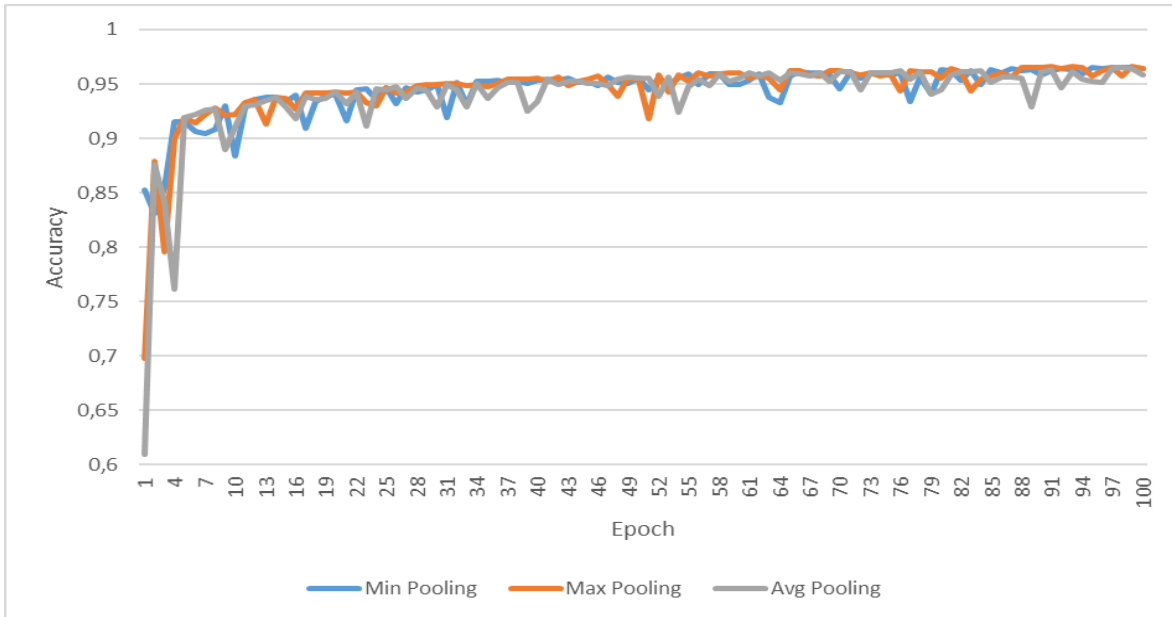


Figure 3.7. Comparison of modified GEINet with different pooling

Finally, different pooling options were evaluated. These options are minimum pooling, maximum pooling and average pooling. Figure 3.7 displays performance comparison of the network using these three pooling methods. According to the results, choice of pooling does not impact the network performance greatly and maximum pooling was selected as pooling method.

3.3 Hyperparameter Optimization

After deciding on number of layers and related regulatory aspects, fine-tuning hyperparameters is considered to provide an important boost to the performance of the architecture. Based on this, several hyperparameters were evaluated with different values. One of the most important hyperparameter is the selection of the optimization algorithm. Choosing a wrong optimization algorithm can drastically reduce the performance of the network. Hence to find out an appropriate one, several common optimization algorithms were evaluated with their initial parameters defined in Keras. These optimization algorithms are SGD [28], RMSProp [80], Adadelta [81], Adamax [79] and Adam [79]. The comparison of these can be seen in Figure 3.8. The results indicate that the performances of optimizers are similar both in terms of accuracy and computational cost. It is considered that Adam requires less intervention to perform better [82]. Therefore, Adam was selected as the optimizer of our architecture.



Figure 3.8. Performance comparison of optimization algorithms

Tuning the learning rate is as important as selecting the correct optimizer. It has capability of directly affecting the learning therefore different learning rate values were evaluated by using Adam optimizer. Figure 3.9 suggests that the network performs similar except lowest and highest learning rates that were tested. Computational cost of these different learning rates does not differ in this network. Thus the best performing learning rate which is 0,00025 was chosen.

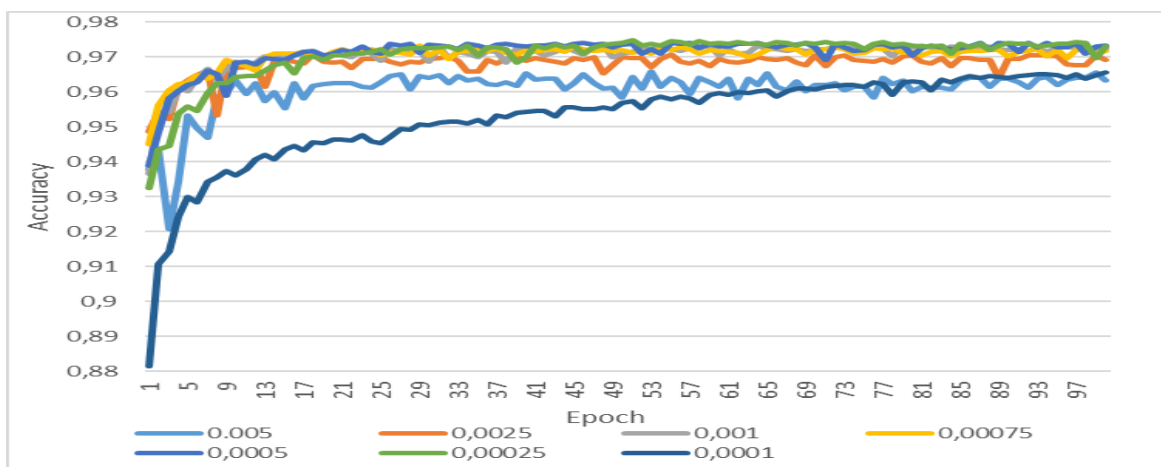


Figure 3.9. Performance comparison of different learning rates

Hyperparameters of convolutional layer were evaluated in order to observe their effect on accuracy and computational overhead. Evaluated hyperparameters are stride, kernel size and kernel number. The results of these with different values can be found in Figure 3.10, 3.11 and 3.12, respectively. Increasing stride reduces the number of trainable parameters therefore helps to reduce overhead. However, a great increase in stride may affect the network's capability of learning in a negative way. Result of different stride values suggests that stride of 3 in this context causes lower capability of learning. Stride of 1 and 2 has similar outcome however using stride of 2 reduces time passed to complete an epoch. Thus stride of 2 was chosen in the convolutional layer. Kernel sizes of 3, 11 and 25 were evaluated and the results indicate that even though the kernel size does not have a drastic effect on the accuracy of the model, larger kernel leads to a better output. However, using too large kernel reduces learning because it inhibits ability to learn from local patches. Considering these, kernel size of 11 was chosen as an optimum. As the final hyperparameter in convolutional layer, the effect of kernel number was evaluated. The results show that kernel number does not make a significant difference in this architecture in terms of both accuracy and computational cost therefore the kernel number was chosen as 20 which adds best to the performance of network. The final state of the network can be found in Table 3.5

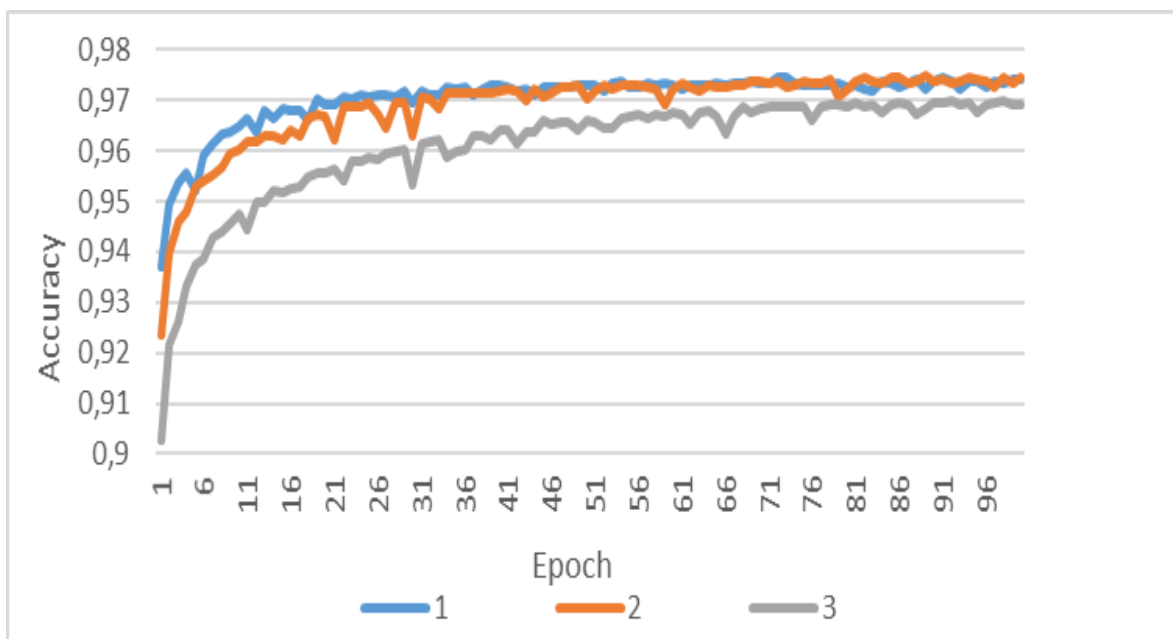


Figure 3.10. Performance comparison of stride on convolutional layer

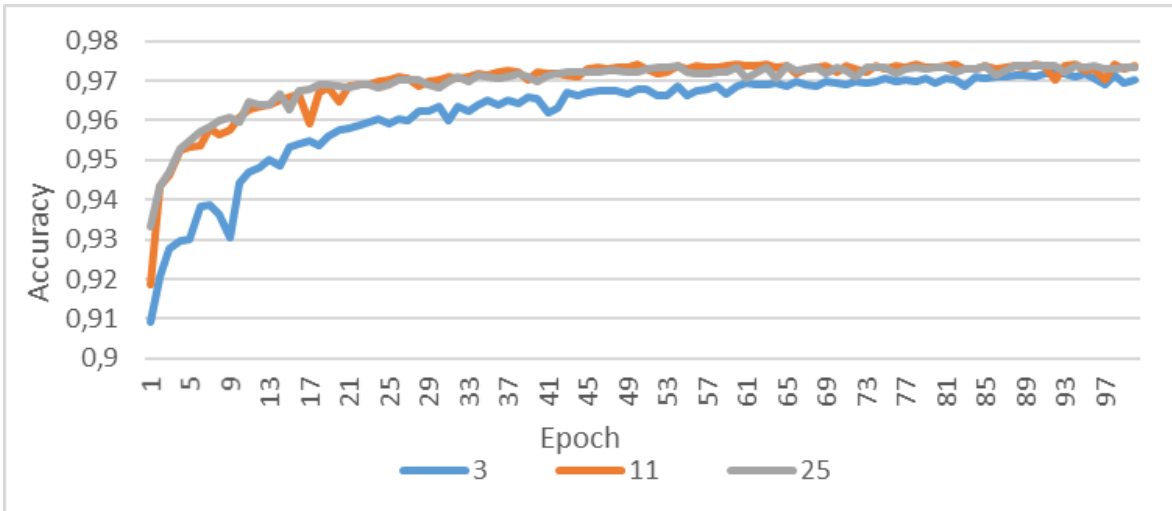


Figure 3.11. Performance comparison of kernel size

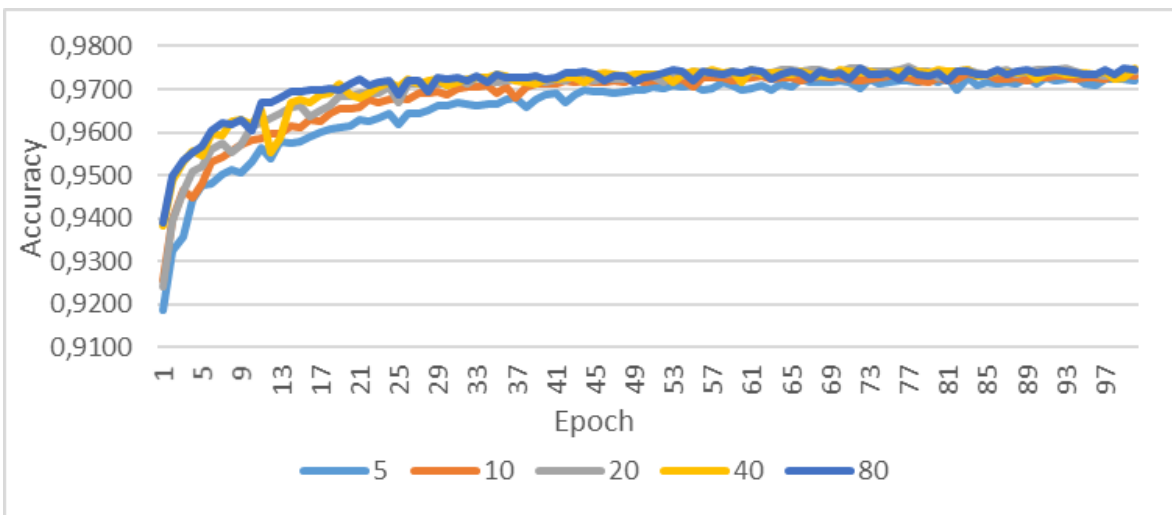


Figure 3.12. Performance comparison of kernel number

Table 3.5. Final network structure of gait based gender recognition

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	11 * 11	20	2	
Max Pooling	2 * 2		2	
FC2 + ReLu (unit number = 512)				
FC3 + Sigmoid (unit number = 1)				

Additionally, the effect of batch size was evaluated. Batch size hyperparameter has capability of affecting the learning of the model therefore different values were evaluated to find out the optimal batch size for this problem.

