GENDER CLASSIFICATION THROUGH CONVNET USING GAIT ENERGY IMAGE AND ONE-SHOT LEARNING

A PROJECT REPORT

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY

IN

INFORMATION SYSTEMS

Submitted by:

Anjali Gahalout

(2K19/ISY/02)

Under the supervision of

Dr. Dinesh Kumar Vishwakarma

Professor

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

JULY, 2021

CANDIDATE'S DECLARATION

I Anjali Gahalout, student of M. Tech., Information Systems (2K19/ISY/02), hereby declare that the project report titled **"Gender classification through ConvNet using Gait Energy Image and One-Shot Learning"** submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in order to fulfil the work for the degree of Master of Technology, is original and not copied from any source without proper citation. I or anybody else has not used this work as a requirement for award of any degree, before.

Place: Delhi

July 23, 2021

Anjali

ANJALI GAHALOUT

2K19/ISY/02

CERTIFICATE

I hereby certify that the Project Report titled **"Gender classification through ConvNet using Gait Energy Image and One-Shot Learning"** which is submitted by Anjali Gahalout (2K19/ISY/02), Department of Information Technology, DTU, Delhi in order to fulfil the requirement for the degree of Master in Technology, is a record of the project work done by the student with my guidance. As far as my best knowledge, this project has not been used for any other degree, to any University.

Place: Delhi

July 23, 2020

Dr. DINESH KUMAR VISHWAKARMA

SUPERVISOR

(Professor)

Department of Information Technology

ABSTRACT

Many human attributes like finger print, iris are being prominently used for identification purpose. Acquisition of these features might be problematic if the person is non-cooperative or not available in vicinity. Another problem that might arise is of cameras. The facial features require high power cameras for clarity and sometimes the angle in which camera is positioned might prove problematic. Another human identification feature that has been gaining popularity is gender. In current scenario facial features are most popularly used for the purpose, but this might face the same problems as stated above. In such cases using gait for the purpose of identification looks promising. In this paper we have used One shot learning with Siamese network for classification of gender using Gait Energy Images. The model achieves 99% accuracy using the CASIA-B database.

ACKNOWLEDGEMENT

The completion of this work has been possible only because many people came forward for their guidance and support. I am extremely grateful to them who helped me reach this point with my work and would like to acknowledge their help for the same.

I am extremely grateful to my supervisor, **Dr. Dinesh Kumar Vishwakarma**, for helping me all along from the beginning. He responded keenly to all my queries and helped in resolving my doubts. I heartily thank him for his help and support, in order to complete this project.

I am also thankful enough to get all the help and support from the whole of the IT Department, who encouraged me and provided their help in much needed times. I would also like to thank the staff from the computer labs who helped me get all the required technical resources.

Anjali

ANJALI GAHALOUT

2K19/ISY/02

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

There are various applications of the visual surveillance like in traffic monitoring, parking monitoring system and banks. In such applications, in which human being is the main focus of attention, certain information must be extracted from the video feed in order to classify and analyze the behavior.

Previously, research utilized images of face and iris, handwriting, finger and palm prints for authentication and authorization of individuals. But there are various limitations that degrade the efficiency of these conventional methods.

1. At present, large distance may prove a hindrance in the aforementioned biometrics. Often, the distance between the objects under surveillance and camera is very large. There may be a possibility of camera being mounted on top of the building or in the corridor ceiling hence, making it difficult to capture the conventional features.

2. We often require cooperation of people for capturing the conventional information. For example, while taking the finger print a scanner is needed which must be used properly and with the consent of the person. Furthermore, iris scanning is much stricter as it requires the person to continuously look through the eyepiece.

3. Problems arises due to authentication and authorization as they too require the user's attention. Hence, if the user does not cooperate then there will be difficulty in collecting the database. Also, if any problem occurs in the system, then the whole process of biometrics needs to be repeated.

The utility of the conventional biometrics is difficult for visual surveillance and so, using human gait for biometrics proves as a fascinating alternative. We can define Gait as the walking demeanor of an individual and it can be easily captured without the person's cooperation from a distance. Surgery [1] and psychology [2] were the first fields to use human gait information. In surgery, it was used for finding the abnormalities in the gait of people so that adequate treatment decisions could be made. For suitable treatment Murray[3] classified the pathologically abnormal gaits into various categories. Comparison of the gait patterns of a patient with that of the control group (which consists of normal gait patterns) helped in achieving this classification. Physical attributes of people and surrounding's factors are certain things that affect the human gait, for example: the viewpoint of the camera affects the measurement of the gait, another issue of visual surveillance is time elapse, carrying status and clothes of the walking figure, walking speed and other factors like injury, image quality, lighting, walking surface and background. Several of these factors have been studied by human ID dataset which provided baseline algorithms. However, the issue of measuring the effect of each factor remains unresolved.

There is also possibility of correlation between the factors for example, variation in speed may arise due to different shoe type or walking surface. These correlations can be very significant because different components of gait may cause variations for example, while walking someone's hand may touch his or her leg. Such wide variations are not possible in traditional biometrics for example, the eye movement below the nose is not possible, simply put the corresponding position of eyes and nose may slightly vary. Difficulty in the measurement of Gait arises due to these large variations but because of them we may be able to gather a lot of information from the gait.

Gait can also be used for the purpose of gender classification. If we see the current pandemic situation then, zero contact biological features collection, unlike voice and footprints is very beneficial. Thus, we can say in coming future gait might take over all the biological features for the purpose of identification. Apart from biometric systems, gait can also prove beneficial in surveillance systems [4] . Like in metros, we can easily ensure the safety of women by monitoring male boarders in the female coach.

The current biometric system has a good accuracy rate. Different human features are captured, preprocessed according to the requirements and are then compared with a database of existing features [5]. This database is created by collecting features of different individuals over a period of time, like our adhaar card system. This database is then used to identify whether the captured human features belong to anyone in the database or not. Thus, using this we are able to confirm if a person is an authorized user or not. Since, a large dataset is present for each individual the accuracy of identification is vastly increased.

However, there are shortcomings to the process of biometric identification as well. For example, while acquiring features like iris and finger prints of an individual for the identification purpose there might be a chance of using fake finger prints or iris. Identity theft and muddling are some attacks that can occur in this phase. During the feature extraction phase there may be attacks like Trojan horse attacks. And while enrolling the data problems like fraud or deliberately changing data might occur. Due to these problems the person might be identified incorrectly. While reconstructing the fingerprints and iris, there is a chance of them being corrupted which might lead to incorrect identification of the individual.

Characteristics like height, build play a major role in individualizing the gait of different people. Thus, gait can be easily utilized in classifying the gender as male or female and can also help in distinguishing whether a person is young or old [6] . At the same time, the shortcomings of the traditional biometric methods can be easily overcome by gait. The biggest advantage of gait is zero contact and its non invasive nature. We do not require the subject's cooperation like in the traditional biometric systems for capturing the biometric feature. The capturing of gait from a safe range without requiring the attention of subject is easily possible. Generally, gait is captured in the form of video which is then used for capturing the frames. These frames are then used for analyzing the motion of the humans which then leads to the identification of individual.

The task of identifying an individual with high accuracy is quite challenging. In restricted environments like banks, airports continuous monitoring is required to avoid threats deftly. In such situations the ability of capturing gait from a distance looks promising. The other promising feature of gait is the easy availability of videos from the surveillance camera. Using these videos we can easily detect and identify the culprit of the threat in such environments.

1.1 GAIT BIOMETRICS

The primary focus of motion analysis field is human recognition through individual. This has resulted in substantial research on gait recognition through videos. In the early stages MLD and reflectors were employed on different human body joints by Cutting and Kozlowski$ and Johansson $ who then observed the uniqueness of gait patterns and sought it in biometric human identification. Presently, the research in gait biometrics is prominently focused on vision based data set. The CASIA Dataset B is the extensively used dataset for the analysis of different gait identification models. This is because the dataset has 124 images with 11 view variations which also include carrying and clothing conditions that effect the gait recognition performance. CASIA has categorized data into: Dataset A, Dataset B, Dataset C and Dataset D. Gait biometric can fit in with the current norms of social distancing. In the recent years, most advancement in gait is done in computer vision field for the purpose of identification. However, gait identification can be done using two approaches:

1.1.1 Model based approach

In model based approach, the tracking or modeling of body components is done to derive the gait signatures. Then these signatures are used for the purpose of individual identification and verification as well as can be used for the purpose of action monitoring [7] [8]. This method is principally based on extracting the prior knowledge. A human body structure is fitted on the walking sequence in every single frame of gait cycle in order to achieve the gait components. This method is easily understood and is vein and scale constant. It is also not affected by noise which makes it most suitable for practical applications. There may arise the problem of self occlusion which can be solved using multi camera gait-acquisition system [9]. In [10], body parameters were used that were static in nature and were collected from static gait frames as well. Height, the space betwixt head and abdomen, abdomen and feet as well as the maximum space betwixt both the feet are used as static parameters. The view invariance of these parameters proves to be beneficial for the purpose of recognition. In [11], silhouette image of the person is divided into 7 segments which were used for calculating the feature vectors of the regions in which ellipses were fit. These were namely, aspect ratio, centroid average and orientation of major axis of ellipse. Yoo et al. [12] proposed a neural network with back propagation mechanism for the automatic gait recognition process. The method consists of extracting 9 body points from silhouette image, which were then used for forming 2D stick images. Further, these stick figures were used for extracting 10 features which were used for training the network for individual recognition.

