University of Padova

Department of Information Engineering

Master Thesis in ICT FOR INTERNET AND MULTIMEDIA

Continual Learning on Gait Recognition

Supervisor
Simone Milani
University of Padova

Co-supervisor
Daniele Mari
University of Padova

Master Candidate
Theophilus Yaw Owusu

Student ID
1216099

Academic Year
2022-2023

“Dedication”

in loving memory of my parents

Abstract

Gait recognition is a critical area of research in biometric identification, which has traditionally required complex instrumentation and systems for detection and analysis. However, wearable devices, such as smart glasses for gait recognition, have the advantages of being less complex and more user-friendly than traditional vision-based systems.

In this thesis an approach for gait recognition is presented using a set of smart glasses for gait data acquisition and deep learning as universal feature extractors. Our study aims to identify individuals based on their walking patterns using smart glasses' accelerometer, gyroscope, and magnetometer data.

The findings suggest that smart glasses with embedded sensors can acquire the required data for training and testing machine learning algorithms, making detecting and recognizing gait in different scenarios possible.

Essentially, the experimental findings also suggest that our approach leads to classification accuracy for multi-class (8 individuals classification) with an average accuracy of 93% and male/female classification with an average accuracy of 97%.

Furthermore, the study addresses the challenge of maintaining the model's performance on previously recognized and classified gaits after learning new ones without being retrained on their data. This challenge was addressed by adopting rehearsal method in continual learning to our model.

Sommario

Il riconoscimento dell'andatura è un'area critica della ricerca nell'identificazione biometrica, che ha tradizionalmente richiesto strumentazione e sistemi complessi per il rilevamento e l'analisi. Tuttavia, i dispositivi indossabili, come gli occhiali intelligenti per il riconoscimento dell'andatura, dimostrano i vantaggi di essere meno complessi e più intuitivi rispetto ai tradizionali sistemi basati sulla visione.

In questo documento, presentiamo il nostro approccio al riconoscimento dell'andatura, che prevede l'utilizzo di occhiali intelligenti per l'acquisizione dei dati sull'andatura e il deep learning come estrattori di caratteristiche universali. Il nostro studio mira a identificare gli individui in base ai loro schemi di deambulazione utilizzando i dati dell'accelerometro, del giroscopio e del magnetometro degli occhiali intelligenti.

I risultati suggeriscono che gli occhiali intelligenti con sensori incorporati possono acquisire i dati richiesti per l'addestramento e il test degli algoritmi di apprendimento automatico, rendendo possibile il rilevamento e il riconoscimento dell'andatura in diversi scenari.

In sostanza, i risultati sperimentali suggeriscono anche che il nostro approccio porta all'accuratezza della classificazione per multi-classe (classificazione di 8 individui) con un'accuratezza media del 93% e classificazione maschio/femmina con un'accuratezza media del 91%.

Inoltre, lo studio affronta la sfida di mantenere le prestazioni del modello su andature precedentemente riconosciute e classificate dopo averne apprese di nuove senza essere riqualificato sui loro dati. Questa sfida è stata affrontata adottando il metodo di prova nell'apprendimento continuo del nostro modello.

Contents

Abstract vi

Sommario vii

List of figures xiii

List of tables xv

Listing of acronyms xvii

1 Introduction 1
- 1.1 Human Gait: The Fundamentals . . . 3
- 1.2 Motivations and Contributions . . . 6
- 1.3 Structure of Thesis . . . 6

2 State-of-the-Arts 7
- 2.1 state-of-the-art on gait recognition . . . 7
  - 2.1.1 silhouette-based gait Recognition . . . 8
  - 2.1.2 3D-based gait recognition . . . 10
  - 2.1.3 deep learning-based gait recognition . . . 13
  - 2.1.4 wearable sensor-based gait recognition . . . 15
- 2.2 state-of-the-art on continual learning . . . 19
  - 2.2.1 Replay methods . . . 19
  - 2.2.2 Regularization-based methods . . . 21
  - 2.2.3 Parameter isolation methods . . . 22