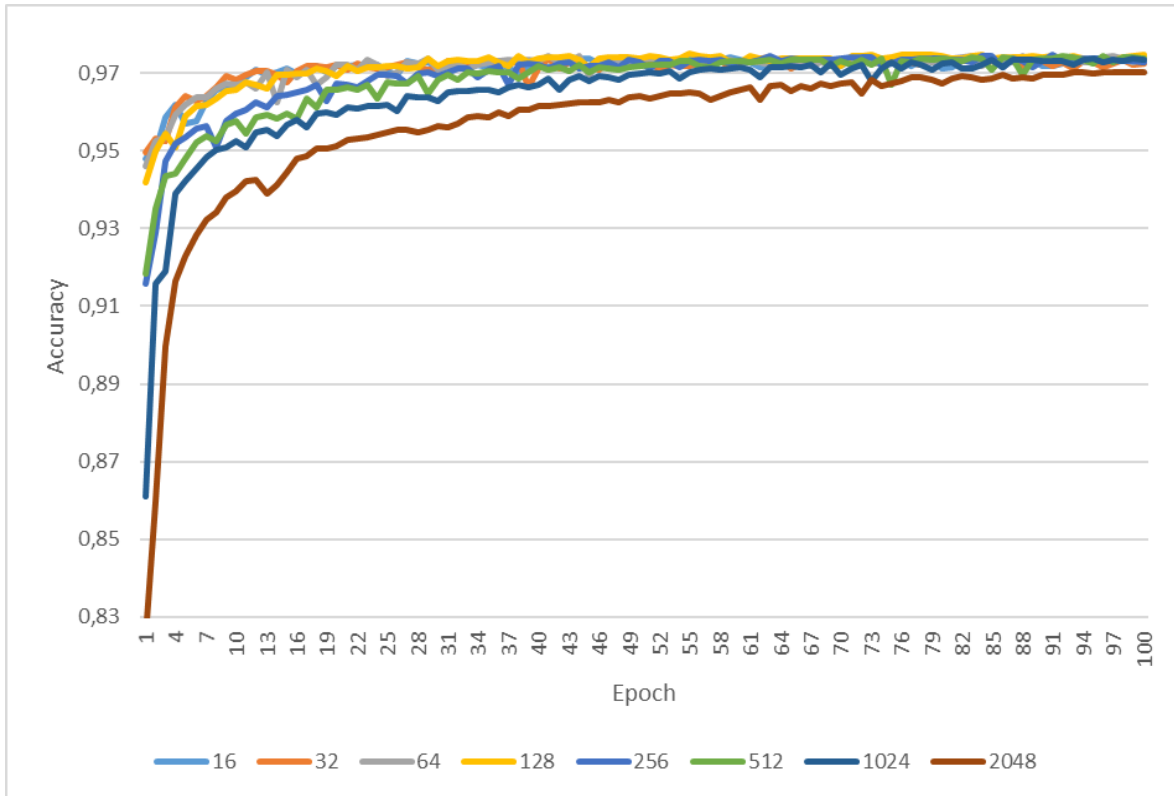


Figure 3.13. Performance comparison of network with different batch size values

Results of these values can be found in Figure 3.13. This suggests that performance of the network decreases as the batch size increases after some point. However, the time to complete an epoch decreases as the batch size increases. According to the results, highest accuracy was achieved with batch size of 128 and the accuracy starts to decrease after that increasing batch size further. Thus, batch size of 128 was selected.

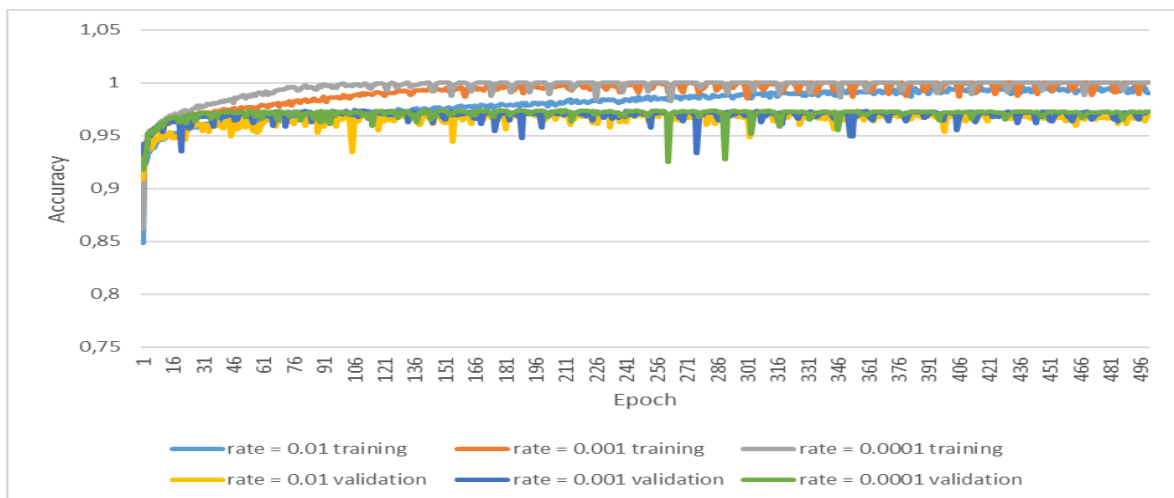


Figure 3.14. Performance of the network with L2 regularization

Finally, effects of L2 and L1 regularization were tested in this problem context and the results can be found in Figure 3.14 and 3.15 respectively. The results suggest that these regularization methods do not contribute to the network and an improvement in accuracy could not be observed.

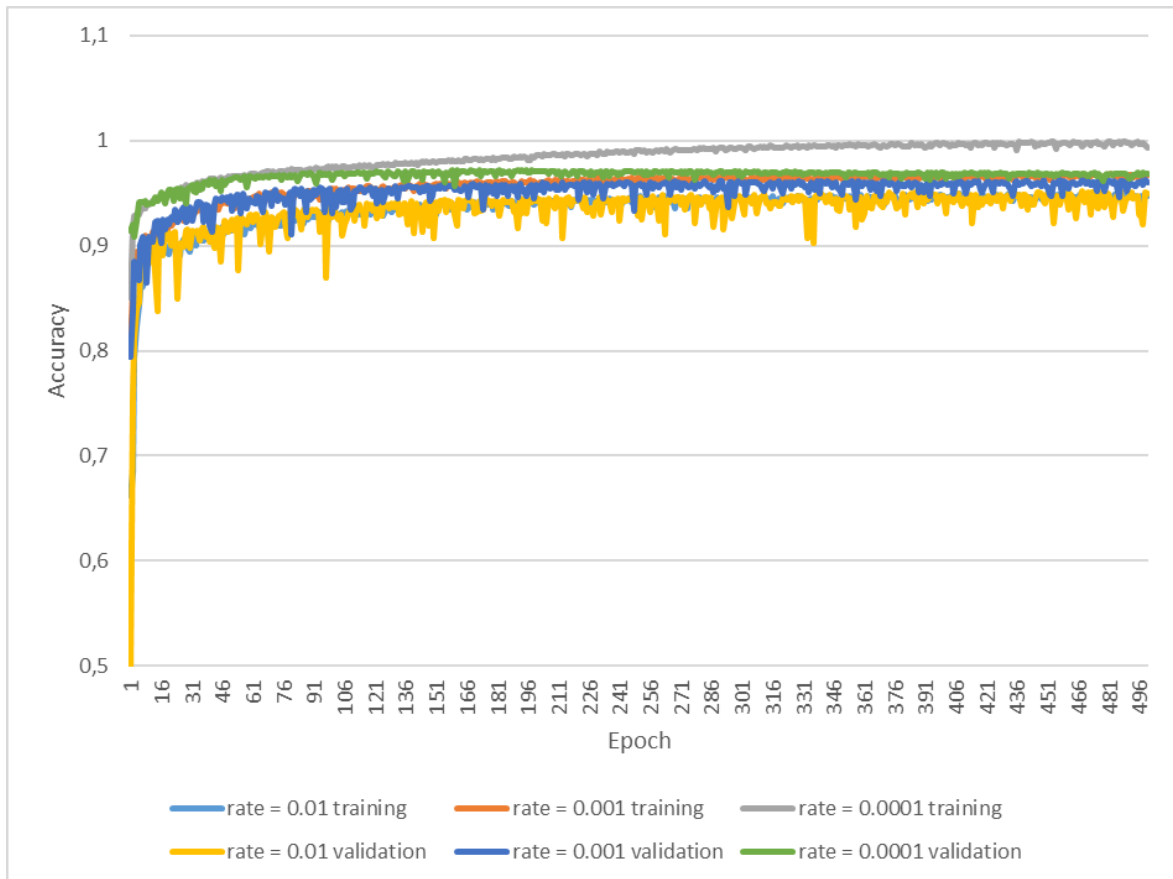


Figure 3.15. Performance of the network with L1 regularization

4 GAIT BASED AGE ESTIMATION

Similar to experiments done for gait based gender recognition, effects of different architectures and hyperparameters on age estimation were explored by taking GEINet as the basis. This time the aim is to come up with a network that can achieve low MAE and again computational time was selected as an evaluation trait. The experiments were conducted using Keras on NVIDIA GeForce GTX 1070.

4.1 Architecture Overview

The CNN architecture that is decided for gait based age estimation task consists of two convolutional layers, one pooling layer and three FC layers as shown in Figure 4.1. GEI is provided as input to the network and the input is convolved with 18 filters which are size of 5 by 5. ReLu is used as the activation function. Convolution layer is followed by max pooling operation with pool size of 2 by 2 and stride of 2. This is followed by a second convolutional layer that has 45 filters which are size of 11 by 11. A max pooling operation follows with pool size of 3 by 3 and stride of 2. After these layers, three FC layer with 512 units are followed. Finally, dense layer with 1 unit is placed and it is followed with linear function. The learning is done by minimizing the MAE and equation 4.1 describes it. Adamax is used as optimization algorithm with learning rate 0.00075 and L2 regularization is employed with rate 0.0001.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - \hat{y}| \quad (4.1)$$

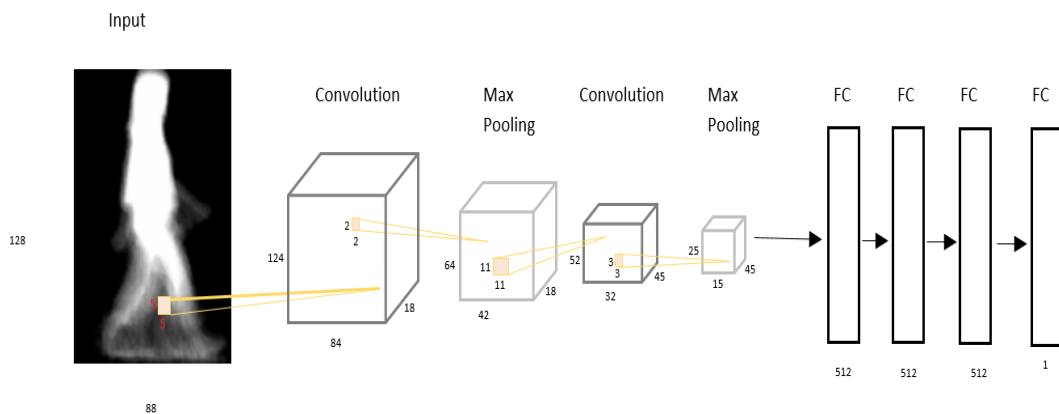


Figure 4.1. Overall architecture of gait based age estimation

4.2 Network Structure

GEINet and AlexNet were evaluated on OULP-Age dataset by modifying their last FC layer to have one unit and to utilize linear function in order to perform regression for age estimation. As GEINet is formed by taking AlexNet as basis, performance comparison of these two networks shows if a larger network is required to tackle the problem. In order to verify if a smaller network than GEINet performs well, second convolutional layer of GEI was removed and experiments were conducted on these three network options. Figure 4.2 shows results of this comparison. The results indicate that GEINet performs the best among three network options with 6.43 years MAE. The time to complete an epoch for AlexNet is 128 seconds, for GEINet is 25 seconds and for the modified GEINet is 20 seconds. Although the modified GEINet and GEINet seem to perform similar in terms of MAE, and modified GEINet comes forward in terms of computational time, the cross evaluations done between network structure options show that additional layers provide a slight gain in performance, and computational cost between the modified GEINet and GEINet can be ignored. Thus, GEINet was chosen among these three architectural options.