1.1.2 Model free approach

In this, we do not require any prior knowledge about different body features as it is solely focused on the shape or behavior of the human silhouette. In this approach, we directly extract the features from the gait silhouette binary part as it is insusceptible to color and quality. Also, it proves easy to track the abnormal human activities like in monitoring the old people [13] [14]. Further, it can prove beneficial for the surveillance purpose for identifying suspicious activities of subjects [15] [16]. The model-free approach is comparatively better than the model based approach in terms of computational cost and also the silhouette quality does not affect the performance of gait recognition methods. However, this method is susceptible to variations in clothing and carrying conditions. For subjugating this problem, Worapan Kusakunniran [17] proposed a method in which the features that prominently identify the gait are directly extracted from the unprocessed videos. A gait model was suggested in [18] which was used for extracting the cadence and stride length from the silhouette images in order to identify an individual. We also need to remove the background for the formation of silhouette images in the preprocessing steps which may lead to time ramification and may also require massive storage space. As a solution to this problem, Gait Energy Image (GEI) which is a spatio-temopral silhouette image representation was proposed by Han and Bhanu [19]. In this, the motion information is represented through a single image containing the temporal information, and is represented by:

$$G(i,j) = \frac{1}{N}\sum_{t=1}^{N} I(i,j,t) \tag{1.1}$$

Here (1.1), N represents the gait cycle frames, t is the frame number at a particular time instance and $I(i,j)$ represents the original (i,j) two-dimensional silhouette images. Fig.1 shows a typical GEI image. The GEI images are computational time and storage space friendly and are also less sensitive to noise.

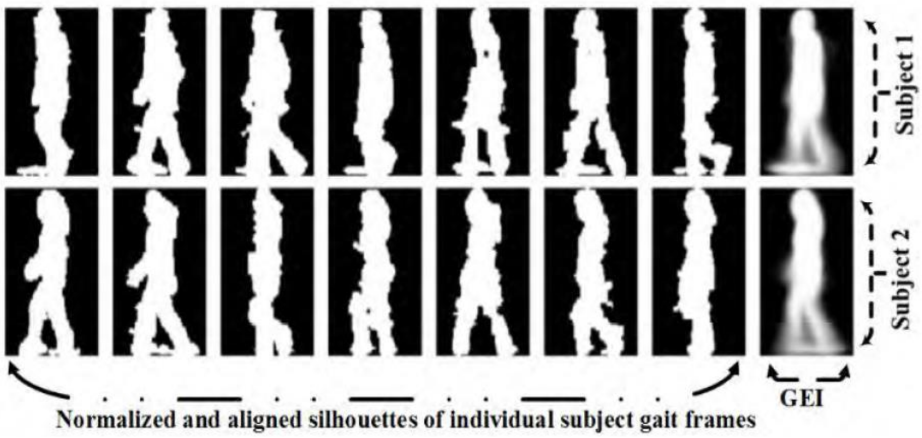


Fig 1.1 Gait Energy Image (GEI) of subjects taken from [19]

1.2 FRAMEWORK FOR GAIT RECOGNITION

Recognition of individuals using gait has become quite a popular research area and has garnered the attention of many researchers. It has various applications like gender classification, age estimation, video surveillance, abnormal action recognition etc. Gait recognition is done using patter recognition, in which the gait of an individual is matched with various available gait cycles. The difference present between them is due to the feature vectors that help in identification of the gait of an individual. The figure below shows the steps in a general framework of recognition system. This framework consists of 2 phases: training and testing. In training, first we need to extract the gait frames of an individual from the videos and then perform pre-processing on them to remove unwanted elements like noise and also perform segmentation to extract the desired object from image. Then the feature vectors of these objects are calculated and provided to model for training. In testing, we pass the gait frames of the test subject through the pre-processing step and its feature vectors are extracted. These feature vectors are then compared with those of the training images and desired result is generated.

a) Gait acquisition: The first step of this system is capturing of the gait frames of the subjects. This can be done using sensor-based or video-based devices. There are two types of sensors: floor sensors and wearable sensors. The floor sensors provide a pressure signal when a subject walk on them, whereas the wearable sensors are attached to the various body joints to capture the dynamic traits like speed, position that can be utilized for gait analysis. The video-based devices capture the gait of the subject using the cameras that can be easily installed at any site. These videos are then used for acquiring gait frames for gathering the gait pattern details that help in identifying the gait of an individual.

b) Pre-processing: This step involves the removal of background elements to obtain the required foreground objects. We can easily represent these methods of background subtraction as:

$$I_f = |I_o - I_b| \tag{1.2}$$

Where (1.2), I_o is the original image and I_f is the foreground image obtained by subtracting the background image, I_b from the original image. Some of the background subtraction methods are erosion, filtering, optical flow, median method and many others.

c) Feature Extraction: Once the required object has ben obtained through background subtraction, feature extraction can be used for obtaining the feature set of such objects.

d) Feature Selection: The traditional methods use entire extracted feature set obtained in the step of pre-processing for the classification purpose as a result the performance deteriorates. This happens because these high dimension features contain some unnecessary features as well. Thus, we need to select proper features from this superset that help in uniquely identifying the gait and also reduces the effect of noise and as a result provide better results.

e) Classification: The final step of gait recognition system is classification of the test images based on the extracted feature sets.The below table shows the most commonly used classifiers for gait recognition:

TABLE 1.1 Different classifiers and their benefits

Classifiers	**Benefits**
kNN	Computation is easy and efficient, when dataset is large
Naive Bayes	Very easy to implement as it requires less training dataset and gives probabilistic prediction
SVM	It provides sparse solution by using kernels in if the local minima is absent
DCNN	It obtains high accuracy but requires GPU and large dataset for the purpose of training

The figure below shows the above discussed gait framework:

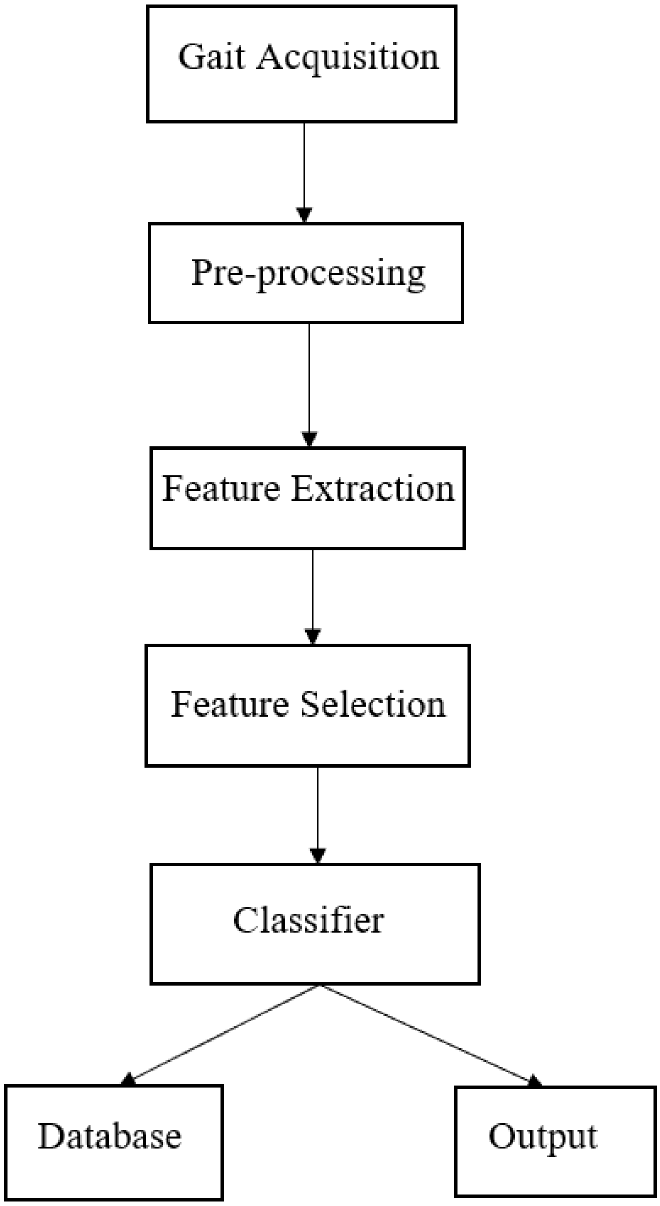


Fig 1.2 A general framework for gait recognition

1.3 GAIT DATASETS

Fundamental development and success of any new application lies solely on the database. A database of variant factors and sufficient size is required to develop and test a robust gait recognition model. Many datasets pertaining to gait recognition have been created and these datasets take into account various factors such as type of clothes the subject wears, whether the subject has additional baggage or not, shadow, type of shoes, view variations, acquisition environment etc. Table 1 shows the various vision based datasets that can be used for recognition purpose using gait. OUISIR (Osaka University) created an ample amount of gait database by including the above mentioned conditions. Currently, they have around 4007 subjetcs in their datbase. CASIA (Chinese Academy of Sciences) has segregated the dataset into 4 categories, namely: CASIA-A, CASIA-B, CASIA-C and CSAIA-D. CASIA-B dataset contains 124 subjects which have been formed by taking into account 11 different viewing directions with each one having an interlude of 18˚.

TABLE 1.2 Different Gait Datasets

Sr.No.	Datasets	Total	Gender Ratio (M/F)	Condition	Camera
1	KY4D Curve Walk	42	-	3 views: front and side	16
2	KY4D Shadow Database	54	-	Varying cloth and carrying conditions	1
3	OUISIR Speed Transition	179	-	Speed changes	1
4	Human Motion & Control lab	1	11/4	5 walking variations	10
5	KIST	113	50/63	Alteration in multiview (8) having consistent speed (3km/hr)	8
6	OUISIR Large Population	4007	2135/1872	-	2
7	AVA-Multiview	20	16/4	Multiview	6
8	Indonesian Database	212	102/110	5 variations: view, baggage, surface, type of shoe and time	1
9	OUISIR Treadmill Speed Variation	200	100/100	Variation in speed	25
10	KY4D Straight walk Database	42	-	3 views: front, side	16
11	TUM-IITKGP	34	-	Both occlusion, baggage, 4 variations in walking	1

12	CASIA-B	124	93/31	Variations in view, clothing, baggage	11
13	CASIA-C	153	130/23	Walking alternation	1
14	CASIA-A	20	-	3 views: front, oblique and lateral	1
15	HID-USF	122	-	2 types of shoes, baggage condition, surfaces, views and time instances	2
16	HID-Georgia Technology	20	-	3 view variations	3
17	HID-UMD Database 1	25	-	4 view variations	1
18	HID-UMD Database 2	55	-	T shape pathway	2
19	HID-UMD Small Database	12	-	5 view variations: 0˚, 15˚, 30˚, 45˚, 60˚	-
20	SOTON small database	12	-	Variations in: shoe (5), clothing (3), speed (3), view with bag	1
21	SOTON large database	115	-	6 view variations: normal track+ treadmill, oblique track+ treadmill	2
22	UCSD	6	-	1 view direction: front parallel	1
23	CMU Body motion	25	-	Walking styles: slow, fast, slanted and walk carrying ball	3

CHAPTER 2

LITERATURE SURVEY

In recent years, the motion analysis and computer vision fields are primarily focusing on using Gait Energy Images (GEI) for individual and gender determination. Using gait for gender recognition improves the computer perception capability and is useful in applications like visual surveillance, that does tracking of objects and their categorization into different classes. Gait based gender classification generates two classes, namely: male and female. This speeds up the process of searching a suspect from a video database and also improves its efficiency, in marketing area it helps the manager in understanding the requirements and interest of his customers and enables him to provide better services to them.

Since the ancient times, many methods and techniques have been used for perceiving the gender of an individual by making use of multiple features. The application domain of gender classification is very vast but the most common one is video surveillance.

Although there are many advances in the field of gender classification, still the machines are unable to provide the desired accurate result. Many factors are responsible for this, for example, while capturing the data for the different gender there might be change in illumination, posture, facial expression or even age. Even the resolution of the image is a very crucial aspect for classification of gender. If the image is of poor resolution or there is a chance of noise being present in image then prediction of gender of the person becomes quite a difficult feat and the accuracy obtained is also not good.

Using biological features for gender determination provides better results. However, acquiring these features is a tedious and difficult task as we require body sensors for this purpose. Also, this process of feature acquisition can be expensive also. Many researches are being done for finding a better solution. Finger prints can also be used as a means of gender identification. Shehu et al [20] utilized Domain Adaptation, Transfer Learning and 2 ResNet-34 models for the purpose of classifying gender.

Another interesting biological feature for gender classification is speech. Liztio et al [21] proposed BPNN (Backpropagation Neural Network) method for classifying gender using speech as input. The main reason of choosing BPNN was its characteristic of learning from non-linear input and generate complex output by adjusting the input data. The method was applied to three different datasets and achieved an accuracy of 95%.

Li et al [22] proposed the method of average gait image for gender identification. They divided the silhouette image into seven segments namely: head, arm, trunk (which also included the chest area), thigh, front leg, back leg and feet and then analyzed the motion of different human parts. Through their research they concluded that head, back-leg and feet did not have a prominent effect on gender determination. They also observed that walking surface, carrying state, arm and thigh hinder the gender determination. For the purpose of classification they used Support Vector Machine (SVM).

A spatio-temopral approach was proposed by Sudha and Bhavani [23] for classification of gender. They did extraction of four anatomical features and five binary moment features after generating the silhouette images and then applied Probabilistic Neural Network (PNN) and SVM to analyze the performance. They adopted the CASIA Dataset B for this purpose.

Hu et al [24] introduced a supervised modeling approach in which temporal dynamics and shape of both the genders are merged into a subsequent model, which provides the spatio-temporal features. Zhang De [25] proposed a fusion of multi view for gender classification, in which tensor data was extracted from GEI for recognition. He employed Principal Component Analysis (PCA) for dimension reduction. Hassan et al [26] proposed a method of using 5/3 wavelet lifting scheme for gait cycle detection and decomposition and Gait Energy Image for gait Feature extraction and an algorithm that contains a combination of Principal Component Aanalysis and C4.5 was used for dimensionality reduction. It was observed that this method gave better results with even higher noise because of illumination. Singh et al [27] proposed a Bayesian Gait-based Network for gender identification of individuals wearing loose fitting clothes. The video is first segregated into frames and then the network extricates human images which is then used for estimating the poses of humans using ScatterNet Hybrid part Affinity Fields Network. The output of this network is presented to 3D ResNet which captures the human motion to determine the gender. Do et al [28] propsed a method in which instead of using Gait Energy Image Average Gait Image (AGI), which does not require the cycle knowledge of GEI, and Lower Average Gait Image (LAGI) was used. A View Point model is used for estimating the view point which helps the system in automatically estimating the view direction. A Distance Signal model is used for unwanted areas removal such as bags to form a silhouette which is attachment. A view-dependent gender classifier is used for gender determination.

In [29] Liu et al replaced the last layer of CNN with SVM to perform gender classification on CASIA-B dataset. In the first part of the method, descriptors of the input image are extracted from the fully connected layer of pre-trained VGGNet-16 model which are then provided as features for the training of SVM. In the second part, model architecture is slightly modified by making use of the hinge loss function. The results show that SVM is able to better classify gender of input image than Softmax.

Choudhary et al [30] proposed a four step method. They first calculated the Gait Energy Image by averaging the sum of all the silhouette images in one gate cycle. Next step involves reducing the size of the generated Gait Energy Image using PCA. In the next step, parameters: footstep length, swing, posture period, speed and height are calculated and appended with the reduced Gait Energy Image. In the final step training and testing of ANN and SVM is done using this reduced feature vector. The accuracy achieved by this method is 98.16%.

In [31], a method based on the angles in images was proposed by Kitchat et al for identifying the gender using the Gait Energy Images. The overall system is divided into two parts: first model for classification of observation angle and second for classification of gender. The angle classifier generates ten angles-based Gait Energy Images. Then the gender classifier makes use of these images for gender classification. The method achieves an accuracy of 90.74% on SIIT-CN-B dataset and an accuracy of 97.58% using the CASIA-B database.

H. M. and R. K. Esther [32] proposed Gait Energy Image Projection model for the purpose of identifying gender. This model gives the value of some gender related characteristics like arm motion, size of the body that do the classification of gender in present gait cycle. This model is then combined with better descriptors like Gait Entropy Images, GGEI (Gradient GEI), D-GEI (Dynamic-GEI) for better classification of gender. For classification purpose SVM was used.

CHAPTER 3

METHODOLOGY

3.1 RESEARCH OBJECTIVE AND PROBLEM FORMULATION

If we consider the present situation of our society, we can clearly see that the crime rate against women has increased drastically. In daily news we can find at least one article pertaining to crimes against women. As, a result we require proctoring of such activities in our day-to-day life. If we consider traditional biometric methods like face, iris, speech then it can prove difficult, as it requires the subject's co-operation as well as it can be easily morphed. If we consider the current pandemic situation, then close contact may not be practical also. Thus, in such cases identifying the gender using gait can prove beneficial. Further, after diagnosing whether the perpetrator is male or female, we can also monitor the actions of the subject and if it looks suspicious, we can alert the authority of the mishap.