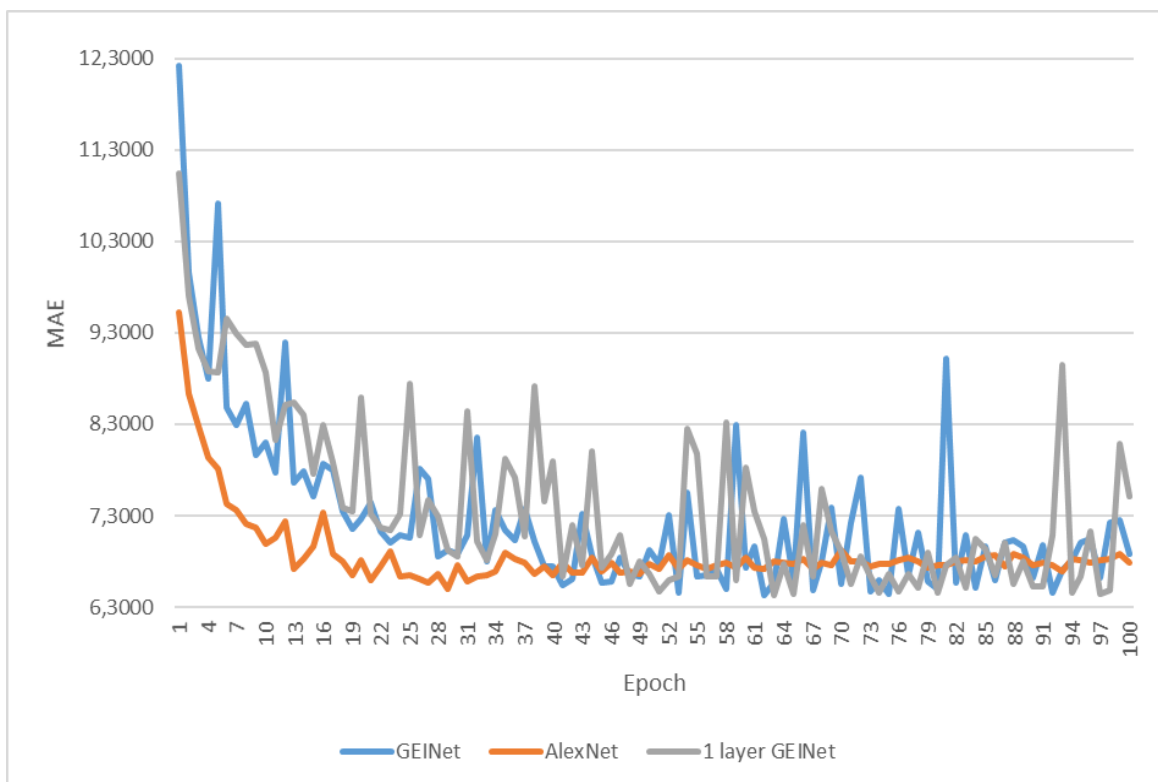


Figure 4.2. Performance comparison of different architectures

Performance difference of GEINet with and without normalization was evaluated in this problem context. GEINet utilizes LRN which helps to adjust the input to be normalized among rest of the inputs and helps to generalize the data. Figure 4.3 displays results of this comparison. Performance of the network during training is also shown in the figure for both scenarios. Results show that test performances of both approaches are close. In the case with normalization, even though the training performance converges better to the test performance than the case without normalization, this does not occur in a way to increase the test performance. Thus, this indicates that LRN does not contribute to the performance of the network. On top of this, an epoch with LRN takes 25 seconds while an epoch without it is 15 seconds therefore the approach without normalization is adopted for this problem context and the network structure after this decision can be found in Table 4.1.

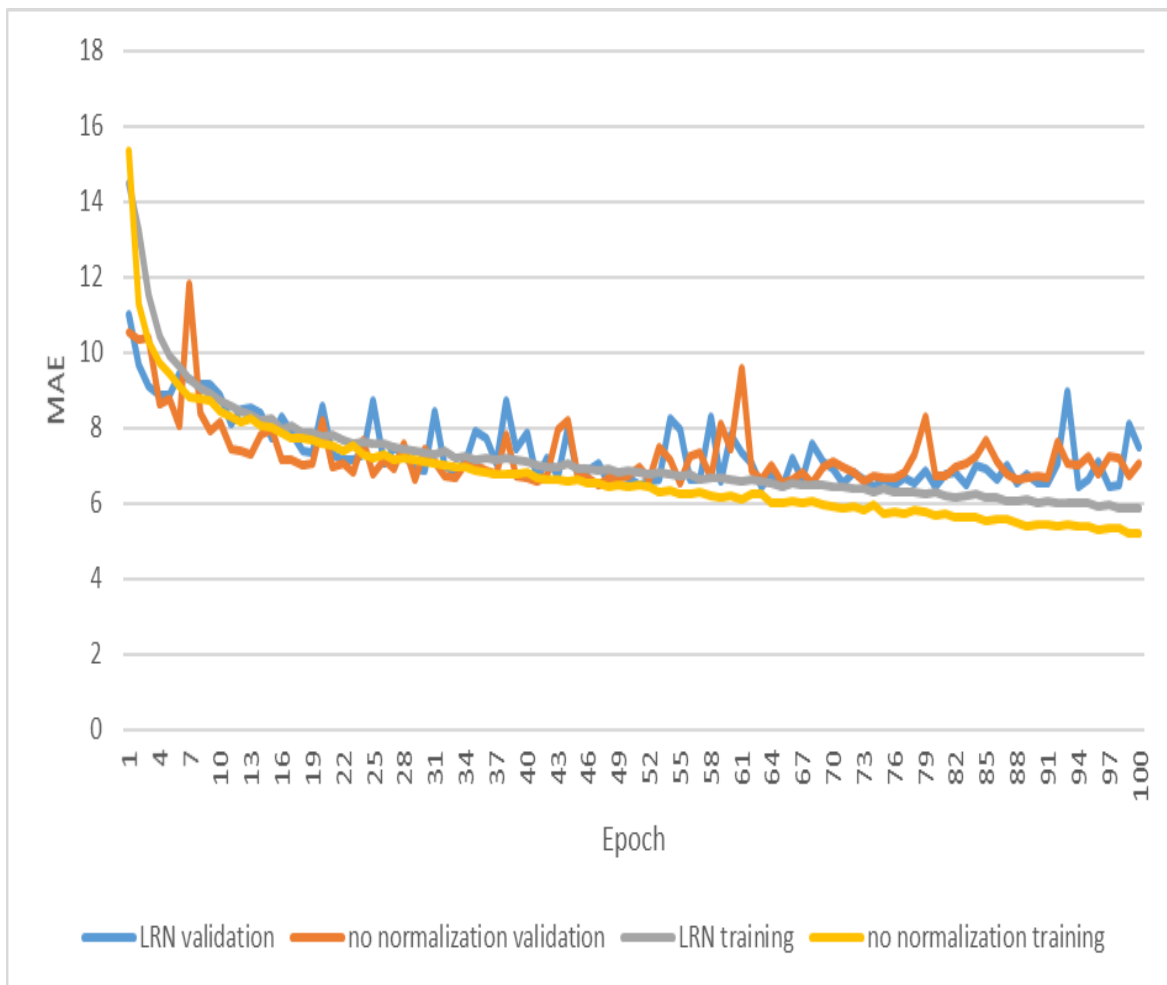


Figure 4.3. Performance comparison of network with normalization and no normalization

Table 4.1. Structure of network without normalization

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	7 * 7	18	1	
Max Pooling	2 * 2		2	
Conv2 + ReLu	5 * 5	45	1	
Max Pooling	3 * 3		2	
FC3 + ReLu (unit number = 1024)				
Dropout				
FC4 + Linear Function (unit number = 1)				

Another experiment done to decide on network structure was testing different combinations of FC layer numbers and their unit numbers. Results of this can be seen in Figure 4.4. In this problem context, the FC layer proved to be an important factor because different combinations of FC create noticeable difference in MAE. Results show that increased number of FC layers up to a point increases the performance. Due to this discovery, experiments were focused on number of FC layers and several combinations were tested. According to the results, the network with single FC layer can achieve lowest 6.34 years MAE, with two FC layers can achieve lowest 5.93 years MAE and three FC layers can achieve 5.90 years MAE. This suggests that increased number of FC layers increases the performance. However, as can be seen from results, further increase in number of FC layers decreases the performance of the network. A four FC layer combination displays decreased performance compared to three FC layer. Results show that best performing combination is three FC layers with unit number being 512. Although performance of two FC layers with unit number 4096 performs similar to the best result, the time to complete an epoch is longer. Thus, three FC layers with unit number 512 option was selected to be utilized and the network can be seen in Table 4.2

Table 4.2. Network structure after FC layer evaluation

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	7 * 7	18	1	
Max Pooling	2 * 2		2	
Conv2 + ReLu	5 * 5	45	1	
Max Pooling	3 * 3		2	
FC3 + ReLu (unit number = 512)				
Dropout				
FC4 + ReLu (unit number = 512)				
Dropout				
FC5 + ReLu (unit number = 512)				
Dropout				
FC6 + Linear Function (unit number = 1)				

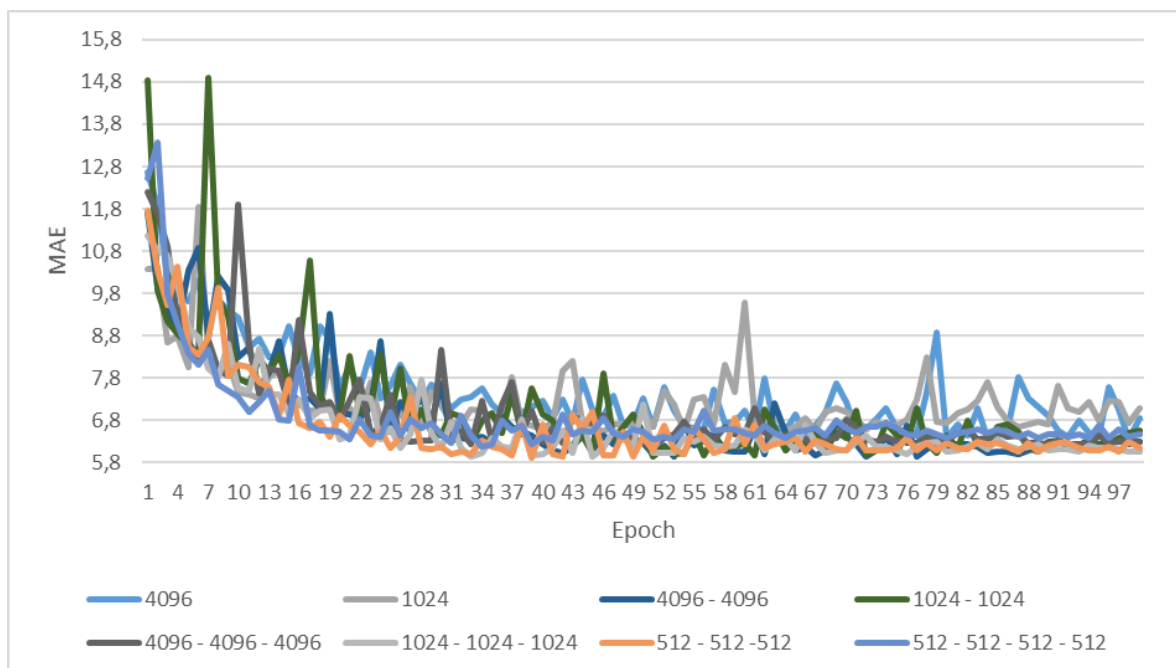


Figure 4.4. Performance comparison of network with different FC layer options

The effects of dropout were evaluated to obtain the final network structure. Results of this evaluation can be seen in Figure 4.5. The figure shows that performance of the network without dropout and with dropout after every FC layer except the last one. In order to observe the effect of dropout on overfitting

problem, training results are included in the figure. Results indicate that the performances of network with and without dropout are similar and the lowest MAE is achieved without using dropout. Furthermore, dropout causes the training performance to approach to the test performance and not the opposite therefore it does not help to overcome the overfitting problem. Thus, the network is designed without the dropout layer and it can be seen in Table 4.3.