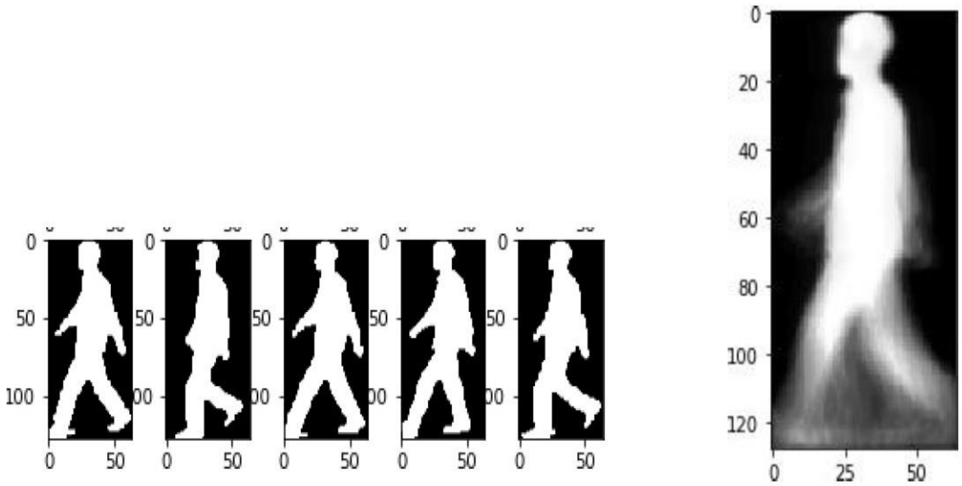
In our project work, we have covered the gender classification aspect of the research objective. Sine the dataset used comprised of silhouette images, it was challenging to train regular neural networks using the same. The networks were unable to identify the prominent features for the purpose of gender classification. Also, the dataset available was small in size. We have used Gait Energy Images for the purpose of training as it covers both spatial and temporal features of the complete gait cycle. We have used One-Shot learning using Siamese network, in order to find similarity between the images. We were able to acquire 99 percent accuracy for gender classification in limited dataset available.

3.2 GAIT ENERGY IMAGE GENERATION

The initial step of the experiment is Gait Energy Image generation. The model free method of gait is susceptible to variations in clothing and carrying conditions. A solution to this problem is GEI which is nothing but a representation of original image using a silhouette image and was proposed by Han and Bhanu [19]. In this, a singular image containing temporal information is used to represent motion information, it is given by (3.1):

$$G(i,j) = \frac{1}{N}\sum_{t=1}^{N} I(i,j,t) \tag{3.1}$$

Here *N* represents the gait cycle frames, *t* is the frame number at a particular time instance and *I(i,j)* represents the original *(i,j)* two-dimensional silhouette images. The GEI have three main characteristics, which are: they are computational time and storage space friendly and are also less sensitive to noise. We have used CASIA-B database for the gender classification task. All the images are first converted into the same dimensions and normalized after which all the frames of a particular subject are combined to form a single image. The result obtained looks as follows:



(a) (b)

Fig 3.1 (a) Sample images from CASIA-B database provided as input for GEI generation (b) Final generated GEI

3.3 METHOD I: DEEP NEURAL NETWORK

For person identification and gender classification through gait energy CNN has been most prominently used. But, the most common belief regarding CNN is that as we increase the layers of a neural network its capacity of capturing complex features also increases, as a result they can be used for representing these complex features better than the shallower features. But, the foremost question is- is this method effective for improving the learning of a model? And what are the advantages and disadvantages of this method? These are some of the important concepts covered in [33] by He et al and many other major concepts are also covered in this paper. This paper explained the concept of ResNet-50 and also covered some essential concepts of DNN (Deep Neural Network).

3.3.1 Degradation problem

The main reason behind the development of ResNet model was to handle the problem of degradation that arises in DNN. When we stack more layers on a network, its accuracy first saturates and then it starts to degrade. We can consider the experiment performed by He et al [33] for better understanding.

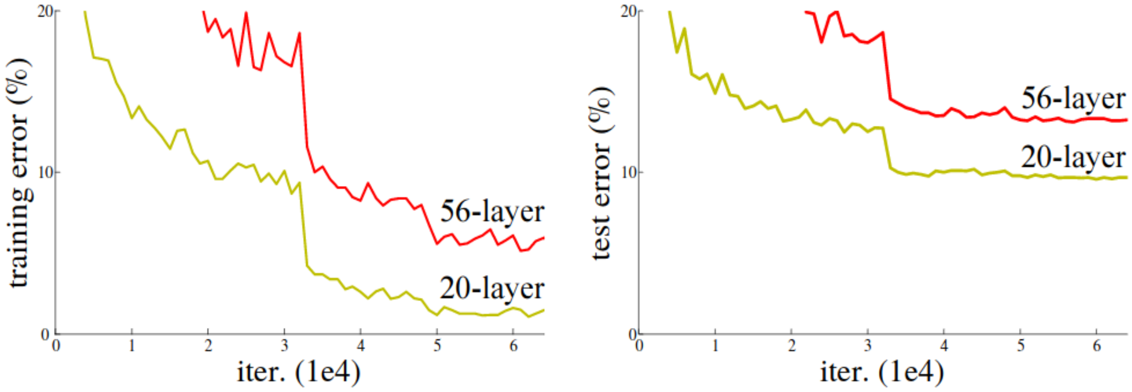


Fig 3.2 Error increases while classifying CIFAR-10 dataset as we increase the layers in Deep Neural Network [33]

As clearly seen from the figure, the training and testing error is more in 56-layer DNN than the 20-layer DNN. With increase in the depth and number of epochs, the error in DNN also increases. If we simplify the problem we can say, we have an adequate deep neural network that calculates adequate strong feature set that is more than enough for the task of image classification. Now if we stack up layers on this already sufficient network then it can prove problematic. If the existing layers of the network can calculate the adequate feature set then the new layers should simply copy those features and forward them. This step might seem simple but it can prove very complex while implementing on DNN.

The above stated solution is the basic idea behind residual block of ResNet. This can be achieved by ensuring the elementary function is always a subset of the complex function. This solution helps in avoiding the problem of degradation. If we consider the input as x and the required input-output mapping as *g(x)* then, rather than handling the complex function we can easily compute the elementary function *f(x)=g(x)-x*. Generally, the optimization of the residual function is much easier than the original function. Also, the optimization of residual function automatically takes care of the identity mapping. The below figure shows the residual block-

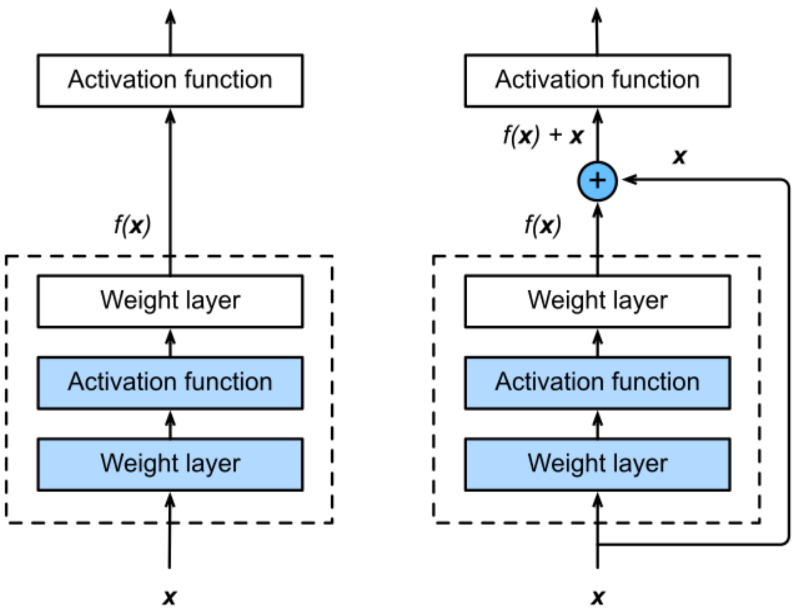


Fig 3.3 Residual block and skip connection in ResNet [33]

This residual block makes sure that when an optimal feature is found, it is as it is passed to the next layer instead of computing a new feature in the next layer. ResNet contains multiple residue blocks which are present at an interval of every two or three network layers.

We first divided the generated gait energy images into training and validation set. The train set consists of 500 images of each male and female and the validation and training set consists of 150 images of each male and female. As the dataset available for training and validation is small we performed augmentation on random images by applying 50 degree rotation, translation on height or width by a factor of 0.2 and horizontally flipping half images. The activation function used for classification is sigmoid instead of softmax as we will be performing the binary classification. The figure below shows the proposed architecture:

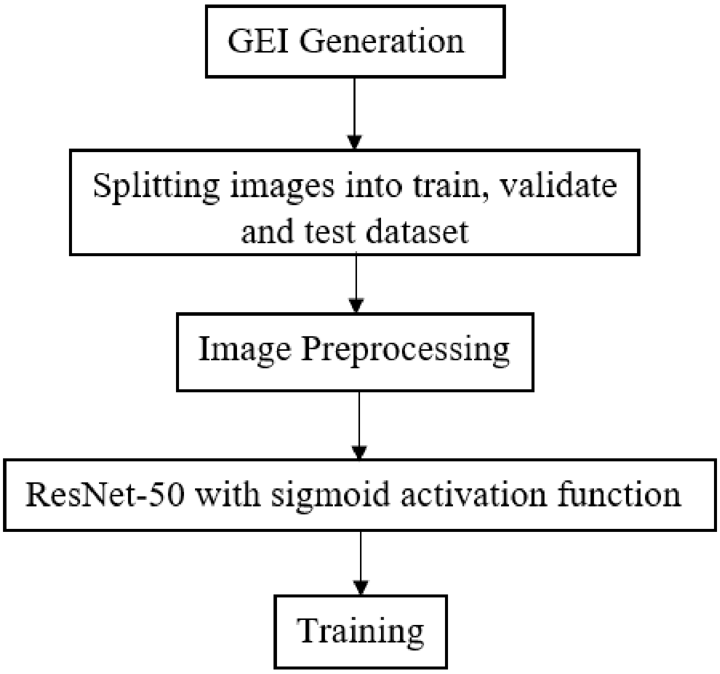


Fig 3.4 Architecture using ResNet-50

This proposed method gives an accuracy of 75%. The problem maybe due to small dataset available for the training purpose or maybe the model is unable to identify the prominent features of gender classification.