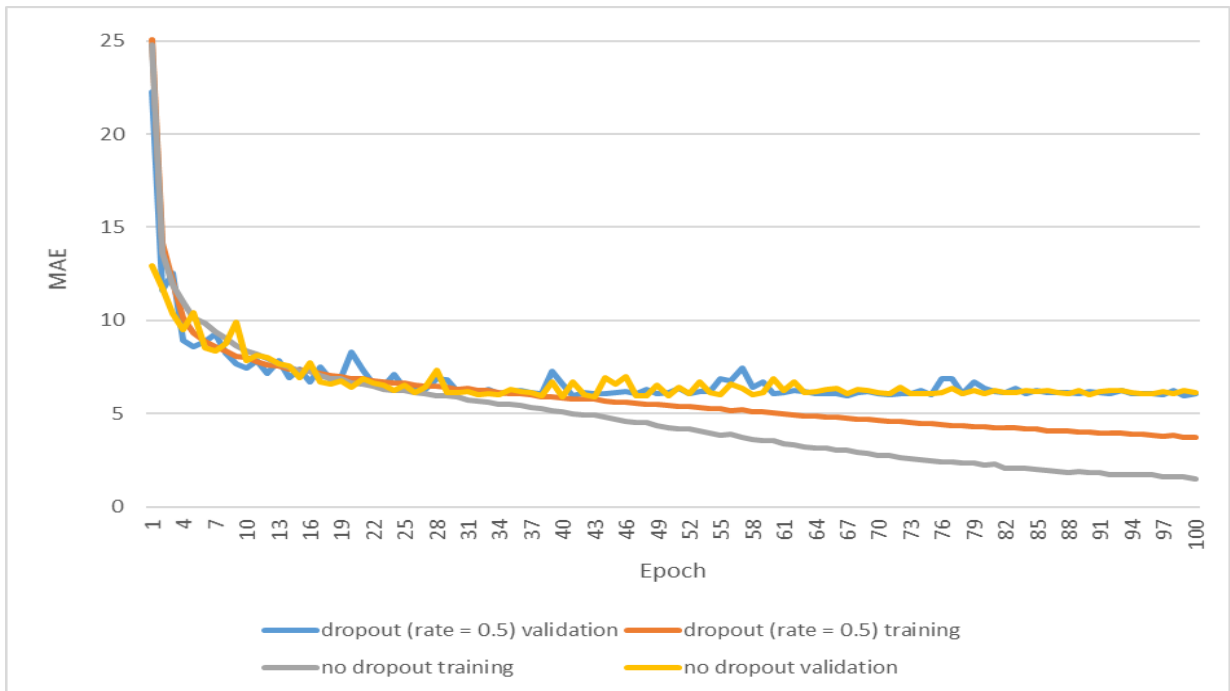


Figure 4.5. Performance comparison of network with dropout and no dropout

Table 4.3. Network structure without dropout

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	7 * 7	18	1	
Max Pooling	2 * 2		2	
Conv2 + ReLu	5 * 5	45	1	
Max Pooling	3 * 3		2	
FC3 + ReLu (unit number = 512)				
FC4 + ReLu (unit number = 512)				
FC5 + ReLu (unit number = 512)				
FC6 + Linear Function (unit number = 1)				

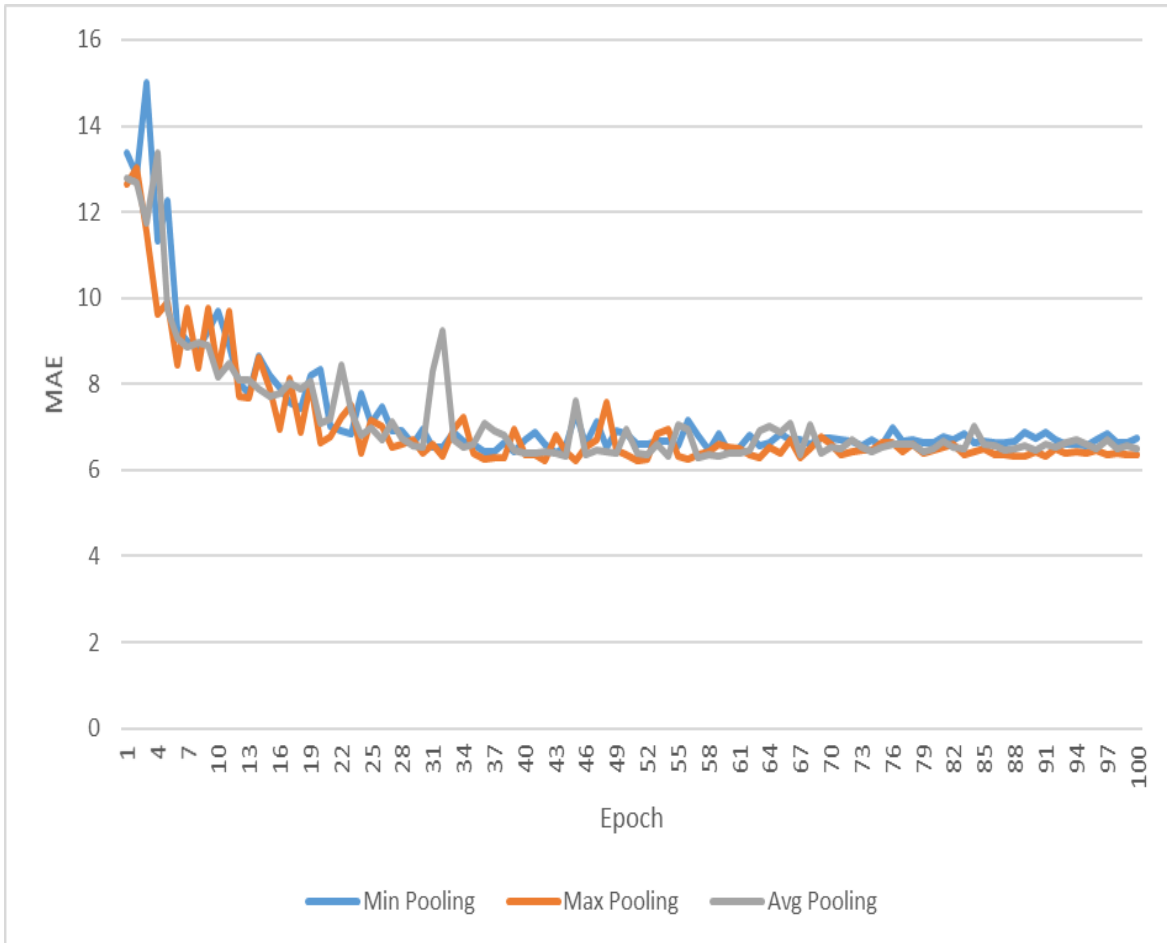


Figure 4.6. Performance comparison of network with different pooling options

Finally, minimum pooling, maximum pooling and average pooling options were evaluated. Figure 4.6 shows performance comparison of the network using these three pooling methods. This suggests that pooling option has no great impact on the outcome of the network and maximum pooling was selected as pooling method.

4.3 Hyperparameter Optimization

After deciding on network structure, several hyperparameters were evaluated. One of the most influencing hyperparameter is the choice of optimization algorithm. For this purpose, age estimation was carried out with different optimization algorithms. These are SDG, RMSProp, Adadelata, Adamax and Adam. This comparison can be seen in Figure 4.7. The results indicate clear performance difference between the optimization algorithms and Adamax performs the best. Computational cost of the algorithms is the same therefore Adamax was chosen as the optimization algorithm.

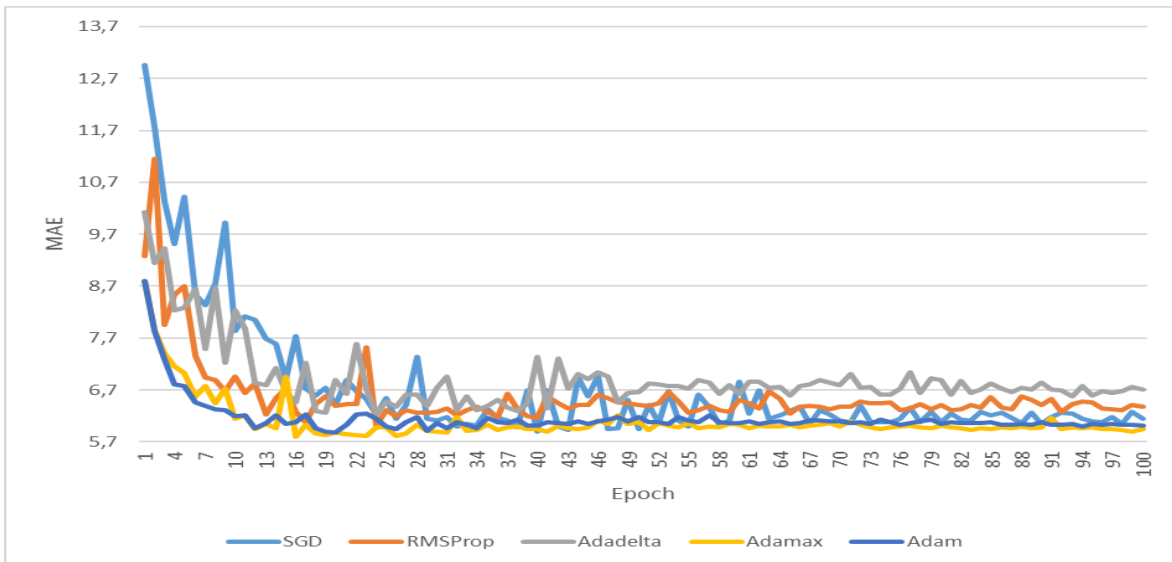


Figure 4.7. Performance comparison of network with different optimization algorithms

Another important hyperparameter is the learning rate set for the optimization algorithm. Different learning rate values were tested using Adamax optimization algorithm and the comparison can be seen in Figure 4.8. The results show that best performance is achieved with learning rate 0.00075. While lower learning rate than this value clearly decreases the performance, as the learning rate increases, performance of the network starts to decline gradually. Considering this, the best performing learning rate which is 0.00075 was set.

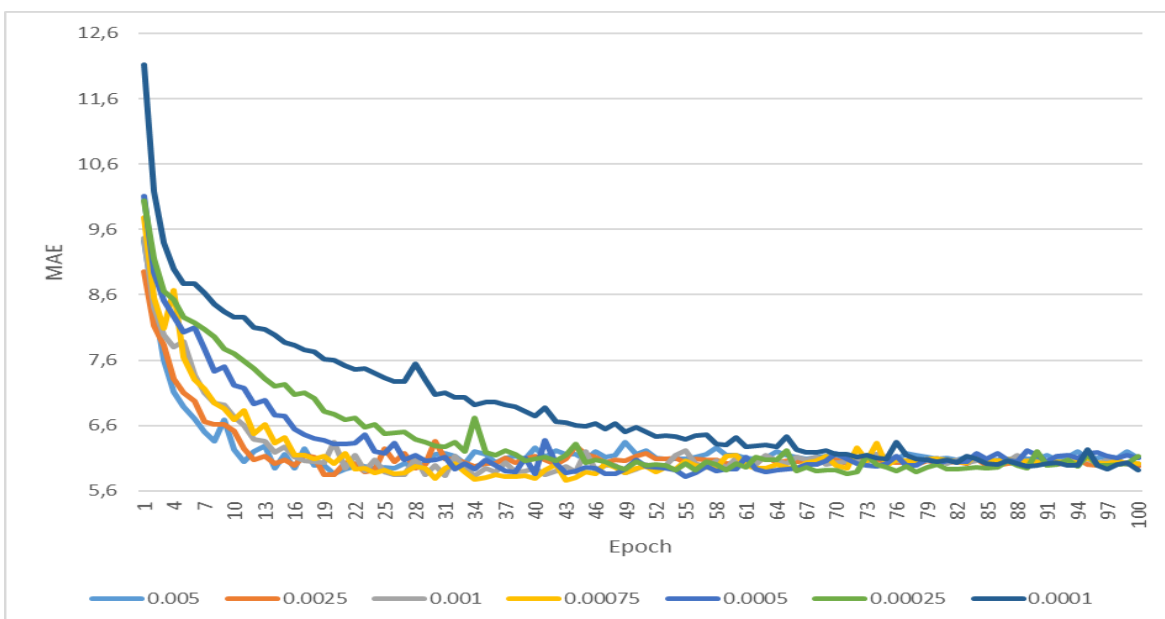


Figure 4.8. Performance comparison of network with different learning rate values

Convolutional layer has several hyperparameters that require fine tuning. The hyperparameters that were tested are stride, kernel size and kernel number. Different values of these hyperparameters were evaluated and results can be found in Figure 4.9, 4.10 and 4.11 respectively. Both convolutional layers were modified to have stride of one, two and three. The results show that as the stride value increases, the performance of the network decreases therefore stride of one was chosen which is the default convolution value. Convolutional layers were changed to have different combinations of kernel size. The first value in series label represents the kernel size of first convolutional layer while the second value represents the kernel size of second convolutional layer. The results indicate that networks with these combinations have analogous outcome and the hyperparameter does not have high impact on the network. However, the network with kernel size of 5 in the first layer and kernel size of 11 in the second layer performs the best therefore this combination was selected. Performance comparison of kernel numbers shows that this hyperparameter does not have great effect on overall performance and different options have similar outcomes. However, as the kernel number increases, computational cost increases as well. Thus, kernel number in first layer was chosen as 18 and in second layer 45. The final network structure can be seen in Table 4.4.

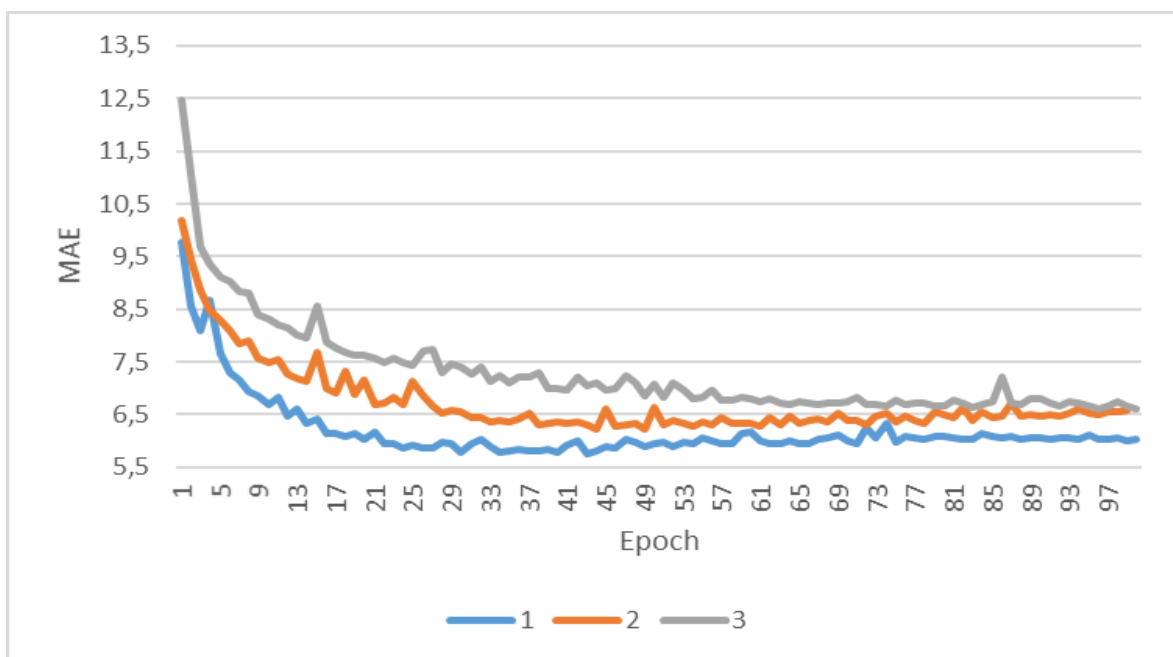


Figure 4.9. Performance comparison of network with stride values in convolutional layers

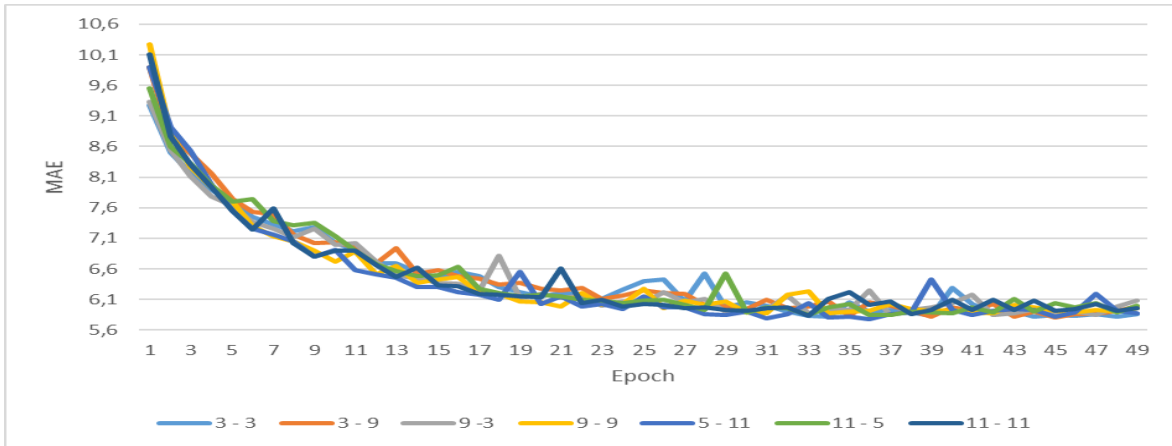


Figure 4.10. Performance comparison of network with kernel size in convolutional layers

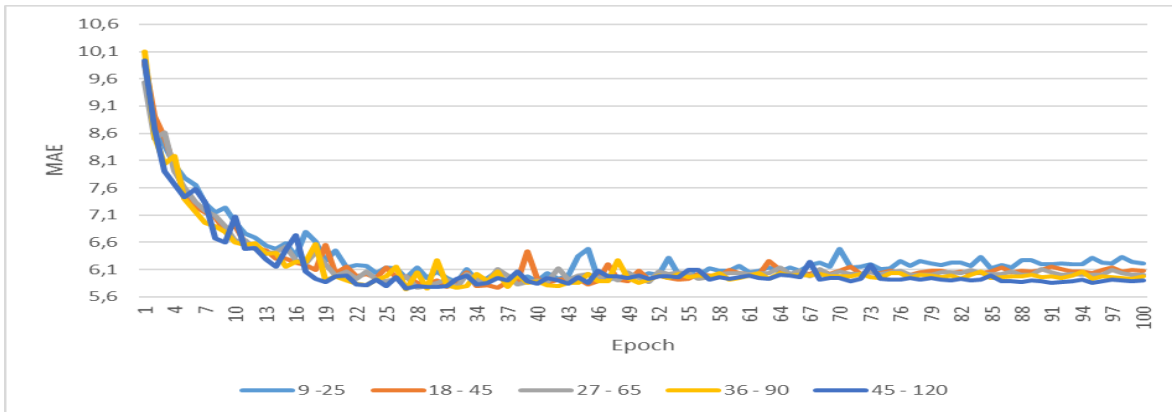


Figure 4.11. Performance comparison of network with kernel number in convolutional layers

Table 4.4. Final network structure

Layer	Filter	Depth	Stride	Padding
Input Image				
Conv1 + ReLu	5 * 5	18	1	
Max Pooling	2 * 2		2	
Conv2 + ReLu	11 * 11	45	1	
Max Pooling	3 * 3		2	
FC3 + ReLu (unit number = 512)				
FC4 + ReLu (unit number = 512)				
FC5 + ReLu (unit number = 512)				
FC6 + Linear Function (unit number = 1)				

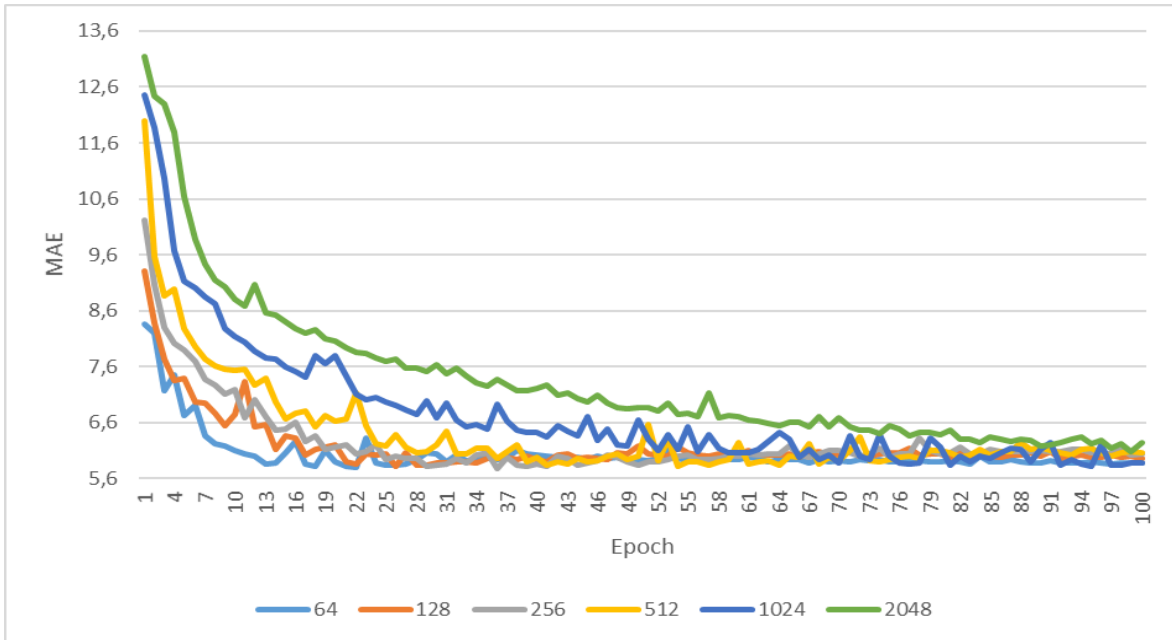


Figure 4.12. Performance comparison of network with different batch sizes

Additionally, the effect of batch size was evaluated. The results can be seen in Figure 4.12. This indicates that as the batch size increases the performance decreases. However, smaller the batch size gets, time to complete an epoch increases. Thus, batch size was selected as 256 because its performance is comparable with lower batch size and has lower computational cost. Furthermore, the lowest MAE was achieved with this batch size.

Finally, effects of L1 and L2 regularization methods on this network were evaluated. Figure 4.13 and Figure 4.14 show effects of L2 and L1 regularization with different rates respectively. This suggests that L2 regularization does not have great impact on overfitting problem. However, utilizing L2 regularization with rate 0.0001 provides slight improvement on network performance. On the other hand, although L1 regularization with rate 0.01 and 0.001 provides a training performance closer to test performance, the test performance does not increase further. In order to verify this, the network was tested with more epochs and rate 0.01. The results of this experiment can be found in Figure 4.15 and it proves that L1 regularization does not contribute to the network. Considering all these, it was decided to employ L2 regularization with rate 0.0001 for slight performance gain.