3.4 METHOD II: COSINE SIMILARITY

To further discern the main cause of low accuracy we used cosine similarity for the purpose of gender classification. The similarity index is nothing but cosine value of angle formed between two or more vectors which tells how similar the vectors are. Mathematically, it is defined as vectors dot product divided by the product of each vector magnitude, written as (3.2):

$$cosine\ similarity = \cos\theta = \frac{X \cdot Y}{\|X\|\|Y\|} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2}\sqrt{\sum_{i=1}^{n} Y_i^2}} \tag{3.2}$$

There are many applications of cosine similarity like finding the similarity between documents [34] and comparing two poses by the angle between their joins to find the similarity between them [35]. This method is a modification of the above method. Instead of training the ResNet-50 model for the purpose of classification we simply used it to extract the feature vector of images. These vectors are then used to create a dictionary of feature of vectors for male and female classes of training dataset. This dictionary is the base dictionary for comparing any new image in order to classify it as male or female. So, the new architecture becomes as:

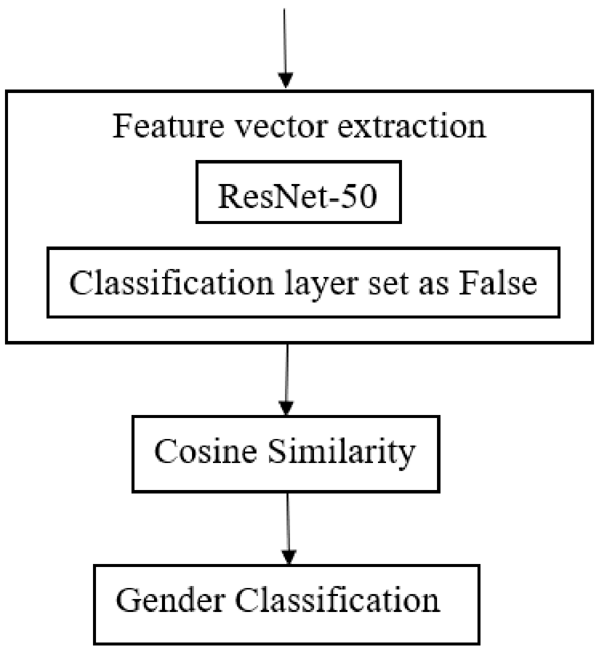


Fig 3.5 Modified Architecture

This method gives an accuracy of 85%. Since, the feature vector used is same as the ResNet-50, it means in the previous method the model was not getting properly trained despite increasing the database using augmentation techniques.

3.5 METHOD III: ONE SHOT LEARNING

The previous methods explained in sections 3.3 and 3.4, were unable to do proper classification due to limited dataset. So, for increasing the accuracy of classification using the limited database one shot learning method was used. The Deep Convolution networks require an abundant amount of labeled data, but sometimes acquiring this amount of data is not feasible. The solution to this problem is one shot learning.

In conventional classification methods, input is provided to multiple layers that process this input and finally classify the image into a category depending upon probability distribution of the image being a part of each class. Let us consider a classification example of four classes of dogs: Labrador, Pomeranian, Spitz and Husky. For each input image, four probabilities are calculated depicting the probability of image being a member of each category. For this purpose, plenty of images are required for each category for the proper training of model. Further, if the model is trained for these four classes then it can't predict correctly for any new class. Suppose, if we want our model to predict a new class of German shepherd then we need to gather a lot of images for that class and retrain the model.

In case of one shot learning, one image is enough for each class. If we try to implement the same example using one shot learning then few images for each class would suffice for the purpose of classification. Rather than directly mapping the input image to one of the four classes, the network takes one more image as reference for input and a similarity score is calculated to find if the 2 input images are of same subject or not. Also, if a new class is to be included then we need only a single image for that class which will be used as referenced input image. Using this reference image similarity value of any new input image can be calculated. Hence, we can say that the network is not learning to directly classify an image to one of the classes instead it is learning a similarity function. One shot learning has been used for developing face recognition system [36], drug breakthrough in cases in which data is sparse [37], and can be useful in banks for the purpose of signature verification [38].

We divide the originally generated gait energy images into training and evaluation sets, with each set consisting of two class folders male and female. First of all we need to load images and provide them in batch form to the model. The loading function returns three tuple variables: the first variable tells how many categories are there in dataset and each category has how many images and what are their dimensions, second is the label which is 0 for male images and 1 for female images, and third is the category which is male and female.

After loading, we need to perform binary classification which is done using supervised learning. Thus, a pair of input and target $(A_i B_i)$ is provided to the model for training. The model takes two images as input and generates similarity value between 0 and 1 as output. Here A_i is set of two images and $B_i = 0$; if the two images are of different gender and $B_i = 1$; if the two images belong to same gender class.

We have used Siamese neural network [39] [40] for one shot learning. The 2 convolution networks shown in the figure below are not dissimilar networks instead they formed by making 2 copies of a network. In other words, they have the same parameters. The 2 input images a1 and a2 are passed through each convolution network to generate output feature vectors f(a1) and f(a2). If both the images belong to the same category then they must have similar feature vectors and if they belong to different categories then they will have different feature vectors. So, both the cases will have a different value for absolute difference between the two feature vectors, as a result the similarity value is also different in both the cases.

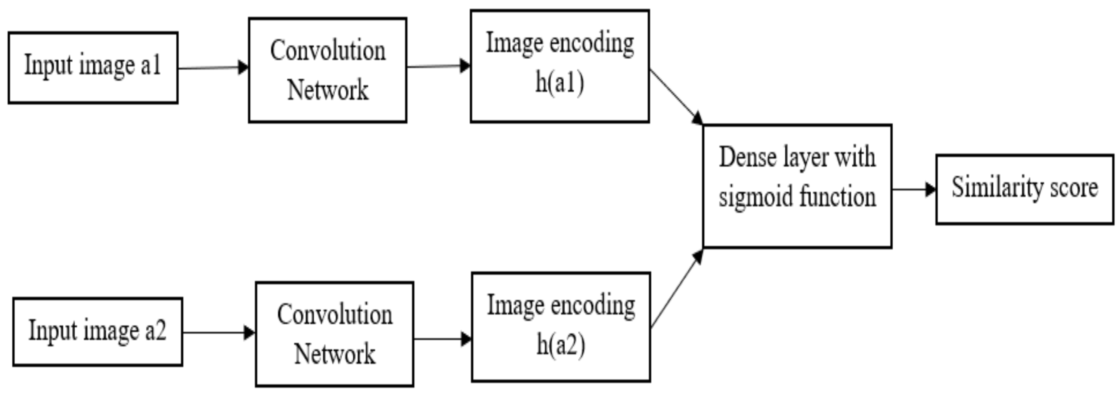


Fig 3.6 Proposed Architecture

Keras does not have any predefined layer for calculating the absolute difference. We did this by incorporating lambda layer in the model which is used for customizing the layer. The different layers in the convolution network and their sizes are shown in the figure below:

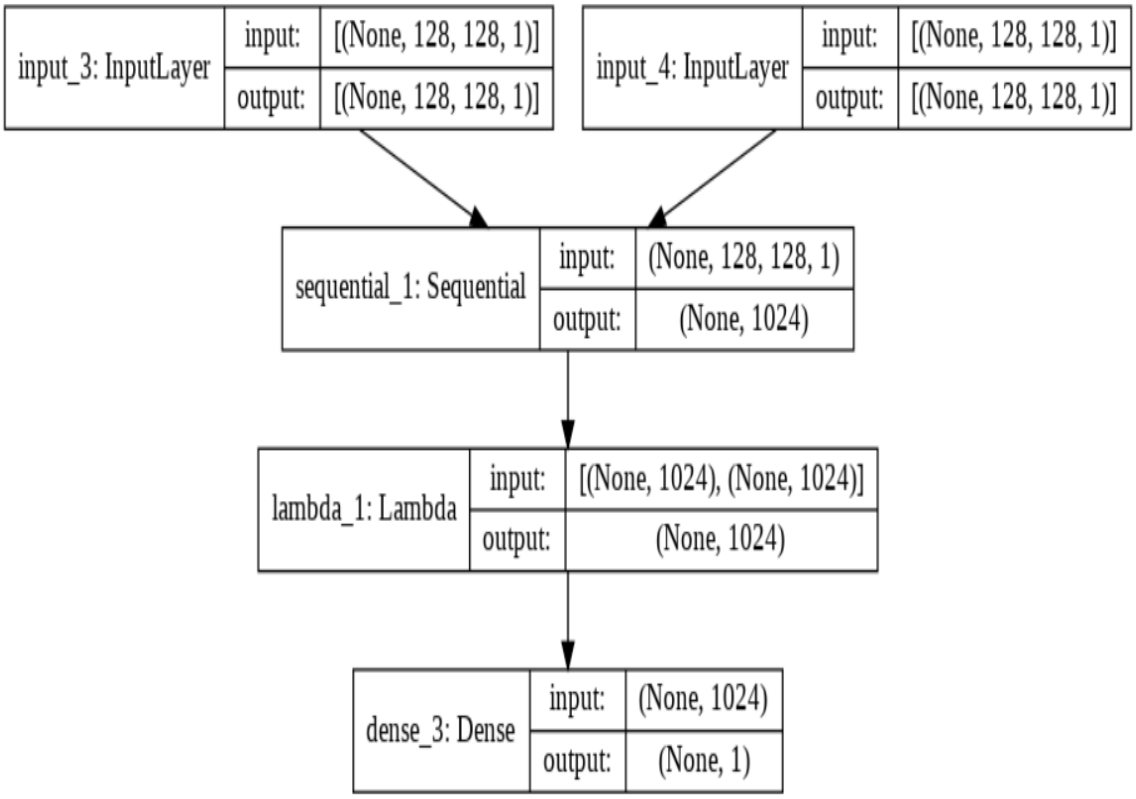


Fig 3.7 Different layers of ConvNet and their shape

For compiling the model, adam and binary cross entropy were used as optimization and loss functions. The learning rate was kept low as on increasing the learning rate the convergence time of the model increased. A total of 1000 iterations were carried with 32 as batch size for the purpose of training. In the wake of every 100 iterations validation of the model was carried out using 2-way one shot learning and accuracy was computed for 150 trials.

3.6 TOOLS USED

In this section, we give an overview of all the software requirements that were necessary for implementing our work, such as the programming language(s), the imported libraries etc.

3.6.1 Programming Platform: Python 3.6

Python is widely used for programming. Developed by a programmer named Guido van Rossum in 1991, it has been extensively developed and used for many large-scale projects.

Also, it is an interpreted language. An interpreted language is a high-level language run and executed by an interpreter (a program which converts the high-level language to machine code and then executing) on the go; it processes the program a little at a time. It involves programming at high level, great for beginners, and a programmer can focus on what to do, and less on how to do that, due to its easy syntax and huge variety of import libraries.

3.6.2 Libraries Used

Multiple libraries and open-source packages, that are required to implement the framework, involves python's open-source OpenCv package, Tensorflow with Keras as backend and other libraries useful for machine learning applications such as scikit and pandas.

OpenCV

Open-CV (Open Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision. In simple language it is library used for Image Processing. It is mainly used to do all the operation related to Images.

It can do a lot of tasks including, read and write Images, Detection of faces and its features, Detection of shapes like Circle, rectangle etc. in an image. E.g. Detection of coin in images, Text recognition in images. e.g. Reading Number Plates, modifying image quality and colours example Instagram, Cam-Scanner, Developing Augmented reality apps and many more. Some of the pros of using open-cv involves:

- One of the powerful, platform independent library and supports C++, Java, Python.
- Provides good support for basic shape detection e.g. Circle, rectangle etc.
- Haar Cascade: Cascade classification is one of the powerful feature of Open-CV, if you are looking for shape or object detection e.g. Face detection
- Video support is good, compatible with almost all the webcams
- Feature detection like detection exact shape of the face like a contour representing lips or eyelash there are many algorithms available like Active shape model, active appearance model
- Widely used in Augmented reality application
- Support camera calibration

Tensor Flow

With the growth of computer science, we saw a boom in the amount of data. As this happened deep learning began to beat all other machine learning. To use this opportunity, Google, to improve its services, thought to use these neural networks:

- Gmail
- Photo
- Google search engine

They came up with TensorFlow so that researchers and developers can work together on AI. In its developed and scaled mode, it gave a chance to a lot of people to bring it to use. The first public version was released in 2015, and the first stable one was released in 2017. Apache Open-Source license has made it open source. We can use, modify and redistribute without any fee. Tensor Flow architecture has three parts for working:

- Pre-processing the data

- Build the model
- Train and estimate the model

The word TensorFlow came up as it uses multi-dimensional arrays called Tensors as inputs. We construct a kind of flowchart called the graph of functions that we want to do on that input. Data is input from one side and we perform multiple operations on it and get an output from the other side. This is the reason it is called TensorFlow as input flows.

Pandas

Pandas is a basic tool for our data. Pandas familiarizes you with your data by cleaning, transforming, and analyzing it. Pandas help us to know our data as we can clean, transform, and analyze the data. Like, if we have to use a dataset in CSV format, Pandas extracts data from the CSV file to a Data-Frame (table). Then we can perform things like:

- Calculation of statistics
- Removing missing values and filtering rows and columns using some criteria
- Visualization of data like plot bars, lines, histograms, bubbles, etc.
- Storing the clean and transformed data into a CSV file or other form of a database

Keras

Although the deep neural networks are growing to become more and more popular, many frameworks are so complex that they have become a barrier to use it. Many high-level APIs have been proposed which are simple and better for developing neural networks, which look same but truly are very different on examining.

One of the highly popular high-level neural networks API is Keras. It supports many backend neural network engines. Keras provides a user-friendly experience which is modular, easily extendible and easy to work with Python. Standalone modules like optimizers, activation functions, cost functions, neural network layers, etc. can be permuted to obtain newly designed models. Modules can be easily added like classes and functions.

Scikit-Learn

Many supervised and unsupervised learning algorithms are provided by Scikit-learn with a good interface in python. It focuses on robustness and support requirements in production systems. It implies focus on ease of use, quality coding, performance, documentation, etc.

Matplotlib

It gives an outstanding visualization in python for 2D plots. Matplotlib is built on NumPy arrays for multiplatform visualization and uses SciPy stack which is for broader use. It was presented by John Hunter in 2002.

Visualization's greatest advantage is that we can visually see large data in easy to understand graphs, etc. It has plots like line, bar, scatter, histogram, etc.

Pickle

Pickle is a module which helps to change or modulate the object structures in a way that is friendly to python and which makes it easy to work upon. All types of python objects can be pickled, with the help of pickle library and then written and stored on the disk.

CHAPTER 4

EXPERIMENTAL RESULTS

The model computes similarity value for each pair of input image which lies between 0 and 1. But we can't judge whether the model is classifying the images correctly just by looking at these values. Thus, the validation and testing of the model is done using 2-way one shot learning. In this the input image is compared with two other images for similarity. Let us consider after the comparison two similarity scores S1 and S2 are generated. If the training of model is done properly then the similarity score of the input image with correct category is greater than other categories. This maximum score is considered as the right prediction and rest are considered wrong predictions. If we repeat this process 'n' times, then we can find the correct predictions percentage by (5.1):

$$percentage = (100 * correct\_pred)/n \tag{5.1}$$

Where k >= total trials and correct_pred >= total correct predictions out of n trials

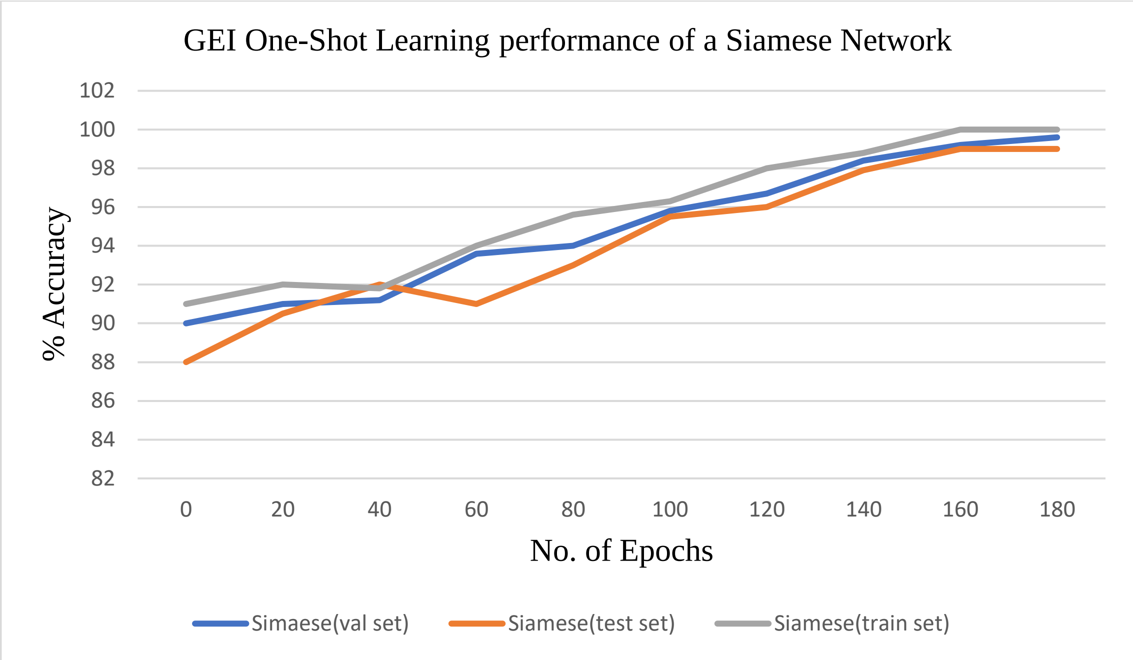


Fig 4.1 Accuracy graph of proposed model

We keep on increasing the number of trials to see if the model works for large number of testing images and see how our model works with the dataset.