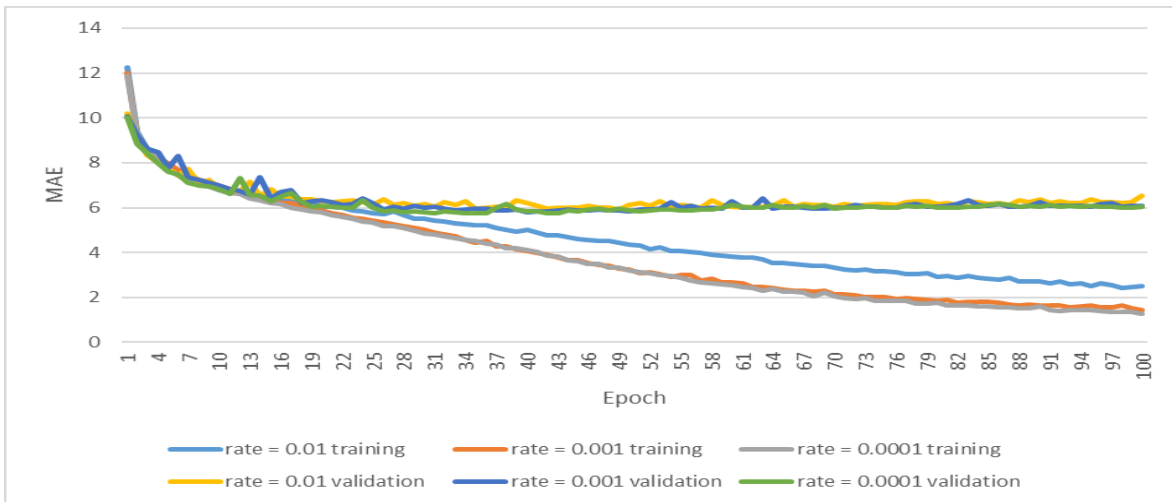


Figure 4.13. Performance of network with L2 regularization

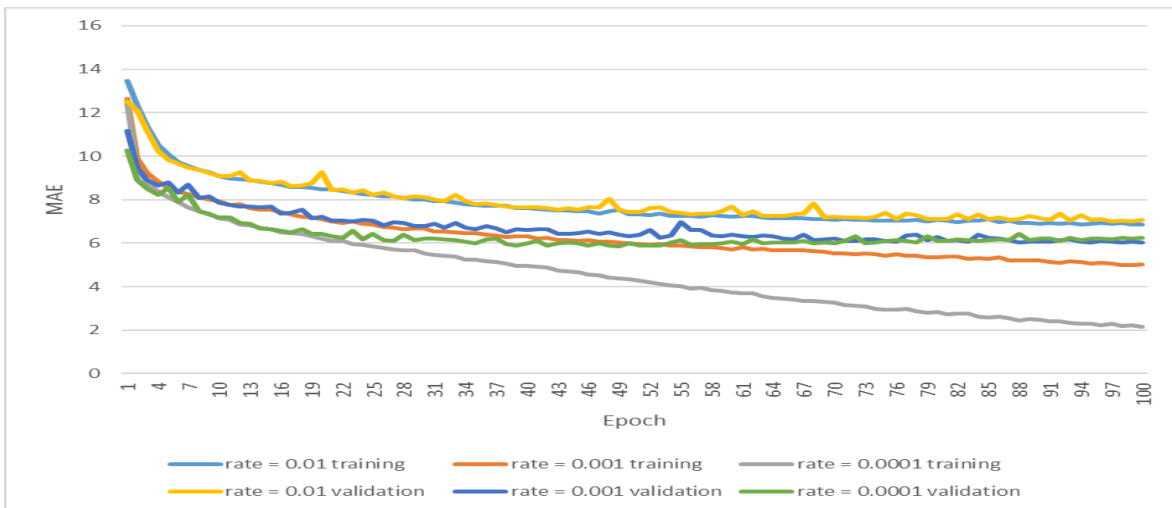


Figure 4.14. Performance of network with L1 regularization

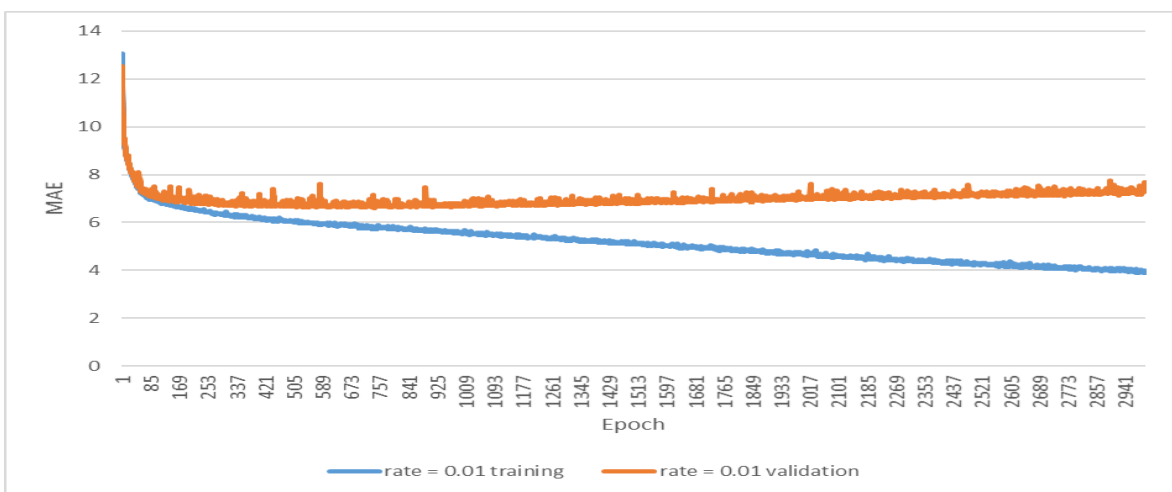


Figure 4.15. Performance of network with L1 regularization and more epochs

5 RESULTS AND DISCUSSION

5.1 Gender Recognition Results

The experiments were conducted on OULP-Age data set which provides balanced distribution of training and test data in terms of gender. Figure 5.1 shows the training and test accuracy obtained with our modified architecture. The highest training accuracy achieved is 99.55% while the highest test accuracy achieved is 99.45%. The results indicate a comparable performance with the existing studies in the literature. This comparison can be seen in Table 5.1. The proposed CNN architecture in this study is able to perform similar to the method proposed in [62] and better than the rest. This indicates that utilizing CNN is feasible in terms of accuracy. An important case to point out in this comparison is that only study [30] employs CNN. This study utilizes a VGGNet-16 [83] based architecture by replacing the last activation function with an SVM. This comparison suggests that an AlexNet based network may be capable of classifying better than a VGGNet based network in gait based gender classification. However, in this study testing data included conditions such as multiple views and carrying object. In gait based recognition studies it is known that such conditions affect performance of methods. Therefore, this accuracy difference cannot prove that VGGNet is not feasible for this problem domain.

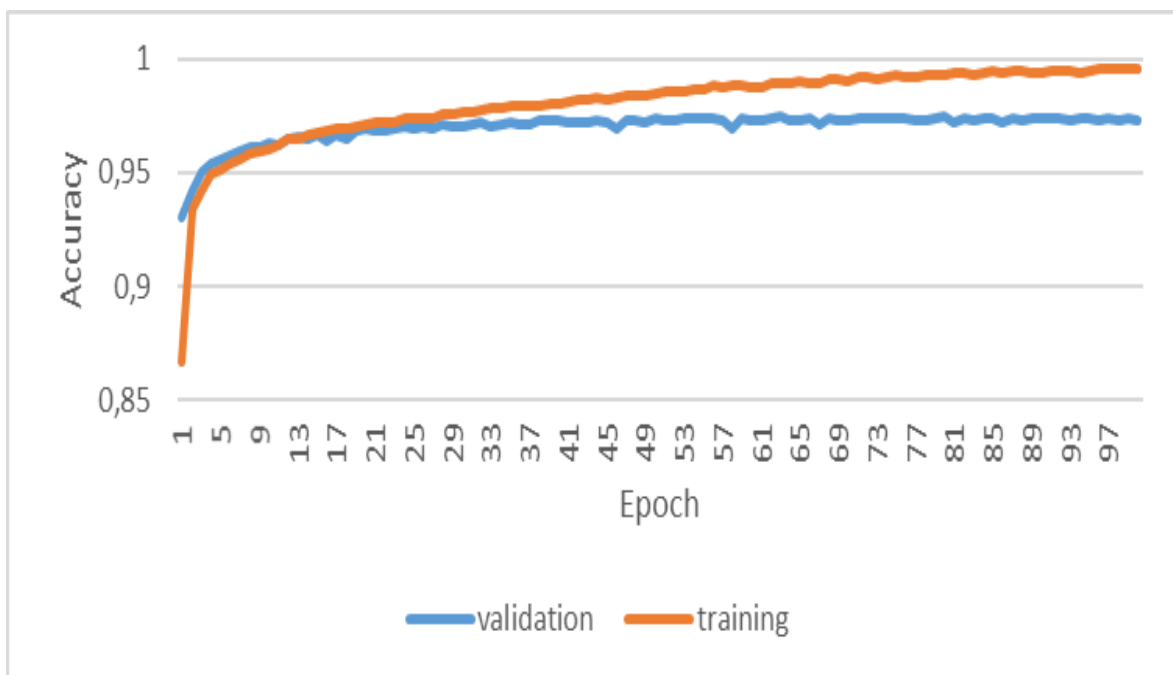


Figure 5.1. Performance of proposed CNN architecture for gender recognition

Table 5.1. Accuracy of different gait based gender recognition methods

Method	Accuracy (%)	Dataset Content
[62]	97.98	124 subjects (93 males – 31 females)
[57]	94.7	124 subjects (93 males – 31 females)
[30]	89.62	62 subjects (31 males – 31 females)
Proposed method	97.45	63,846 subjects (31,093 males – 32,753 females)

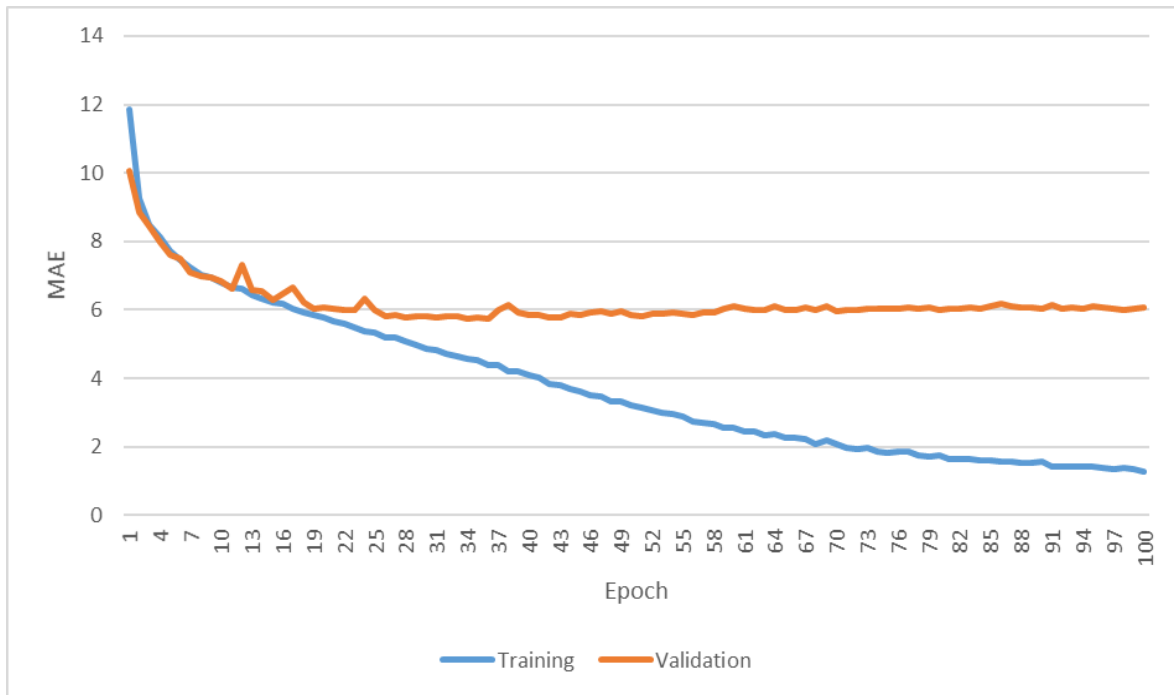


Figure 5.2. Performance of proposed CNN architecture for age estimation

5.2 Age Estimation Results

Several architectural options were evaluated and hyperparameters were tuned for gait based age estimation problem. The results of the final architecture can be found in Figure 5.2.

Performance comparison of proposed architecture and existing studies in the

literature can be found in Table 5.2. There are some important points to note in this comparison. Methods proposed in studies [74] and [77] are performed on a dataset which has 79 subjects with age of these subjects ranging between 15 and 39 years. However, the dataset used in this study consists of more than 60,000 subjects and age ranges from 2 to 90 years. Although these methods perform better than the proposed method in this study, they are tested in a dataset that is smaller and has narrower age range. On the other hand the proposed method in [76] was evaluated with the same dataset used in this study therefore this comparison provides a better insight about the performance of the proposed method. Considering these the proposed method in this study shows a comparable performance. In the literature, there were no attempts to estimate age using CNN with GEI being the input. These results show that using CNN is a viable option to solve the gait based age estimation problem.

The final results were evaluated in terms of actual age of the subjects against the predicted age values in order to assess how the proposed architecture performs on different age groups. The performance of the architecture can be seen in Figure 5.3. This suggests that as the age increases the prediction capability of the network gets worse which can be explained by the age distribution of the dataset. The number of subjects in the dataset decreases as the age increases which causes the network to output off prediction values.