TABLE 4.1 Comparison table of existing and proposed method

Sr.No.	Method	Model	Accuracy
1	Liu et al [29]	VGGNet-16 and SVM	89.62
2	Choudhary et al [30]	ANN and SVM	98.16
3	Kitchat et al [31]	CNN	97.58
4	H. M. and R. K. Esther [32]	Gait Energy Image Projection model	96
5	**Proposed Method**	**One Shot Learning**	**99**

The accuracy obtained by our proposed model is 99% which is better than the current existing methods for identifying gender using gait energy image. But the biggest advantage of this model is it does not require large database for training purpose and if new classification class needs to be added then it can be easily done by providing few images for that class

CHAPTER 5

CONCLUSION

Gait Biometric is currently dominating the research area because it has non-perceivable and unobtrusive characteristics, and can prove beneficial in the field of visual surveillance. The report contains a survey of the existing methods that have been employed in the field of gait biometrics and also gives an insight to the various methods that are being used for gender classification through gait biometrics. KNN has been extensively used by the researchers for classification, but currently Deep learning is being explored for gait recognition and the results are promising.

Gait is an identification feature whose acquisition can be easily done from certain interval without the subject's cooperation. Since, we are interested in walking cycle we can convert the colored images captured from the frames of video can be converted to silhouette images, as it can reduce the computation. We have used the CASIA-B dataset which already contains the silhouette images. We have first formed the gait energy images by averaging the sum of all the silhouette images of a subject. These images are provided as a pair of input to the model. The siamese network instead of directly mapping the input image to a class finds the similarity quotient between the images. We get a classification accuracy of 99% with the suggested method

We have done work on normal walking silhouettes of the subjects. In the future, the proposed method can be used for classifying the gender even if there is change in clothing condition, environment or the subject may be carrying additional baggage.

Further, it can be concluded that further work is needed to provide accuracy in case of change in conditions. The accuracy so far achieved is under normalized conditions only and performance deteriorates because of view variation, speed variation carrying condition, appearance change and occlusion. These are the fields that can be explored in future.

5.1 LIMITATIONS

Behavioral and physical properties of individual are very efficient methods for automatic identification and verification in surveillance and applications in security. Biometric traits like iris, fingerprints and face have proven beneficial for human verification or identification, but they suffer from perceivable nature and obtrusiveness. Gait Biometrics has become popular because of its non-invasive, non-perceivable and unobtrusive nature. Factors like walking surface, footwear, carrying condition, leg injuries, clothing condition etc. can change the walking pattern of a person and affect the effectiveness of gait. There are 2 main research challenges in gait: occlusion and view and appearance changes.

5.1.1 Occlusion

Generally a full gait cycle is required for gait recognition, but due to occlusion the acquisition of the full gait cycle proves challenging. For eg. one or more person walking together, a person walking in front of other may lead to occlusion as shown in Figure 2. Thus, extraction of an entire gait cycle proves challenging. Static occlusion (static obstacle) and Dynamic occlusion (when two or more people walk together) are two types of occlusions in gait.

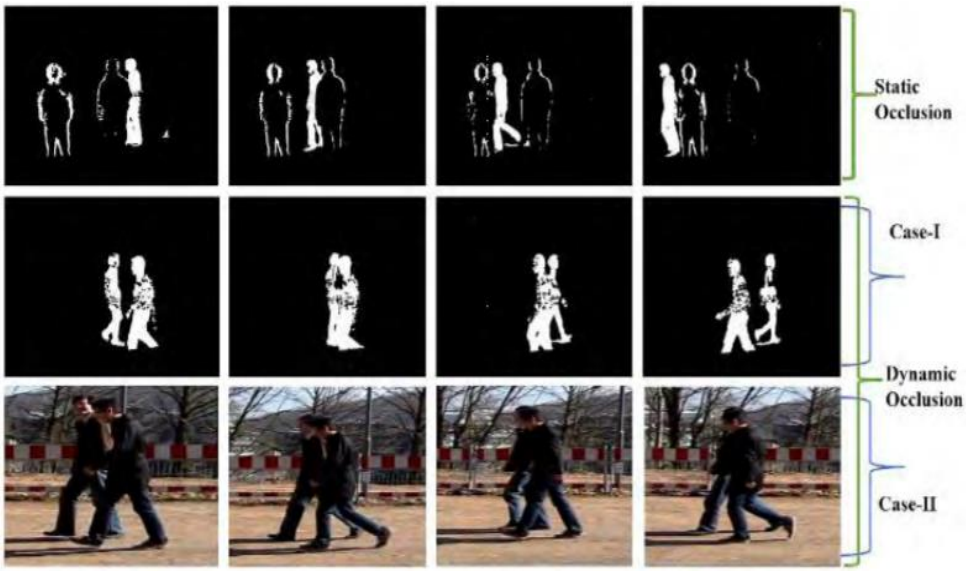


Fig 5.1 Types of occlusion [41]. The first row shows static occlusion while the second and third show dynamic occlusion

a) Static occlusion occurs due to many reasons like people standing in the view or other non-moving hurdles like pillars, planks etc.

b) There are two types of dynamic occlusion: 1) firstly, when one or more walker come into the path of the subject's gait path 2) secondly, more than one person walk together.

A novel dataset was proposed by Hofmann et al [41] to deal with the problem of both the occlusions in recognition of gait. Based on the GEI and color histogram they proposed two algorithms. They found that color histogram was subject to more variations because of variations in clothing conditions on the other hand GEI proved as an efficient method for recognition of gait as it is appearance-based feature invariant.

5.1.2 View and Appearance changes

The view changes and appearance changes can vastly impact gait recognition and gender recognition performance. Clothing may prove beneficial for gender determination in some cases but there is a possibility of someone wearing loose-fitting clothes that may cause an error in determining the gender. Similarly, the view from which the subject dataset was captured and the view from which the testing dataset was captured plays a very important role. This issue can be easily fixed by capturing training dataset from all possible views. So, if some different view input is given it can be easily estimated from the existing database. Another method could be capturing dataset from all 360-degree angles.

The problems in gait recognition often occur due to the changes in view angles. Three types of view angles of acquiring gait are discussed as follows:

1. When the view angle for acquisition of both gallery gait and probe is the same.
2. When the view angle for acquisition of both gallery gait and probe is different.
3. When the probe is captured from a single view and gallery gait is captured from multiple views.

CHAPTER 6

FUTURE WORK

A lot of research work has been done on gender recognition and gender classification and many favorable results have been achieved for the same, but there are still many issues in the real-world application of using gait as a biometrics approach for identification and gender classification. Some of the future perspectives are as follows:

6.1 CREATION OF GAIT DATASET

In recent years many gait databases have been developed but all these databases are susceptible to demographic and geographic conditions. Moreover, researchers develop the databases according to their rules and conditions. The databases so far formed are under controlled environments and also comprise of single gait.

Thus, recognizing a subject using his or her gait in real-world can prove as a new area of research for researchers. Such as identifying one or more subjects walking alone or together in a large crowd of people. This requires a challenging research in which we need to cover the possibility of occlusion and also development of such a database could be challenging as we also need to consider other factors like clothing conditions, baggage, view variations etc.

6.2 VIEW VARIATIONS

View variations is a main research issue for model free approach of gait recognition using silhouette images. The researches so far done have achieved promising results in the case of side views, but performance deteriorates in case of other views and angles. Thus, this can be treated as a further research area. New approaches can be proposed for resolving this issue, like capturing the datasets from all possible views or even developing some view-invariant sophisticated algorithms which identifies the subject from different views with minimum error rate.

6.3 CHANGE IN APPEARANCE

The human walking style changes a lot based on the type of surface they are walking on, the type of shoes or clothes they are wearing or even due to carrying bags. These condition changes can prove to be new research area for the researches, they should identify features which are insusceptible to such changes in appearance. Further, datasets with such severe appearance changes can be created. For example, if we consider the Indian attire saree, it affects the gait cycle more than the western style like skirts and jeans. So, far there is no such dataset in existence.

6.4 ADAPTIVE MODELS FOR BACKGROUND

Gait recognition in unrestricted environment is still a major research area for researchers. The traditional methods so far perform the segmentation of human parts of silhouette images after subtracting the background from the images. Thus, it opens the possibility of developing models that can easily adapt to the background changes and hence be able to identify the subject under the unrestricted conditions like changing background, conglomerate background, shadows in background, change in lighting conditions and even occlusion.

6.5 REDUCING FEATURE SPACE

The works so far done in gait recognition consider a general feature vector instead of identifying specific features that have higher probability of identifying any individual. Genetic algorithm was even proposed for optimizing the selection of identification features [42], however more work is still needed for improving the efficiency. For this purpose, optimized methods can be used for extracting the specific features that help in improving the accuracy.

REFERENCES

[1] N. M. Bora, G. V. Molke, and H. R. Munot, "Understanding human gait: A survey of traits for biometrics and biomedical applications," 2016, doi: 10.1109/ICESA.2015.7503444.

[2] B. Stephens-Fripp, F. Naghdy, D. Stirling, and G. Naghdy, "Automatic Affect Perception Based on Body Gait and Posture: A Survey," *Int. J. Soc. Robot.*, 2017, doi: 10.1007/s12369-017-0427-6.

[3] and R. C. K. Murray, M. Pat, A. Bernard Drought, "Walking patterns of normal men." JBJS 46.2 (1964): 335-360.

[4] M. H. Nadhif, A. P. Hadiputra, Y. Whulanza, and S. Supriadi, "Gait analysis for biometric surveillances using kinect^{TM}: A study case of axial skeletal movements," 2019, doi: 10.1109/QIR.2019.8898273.

[5] P. Akulwar and N. A. Vijapur, "Secured Multi Modal Biometric System : A Review," 2019, doi: 10.1109/I-SMAC47947.2019.9032628.

[6] K. A. Boyer, R. T. Johnson, J. J. Banks, C. Jewell, and J. F. Hafer, "Systematic review and meta-analysis of gait mechanics in young and older adults," *Exp. Gerontol.*, 2017, doi: 10.1016/j.exger.2017.05.005.

[7] C. Dhiman, M. Saxena, and D. K. Vishwakarma, "Skeleton-based view invariant deep features for human activity recognition," 2019, doi: 10.1109/BigMM.2019.00-21.

[8] C. D. , D.K. Vishawakarma, P. Aggarwal, "Skeleton based Activity Recognition by Fusing Part-wise Spatio-temporal and Attention Driven Residuesitle," *arXiv preprint arXiv:1912.00576 (2019).* .

[9] S. Choi, J. Kim, W. Kim, and C. Kim, "Skeleton-Based Gait Recognition via Robust Frame-Level Matching," *IEEE Trans. Inf. Forensics Secur.*, 2019, doi: 10.1109/TIFS.2019.2901823.

[10] "Audio- and Video-Based Biometric Person Authentication - Third International Conference, AVBPA 2001, Proceedings," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2001.

[11] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," 2002, doi: 10.1109/AFGR.2002.1004148.

[12] J. H. Yoo, K. Y. Moon, D. Hwang, and M. S. Nixon, "Automated human recognition by gait using neural network," 2008, doi: 10.1109/IPTA.2008.4743792.

[13] C. Dhiman and D. K. Vishwakarma, "High Dimensional Abnormal Human Activity Recognition Using Histogram Oriented Gradients and Zernike Moments," 2018, doi: 10.1109/ICCIC.2017.8524372.

[14] C. Dhiman and D. K. Vishwakarma, "A robust framework for abnormal human action recognition using\boldsymbol{\mathcal{r}} -transform and zernike moments in depth videos," *IEEE Sens. J.*, 2019, doi: 10.1109/JSEN.2019.2903645.

[15] D. K. Vishwakarma and C. Dhiman, "A unified model for human activity recognition using spatial distribution of gradients and difference of Gaussian kernel," *Vis. Comput.*, 2019, doi: 10.1007/s00371-018-1560-4.

[16] C. Dhiman and D. K. Vishwakarma, "View-Invariant Deep Architecture for Human Action Recognition Using Two-Stream Motion and Shape Temporal Dynamics," *IEEE Trans. Image Process.*, 2020, doi: 10.1109/TIP.2020.2965299.

[17] W. Kusakunniran, "Recognizing Gaits on Spatio-Temporal," *Ieee Trans. Inf. Forensics Secur.*, 2014.

[18] C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification," 2002, doi: 10.1109/AFGR.2002.1004182.

[19] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, doi: 10.1109/TPAMI.2006.38.

[20] Y. I. Shehu, A. Ruiz-Garcia, V. Palade, and A. James, "Detailed Identification of

Fingerprints Using Convolutional Neural Networks," 2019, doi: 10.1109/ICMLA.2018.00187.

[21] L. Maulisa Liztio, C. Atika Sari, D. R. Ignatius Moses Setiadi, and E. Hari Rachmawanto, "Gender identification based on speech recognition using backpropagation neural network," 2020, doi: 10.1109/iSemantic50169.2020.9234237.

[22] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu, "Gait components and their application to gender recognition," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 38, no. 2, pp. 145–155, 2008, doi: 10.1109/TSMCC.2007.913886.

[23] L. R. Sudha and R. Bhavani, "An efficient spatio-temporal gait representation for gender classification," *Appl. Artif. Intell.*, vol. 27, no. 1, pp. 62–75, 2013, doi: 10.1080/08839514.2013.747373.

[24] M. Hu, Y. Wang, Z. Zhang, and D. Zhang, "Gait-based gender classification using mixed conditional random field," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 41, no. 5, pp. 1429–1439, 2011, doi: 10.1109/TSMCB.2011.2149518.

[25] Z. De, "Research on Gait-Based Gender Classification via Fusion of Multiple Views," *Int. J. Database Theory Appl.*, vol. 8, no. 5, pp. 39–50, 2015, doi: 10.14257/ijdta.2015.8.5.04.

[26] O. M. S. Hassan, A. M. Abdulazeez, and V. M. T?ryak?, "Gait-Based Human Gender Classification Using Lifting 5/3 Wavelet and Principal Component Analysis," *ICOASE 2018 - Int. Conf. Adv. Sci. Eng.*, pp. 173–178, 2018, doi: 10.1109/ICOASE.2018.8548909.

[27] A. Singh, A. Kumar, and A. Jain, "Bayesian gait-based gender identification (BGGI) network on individuals wearing loosely fitted clothing," *Proc. - 2019 Int. Conf. Comput. Vis. Work. ICCVW 2019*, pp. 1828–1835, 2019, doi: 10.1109/ICCVW.2019.00227.

[28] T. D. Do, V. H. Nguyen, and H. Kim, "Real-time and robust multiple-view gender classification using gait features in video surveillance," *Pattern Anal. Appl.*, vol. 23, no. 1, pp. 399–413, 2020, doi: 10.1007/s10044-019-00802-6.

[29] T. Liu, X. Ye, and B. Sun, “Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition,” 2019, doi: 10.1109/CAC.2018.8623118.

[30] S. Choudhary, C. Prakash, and R. Kumar, “A hybrid approach for gait based gender classification using GEI and spatio temporal parameters,” 2017, doi: 10.1109/ICACCI.2017.8126100.

[31] K. Kitchat, N. Khamsemanan, and C. Nattee, “Gender classification from gait silhouette using observation angle-based GEIs,” 2019, doi: 10.1109/CIS-RAM47153.2019.9095797.

[32] M. Hema and K. Esther Rachel, “Gait energy image projections based on gender detection using support vector machines,” 2020, doi: 10.1109/ICCES48766.2020.09137900.

[33] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2016, doi: 10.1109/CVPR.2016.90.

[34] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, “Cosine similarity to determine similarity measure: Study case in online essay assessment,” 2016, doi: 10.1109/CITSM.2016.7577578.

[35] S. K. Biswas and P. Milanfar, “One Shot Detection with Laplacian Object and Fast Matrix Cosine Similarity,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2016, doi: 10.1109/TPAMI.2015.2453950.

[36] S. Chanda, A. C. Gv, A. Brun, A. Hast, U. Pal, and D. Doermann, “Face recognition - A one-shot learning perspective,” 2019, doi: 10.1109/SITIS.2019.00029.

[37] H. A.-T. , Vijay Pande, Bharath Ramsundar, Aneesh S. Pappu, “Low data drug discovery with one-shot learning.” ACS central science 3.4, pp. 283–293, 2017.

[38] C. S. Vorugunti, G. S. Devanur, P. Mukherjee, and V. Pulabaigari, “OSVNet: Convolutional siamese network for writer independent online signature verification,” 2019, doi: 10.1109/ICDAR.2019.00236.

[39] Z. Zhu, Q. Wang, B. Li, W. Wu, J. Yan, and W. Hu, “Siamese Neural Networks

for One-shot Image Recognition Gregory," *ICML Deep Learn. Work.*, 2018.

[40] A. V. J. Bertinetto, J.F. Bertinetto, J. Valmadre, P.H. Torr, "Learning feed-forward one-shot learners," 2016.

[41] M. Hofmann, S. Sural, and G. Rigoll, "Gait recognition in the presence of occlusion: A new dataset and baseline algorithms," 2011.

[42] F. Tafazzoli, G. Bebis, S. Louis, and M. Hussain, "Genetic feature selection for gait recognition," *J. Electron. Imaging*, 2015, doi: 10.1117/1.jei.24.1.013036.