Table 5.2. MAE of different gait based age estimation methods

Method	MAE	Dataset Content
[74]	4.28	79 subjects with age range [15-39]
[77]	4.35	79 subjects with age range [15-39]
[73]	6.76	4096 subjects with age range [1-94]
[76]	6.78	63,846 subjects with age range [2-90]
[75]	8.20	1728 subjects with age range [1-94]
Proposed method	5.74	63,846 subjects with age range [2-90]

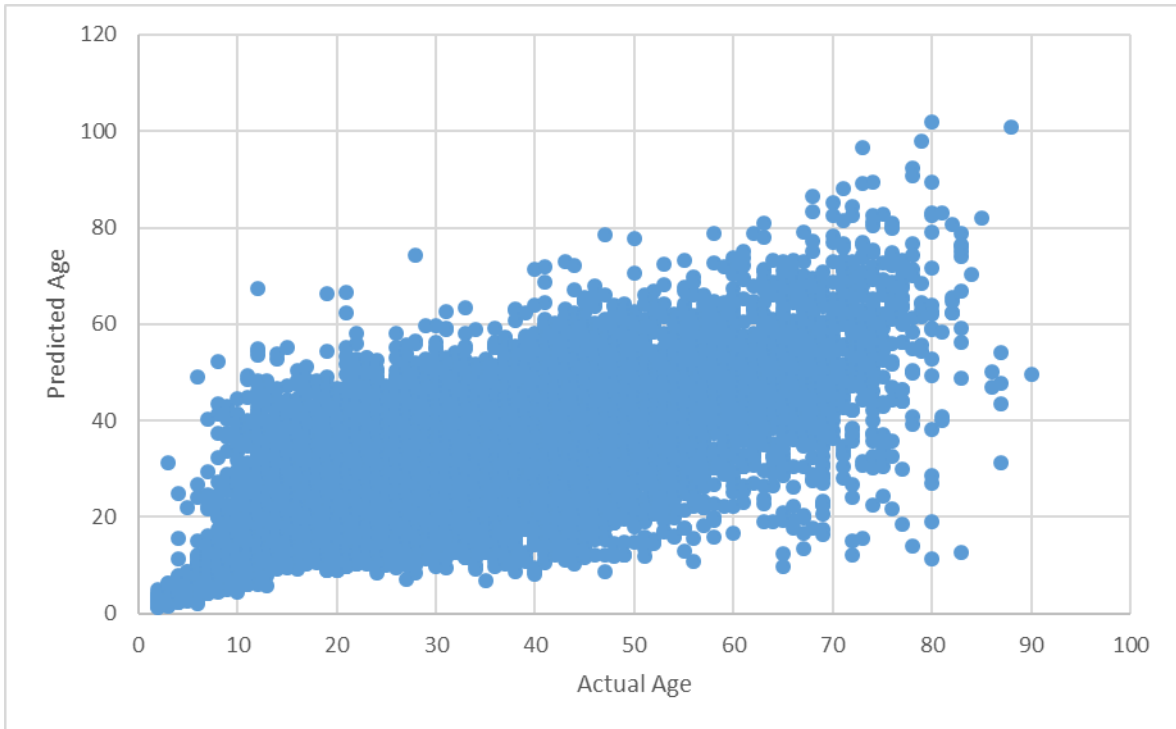


Figure 5.3. Actual age against predicted age

5.3 Discussion

The results obtained to attain the architecture for gender recognition problem can give important insight about the problem and how CNNs can relate to it. The results indicate that it is not necessary to have a deep network. A deep network is generally capable of capturing high level and more complex features [22]. The first layer generally outputs low level features such as edges. The proposed architecture consists of a single convolutional layer because adding more layers does not contribute to the goal. This suggests that gender classification made on GEI can be achieved by simple features such as edges. This finding is anticipatable because GEI is a feature descriptor which means that available information is already processed to be represented in a form that is easier to interpret. Without GEI a deeper network may be required because background noise should be omitted by the network and important temporal gait aspects such as stride frequency should be considered. These are already sorted with GEI allowing us to utilize a smaller network for gender recognition.

Another result that can be inferred is the effect of normalization. Our results

indicate that introducing normalization does not contribute to the goal but adds to the computational overhead. Normalization is considered to be useful to handle sparse data by scaling the input and aids generalization. Datasets that have variance may benefit from normalization. However, in this problem the dataset was constructed in a controlled environment and a preprocessing was applied on the raw data. Images in the dataset were taken from single view and the dataset consists of single channel images. In addition, the dataset is balanced in terms of gender and age. Considering all these, introducing a normalization layer only adds to the computational overhead in the present case.

A better result could be obtained for age estimation problem by having a larger network than the one proposed for gender recognition problem. Additional convolutional layers help to increase performance. This suggests that more complex features are required to have a successful outcome compared to gender recognition problem. Moreover, this indicates that age estimation problem is more challenging compared to gender recognition. However, it is found that a deeper network such as AlexNet does not improve the performance further. This again can be explained by usage of GEI as input. Furthermore, it is observed that additional FC layers contribute to the performance more than other hyperparameters. Two FC layers reduce MAE by 6.54% and three FC layers reduce MAE by 6.88% compared to single FC layered network. This suggests that a CNN architecture with deeper FC layers has higher capability of learning from the extracted feature set.

Another important observation that can be deduced from age estimation results is that the proposed network suffers overfitting. There are several ways to tackle overfitting problem. Using regularization methods and increasing size of dataset are some of these [84]. In order to regulate the network, dropout method was used. However, experimental results show that dropout does not help to reduce overfitting. Additionally, L1 and L2 regularization methods were applied to perform weight decay and to decrease overfitting. Although L2 regularization helps to improve outcome of the network, the results show that these methods do not aid on overfitting problem greatly. These findings suggest that more data is required for gait based age estimation with CNN. Increasing the dataset size or applying

data augmentation may help to overcome overfitting problem and therefore increase the performance of the network.

6 CONCLUSION

A model free approach was utilized to tackle the problem of gait based gender recognition and age estimation problem. In this approach, GEI which is a gait descriptor was used as the data and CNN was adopted as the method to learn the attributes of the problem.

Different CNN architectures were tested and GEINet showed promising results. Based on this architecture, several experiments were conducted to evaluate effects of different architectural designs and hyperparameters in gait based gender recognition and age estimation domain. OULP-Age data set which is the largest gait data set currently available and balanced in terms of gender and age distribution was used for experiments. The data set consists of GEI therefore in this study we evaluated possibility of using CNN for gender recognition and age estimation with GEI being the input.

Results were evaluated in terms of accuracy for gender recognition and of MAE for age estimation; however computational cost was regarded as well in case comparison of these metrics do not yield considerable difference. The final network for gender recognition is able to classify the gender with 97.45% accuracy and an epoch takes approximately 7 seconds with the available resources. The proposed network for age estimation is able to estimate age with 5.74 MAE and time to complete and epoch is approximately 15 seconds. Considering both these outputs, proposed architectures provides improved performance and decreased computational cost compared to the initial architectures they are based on.

The gender recognition and age estimation problems are understudied with CNNs and therefore different CNN architectures and hyperparameters were tested out in order to find an optimal architecture for both problems. Results obtained during this process can provide intuition on how layer and hyperparameter choices such as optimization algorithms can affect the performance of the model in these problem domains. On top of this, experiments yield an understanding on using GEI for this purpose. Not only does this study provide guidance on constructing networks for these problems, it also provides CNN architectures that have

performance comparable to existing studies.

In the future, attempts to improve the performance of networks for both problems can be made. As [72] suggests, there may be a correlation between age and gender therefore both networks can benefit from both traits and may yield better performance.

REFERENCES

- [1] JAIN, Anil K., RUUD, Bolle, and SHARATH, Pankanti, *Biometrics : personal identification in networked society*, Kluwer Academic Publishers, p.4, 1998.
- [2] BOULGOURIS, Nikolaos V., HATZINAKOS, Dimitrios, and PLATANIOTIS, Konstantinos N., *Gait recognition: a challenging signal processing technology for biometric identification*, *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 78–90, 2005.
- [3] KOZLOWSKI, Lynn T. and CUTTING, James E., *Recognizing the sex of a walker from a dynamic point-light display*, *Perception & Psychophysics*, vol. 21, no. 6, pp. 575–580, 1977.
- [4] GRABINER, Penny C., BISWAS, S. Tin., and GRABINER, Mark D., *Age-related changes in spatial and temporal gait variables*, *Archives of Physical Medicine and Rehabilitation*, vol. 82, no. 1, pp. 31–35, 2001.
- [5] ELBLE, Rodger, THOMAS, Susan S., HIGGINS, Christopher, and COLLIVER, Jerry, *Stride-dependent changes in gait of older people*, *Journal of Neurology*, vol. 238, no. 1, pp. 1–5, 1991.
- [6] NIXON, Mark S. and CARTER, John N., *Automatic Recognition by Gait*, *Proceedings of the IEEE*, vol. 94, no. 11, pp. 2013–2024, 2006.
- [7] WANG, J., SHE, M., NAHAVANDI, S., and KOUZANI, A., *A Review of Vision-Based Gait Recognition Methods for Human Identification*, 2010 International Conference on Digital Image Computing: Techniques and Applications, Sydney, NSW, Australia, pp. 320–327, 2010.
- [8] WAGG, D.K. and NIXON, M., *On automated model-based extraction and analysis of gait*, Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, South Korea, pp. 11–16, 2004.
- [9] DUPUIS, Yohan, SAVATIER, Xavier, and VASSEUR, Pascal, *Feature subset selection applied to model-free gait recognition*, *Image and Vision Computing*, vol. 31, no. 8, pp. 580–591, 2013.
- [10] RIDA, Imad, JIANG, Xudong, and MARCIALIS, Gian L., *Human Body Part Selection by Group Lasso of Motion for Model-Free Gait Recognition*, *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 154–158, 2016.
- [11] LIU, Y., COLLINS, R., and TSIN, Y., *Gait Sequence Analysis Using Frieze Patterns*, European Conference on Computer Vision, Copenhagen, Denmark, pp. 657–671, 2002.

- [12] DECANN, B. and ROSS, A., Gait curves for human recognition, backpack detection, and silhouette correction in a nighttime environment, SPIE Conference on Biometric Technology for Human Identification, Orlando, USA, p. 76670Q, 2010.
- [13] HAN, Ju and BHANU, Bir, Individual recognition using gait energy image, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 2, pp. 316–322, 2006.
- [14] CHEUNG, Ada S., GRAY, Hans, SCHACHE, Anthony G., HOERMANN, Rudolf, LIM JOON, Daryl, ZAJAC, Jeffrey D., PANDY, Marcus G., and GROSSMANN, Mathis, Androgen deprivation causes selective deficits in the biomechanical leg muscle function of men during walking: a prospective case-control study, Journal of Cachexia, Sarcopenia and Muscle, vol. 8, no. 1, pp. 102–112, 2017.
- [15] SARKAR, Sudeep, PHILLIPS, P. Jonathon., LIU, Zongyi, VEGA, Isidro R., GROTHOR, Patrick, and BOWYER, Kevin W., The humanID gait challenge problem: data sets, performance, and analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 2, pp. 162–177, 2005.
- [16] SHIQI, Y, DAOLIANG, T., and TIENIU, T., A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition, 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, pp. 441–444, 2006.
- [17] SHUTLER, J., GRANT, M., NIXON, M.S., and CARTER, J.N., On a Large Sequence-Based Human Gait Database. Applications and Science in Soft Computing, LOFTI, Ahamad and GARIBALDI, Johnatan M. (eds), Springer, Berlin, pp. 339–346, 2004.
- [18] HOFMANN, Martin, GEIGER, Jürgen, BACHMANN, Sebastian, SCHULLER, Björn, and RIGOLL, Gerhard, The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits, Journal of Visual Communication and Image Representation, vol. 25, no. 1, pp. 195–206, 2014.
- [19] XU, Chi, MAKIHARA, Yasushi, OGI, Gakuto, LI, Xiang, YAGI, Yasushi, and LU, Jianfeng, The OU-ISIR Gait Database comprising the Large Population Dataset with Age and performance evaluation of age estimation, IPSJ

- Transactions on Computer Vision and Applications, vol. 9, no. 1, p. 24, 2017.
- [20] MAKIHARA, Y. and YAGI, Y., Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation, 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, pp. 1–4, 2008.
- [21] LECUN, Yann, HAFFNER, Patrick, BOTTOU, Léon, and BENGIO, Yoshua, Object Recognition with Gradient-Based Learning, Lecture Notes in Computer Science, vol. 1681, pp. 319–345, 1999.
- [22] GU, Jiuxiang, WANG, Zhenhua, KUEN, Jason, MA, Lianyang, SHAHROUDY, Amir, SHUAI, Bing, LIU, Ting, WANG, Xingxing, WANG, Gang, CAI, Jianfei, and CHEN, Tsuhan, Recent advances in convolutional neural networks, Pattern Recognition, vol. 77, pp. 354–377, 2018.
- [23] SHIRAGA, K., MAKIHARA, Y., MURAMATSU, D., ECHIGO, T., and YAGI, Y., GEINet: View-invariant gait recognition using a convolutional neural network, 2016 International Conference on Biometrics (ICB), Halmstad, Sweden, pp. 1–8, 2016.
- [24] KRIZHEVSKY, A., SUTSKEVER, I., and HINTON, G.E., ImageNet Classification with Deep Convolutional Neural Networks, International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, USA, pp. 1097–1105, 2012.
- [25] NAIR, V. and HINTON, G.E., Rectified linear units improve restricted boltzmann machines, International Conference on Machine Learning, Haifa, Israel, pp. 807–814, 2010.
- [26] HINTON, G.E., SRIVASTAVA, N., KRIZHEVSKY, A., SUTSKEVER, I., and SALAKHUTDINOV, R.R., Improving neural networks by preventing co-adaptation of feature detectors, <https://arxiv.org/abs/1207.0580>, 16 June 2019.
- [27] NG, A.Y., Feature selection, L1 vs. L2 regularization, and rotational invariance, Twenty-first international conference on Machine learning - ICML '04, New York, New York, USA, p. 78, 2004.
- [28] BOTTOU, L. and BOUSQUET, O., The Tradeoffs of Large Scale Learning, 20th International Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, pp. 161–168, 2007.
- [29] MARIN-JIMENEZ, M.J., CASTRO, F.M., GUIL, N., DE LA TORRE, F., and

- MEDINA-CARNICER, R., Deep multi-task learning for gait-based biometrics, 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, pp. 106–110, 2017.
- [30] LIU, T., YE, X., and SUN, B., Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition, 2018 Chinese Automation Congress (CAC), Xi'an, China, pp. 3477–3481, 2018.
- [31] HU, M., WANG, Y., ZHANG, Z., and WANG, Y., Combining Spatial and Temporal Information for Gait Based Gender Classification, 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, pp. 3679–3682, 2010.
- [32] MAKIHARA, Y., MANNAMI, H., and YAGI, Y., Gait Analysis of Gender and Age Using a Large-Scale Multi-view Gait Database, 10th Asian Conference on Computer Vision, Queenstown, New Zealand, pp. 440–451, 2011.
- [33] CHEN, L., WANG, Y., and WANG, Y., Gender Classification Based on Fusion of Weighted Multi-View Gait Component Distance, 2009 Chinese Conference on Pattern Recognition, Nanjing, China, pp. 1–5, 2009.
- [34] LU, Jiwen and TAN, Yap-Peng, Uncorrelated discriminant simplex analysis for view-invariant gait signal computing, *Pattern Recognition Letters*, vol. 31, no. 5, pp. 382–393, 2010.
- [35] CHANG, C. and WU, T., Using gait information for gender recognition, 2010 10th International Conference on Intelligent Systems Design and Applications, Cairo, Egypt, pp. 1388–1393, 2010.
- [36] TROJE, Nikolaus F., Decomposing biological motion: A framework for analysis and synthesis of human gait patterns, *Journal of Vision*, vol. 2, no. 5, p. 2, 2002.
- [37] HUANG, G. and WANG, Y., Gender Classification Based on Fusion of Multi-view Gait Sequences, *Computer Vision – ACCV 2007*, Nanjing, China, pp. 462–471, 2007.
- [38] ZHANG, D. and WANG, Y., Investigating the separability of features from different views for gait based gender classification, 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, pp. 1–4, 2008.
- [39] ZHANG, D. and WANG, Y., Using multiple views for gait-based gender classification, *The 26th Chinese Control and Decision Conference (2014 CCDC)*, Changsha, China, pp. 2194–2197, 2014.

- [40] LU, J., WANG, G., and HUANG, T.S., Gait-based gender classification in unconstrained environments, 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, pp. 3284–3287, 2012.
- [41] LU, Jiwen, WANG, Gang, and MOULIN, Pierre, Human Identity and Gender Recognition From Gait Sequences With Arbitrary Walking Directions, IEEE Transactions on Information Forensics and Security, vol. 9, no. 1, pp. 51–61, 2014.
- [42] ZAKI, Mohamed H. and SAYED, Tarek, Using automated walking gait analysis for the identification of pedestrian attributes, Transportation Research Part C: Emerging Technologies, vol. 48, pp. 16–36, 2014.
- [43] CHANG, P., TIEN, M., WU, J., and HU, C., Real-time Gender Classification from Human Gait for Arbitrary View Angles, 2009 11th IEEE International Symposium on Multimedia, San Diego, USA, pp. 88–95, 2009.
- [44] KALAISELVAN, C. and RAJA, A. Sivanantha, Robust Gait-Based Gender Classification for Video Surveillance Applications, Applied Mathematics & Information Sciences, vol. 11, no. 4, pp. 1207–1215, 2017.
- [45] GUAN, Yu and WEI, Xingjie, On the Generalization Power of Face and Gait Gender Recognition Methods, International Journal of Digital Crime and Forensics, vol. 6, no. 1, pp. 1–8, 2014.
- [46] FÉLEZ, Raúl Martín, GARCÍA, Vicente, and SÁNCHEZ, J. Salvador, Gait-based Gender Classification Considering Resampling and Feature Selection, Journal of Image and Graphics, vol. 1, no. 2, pp. 85–89, 2013.
- [47] BORRÀS, R., LAPEDRIZA, À., and IGUAL, L., Depth Information in Human Gait Analysis: An Experimental Study on Gender Recognition, International Conference Image Analysis and Recognition, Aveiro, Portugal, pp. 98–105, 2012.
- [48] IGUAL, Laura, LAPEDRIZA, Àgata, and BORRÀS, Ricard, Robust gait-based gender classification using depth cameras, EURASIP Journal on Image and Video Processing, vol. 2013, no. 1, p. 1, 2013.
- [49] CAIFENG S., SHAO G., and MCOWAN, P.W., Learning gender from human gaits and faces, 2007 IEEE Conference on Advanced Video and Signal Based Surveillance, London, UK, pp. 505–510, 2007.
- [50] ZHANG, D. and WANG, Y., Gender recognition based on fusion on face and gait information, 2008 International Conference on Machine Learning and

- Cybernetics, Kunming, China, pp. 62–67, 2008.
- [51] LEE, L. and GRIMSON, W.E.L., Gait analysis for recognition and classification, Fifth IEEE International Conference on Automatic Face Gesture Recognition, Washington, DC, USA, pp. 155–162, 2002.
 - [52] DAVIS, J.W. and HUI, G., Gender Recognition from Walking Movements using Adaptive Three-Mode PCA, 2004 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, pp. 9–9, 2004.
 - [53] DECANN, B., ROSS, A., and CULP, M., On Clustering Human Gait Patterns, 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, pp. 1794–1799, 2014.
 - [54] YU, Shiqi, TAN, Tieniu, HUANG, Kaiqi, JIA, Kui, and WU, Xinyu, A Study on Gait-Based Gender Classification, IEEE Transactions on Image Processing, vol. 18, no. 8, pp. 1905–1910, 2009.
 - [55] LI, Xuelong, MAYBANK, Stephen. J., YAN, Shuicheng, TAO, Dacheng, and XU, Dong, Gait Components and Their Application to Gender Recognition, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 2, pp. 145–155, 2008.
 - [56] YOO, J., HWANG, D., and NIXON, M.S., Gender Classification in Human Gait Using Support Vector Machine, 7th international conference on Advanced Concepts for Intelligent Vision Systems, Antwerp, Belgium, pp. 138–145, 2005.
 - [57] FÉLEZ, R.M., MOLLINEDA, R.A., and SANCHEZ, J.S., Towards a More Realistic Appearance-Based Gait Representation for Gender Recognition, 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, pp. 3810–3813, 2010.
 - [58] HU, Maodi, WANG, Yunhong, ZHANG, Zhaoxiang, and ZHANG, De, Gait-Based Gender Classification Using Mixed Conditional Random Field, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 41, no. 5, pp. 1429–1439, 2011.
 - [59] HANDRI, Santoso, NOMURA, Shusaku, and NAKAMURA, Kazuo, Determination of Age and Gender Based on Features of Human Motion Using AdaBoost Algorithms, International Journal of Social Robotics, vol. 3, no. 3, pp. 233–241, 2011.
 - [60] OSKUIE, F.B. and FAEZ, K., Gender Classification Using a Novel Gait

- Template: Radon Transform of Mean Gait Energy Image, 8th International Conference on Image Analysis and Recognition, Burnaby, Canada, pp. 161–169, 2011.
- [61] HU, M. and WANG, Y., A New Approach for Gender Classification Based on Gait Analysis, 2009 Fifth International Conference on Image and Graphics, Xi'an, Shanxi, China, pp. 869–874, 2009.
- [62] HASSAN, S, MOHAMMED, O., ABDULAZEEZ, A.M., and TIRYAKI, V.M., Gait-Based Human Gender Classification Using Lifting 5/3 Wavelet and Principal Component Analysis, 2018 International Conference on Advanced Science and Engineering (ICOASE), Duhok, Iraq, pp. 173–178, 2018.
- [63] LIVNE, Micha, SIGAL, Leonid, TROJE, Nikolaus F., and FLEET, David J., Human attributes from 3D pose tracking, *Computer Vision and Image Understanding*, vol. 116, no. 5, pp. 648–660, 2012.
- [64] LAWSON, W., DURIC, Z, and WECHSLER, H., Gait Analysis using Independent Components of image motion, 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, Amsterdam, Netherlands, pp. 1–6, 2008.
- [65] ARAI, K. and ANDRIE, R., Gender Classification with Human Gait Based on Skeleton Model, 2013 10th International Conference on Information Technology: New Generations, Las Vegas, USA, pp. 113–118, 2013.
- [66] HANDRI, Santoso and NAKAMURA, Kazuo, Gender and Age Classification Based on Pattern of Human Motion Using Choquet Integral Agent Networks, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 13, no. 4, pp. 481–488, 2008.
- [67] MANSOURI, Nabila, ISSA, Mohammed A., and JEMAA, Yousra B., Gait features fusion for efficient automatic age classification, *IET Computer Vision*, vol. 12, no. 1, pp. 69–75, 2018.
- [68] BEGG, Rezaul K., PALANISWAMI, Marimuthu, and OWEN, Brendan, Support Vector Machines for Automated Gait Classification, *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 5, pp. 828–838, 2005.
- [69] ZHANG, D., WANG, Y., and BHANU, B., Age Classification Base on Gait Using HMM, 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, pp. 3834–3837, 2010.
- [70] DAVIS, J.W., Visual Categorization of Children and Adult Walking Styles,

- Third International Conference on Audio- and Video-Based Biometric Person Authentication, Halmstad, Sweden, pp. 295–300, 2001.
- [71] CHUEN, B.K.Y., CONNIE, T., SONG, O.T., and GOH, M., A preliminary study of gait-based age estimation techniques, 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Hong Kong, China, pp. 800–806, 2015.
- [72] LU, Jiwen and TAN, Yap-Peng, Gait-Based Human Age Estimation, IEEE Transactions on Information Forensics and Security, vol. 5, no. 4, pp. 761–770, 2010.
- [73] PUNYANI, Prachi, GUPTA, Rashmi, and KUMAR, Ashwani, Human age-estimation system based on double-level feature fusion of face and gait images, International Journal of Image and Data Fusion, vol. 9, no. 3, pp. 222–236, 2018.
- [74] LU, Jiwen and TAN, Yap-Peng, Ordinary Preserving Manifold Analysis for Human Age and Head Pose Estimation, IEEE Transactions on Human-Machine Systems, vol. 43, no. 2, pp. 249–258, 2013.
- [75] MAKIHARA, Y., OKUMURA, M., IWAMA, H., and YAGI, Y., Gait-based age estimation using a whole-generation gait database, 2011 International Joint Conference on Biometrics (IJCB), Washington, DC, USA, pp. 1–6, 2011.
- [76] LI, Xiang, MAKIHARA, Yasushi, XU, Chi, YAGI, Yasushi, and REN, Mingwu, Gait-based human age estimation using age group-dependent manifold learning and regression, Multimedia Tools and Applications, vol. 77, no. 21, pp. 28333–28354, 2018.
- [77] LU, J. and TAN, Y., Cost-sensitive subspace learning for human age estimation, 2010 IEEE International Conference on Image Processing, pp. 1593–1596, 2010.
- [78] CHOLLET, F., Keras, <https://keras.io/>, 10 May 2019.
- [79] KINGMA, D.P. and BA, J., Adam: A Method for Stochastic Optimization, <https://arxiv.org/abs/1412.6980>, 10 June 2019.
- [80] HINTON, G., SRIVASTA, N., SWERSKY, K., Lecture 6a Overview of mini-batch gradient descent, https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf, 10 June 2019.
- [81] ZEILER, M.D., ADADELTA: An Adaptive Learning Rate Method, <https://arxiv.org/abs/1212.5701>, 10 June 2019.

- [82] GOODFELLOW, Ian, YOSHUA Bengio, and AARON Courville, Deep Learning, 1st ed., MIT Press, p. 302, 2016.
- [83] SIMONYAN, K. and ZISSERMAN, A., Very Deep Convolutional Networks for Large-Scale Image Recognition, <https://arxiv.org/abs/1409.1556>, 15 June 2019.
- [84] NIELSEN, Michael A., Neural Networks and Deep Learning, Determination Press, p.78, 2018.