3 Smart Glasses 25
- 3.1 Smart Glasses . . . 25
- 3.2 Data Gathering . . . 27
  - 3.2.1 Embedded Sensors . . . 27
  - 3.2.2 Raw Dataset . . . 30

4 Methods 33
- 4.1 Data Preprocessing . . . 33
- 4.2 Processing Structure . . . 34
  - 4.2.1 Filtering . . . 34

4.2.2 Walking cycle extraction 36
4.2.3 Normalization 36
4.3 Convolutional neural network 37
4.3.1 Dataset 37
4.3.2 CNN architecture 38

5 Continual Learning on Gait Recognition **41**
5.1 Rehearsal Method 41
5.1.1 Rehearsal Process 42
5.1.2 Evaluation 43

6 Results **45**
6.1 Gait Recognition 45
6.2 Continual Learning 49

7 Conclusion **53**

References **55**

Acknowledgments **59**

Listing of figures

1.1 Smart Wearable Users(millions), 2019-2026 1
1.2 Gait phases of a normal gait. 4

2.1 Examples of extracted, silhouette frames in different human walking sequences. 8
2.2 The bottom row ((f)-(j)) shows sample silhouette frames with a variety of segmentation errors. The raw image corresponding to each silhouette is shown on the top row ((a)-(e)[1]. 9
2.4 block diagram depicting the steps taken to develop the library for the human gait in[2]. 13
2.6 System design & Operation of IoT-based sensor shoe [3]. 18

3.1 Labeled component of the Smart Glasses. 26
3.2 Roll, pitch, and yaw angles 28

4.1 Raw gait before and after preprocessing. 35
4.2 Template extraction using the accelerometer magnitude $a_{mag}(i)$. 37

5.1 The Rehearsal Method: Continual learning Process 42

6.1 Training and Validation Accuracy Person Identification 46
6.2 Training and Validation Accuracy for Sex Identification 47
6.3 Classification Accuracies for Person Identification 48
6.4 Classification Accuracies for Sex Identification 48
6.5 Accuracies for first task 50
6.6 Accuracies for incremental second task 50
6.7 Accuracies for incremental third task 51

Listing of tables

3.1 Features functionality & LED Behavior. 27
3.2 Sample data of raw gait signal from a subject. 31

4.1 Preprocessed Gait data from Subjects. 38

6.1 Tabular Representation of Person Identification. 47
6.2 Tabular Representation of Sex Identification. 47

Listing of acronyms

IMU .... Inertial Measurement Unit

SotA .... State-of-the-Art

LwF .... Learning without Forgetting

EWC .... Elastic Weight Consolidation

WiFi .... Wireless Fidelity

WD .... Wavelet Descriptors

3D .... Three-Dimensional

CSI .... Channel State Information

NN .... Nearest-Neighbor

EER .... Equal Error Rate

CIFAR-100 .... Canadian Institute for Advanced Research-100

ILSVRC2012 .. ImageNet Large Scale Visual Recognition Challenge 2012

VGG-16 .... Visual Geometry Group-16

ResNets .... Residual Neural Networks

DenseNets ..... Densely Connected Neural Networks

LED .... Light-Emitting Diode

PC .... Personal Computer

MEMS .... Micro Electro-mechanical Systems

CSV .... Comma-Separated Values

EMG .... Electromyograph

API .... Application Programming Interface

SVM .... Support Vector Machine

PCA .... Principal Component Analysis

CNN .... Convolutional Neural Network

FC .... Fully Connected

LRelu .... LeakyReLU

Adam .... Adaptive Moment Estimation

CL .... Continual Learning

1 Introduction

Smart wearable technologies, also known as smart wearable devices, such as fit-bands, smart-watches and smart glasses have been rapidly evolving in recent years fostering improvements in sensors and wireless technologies.

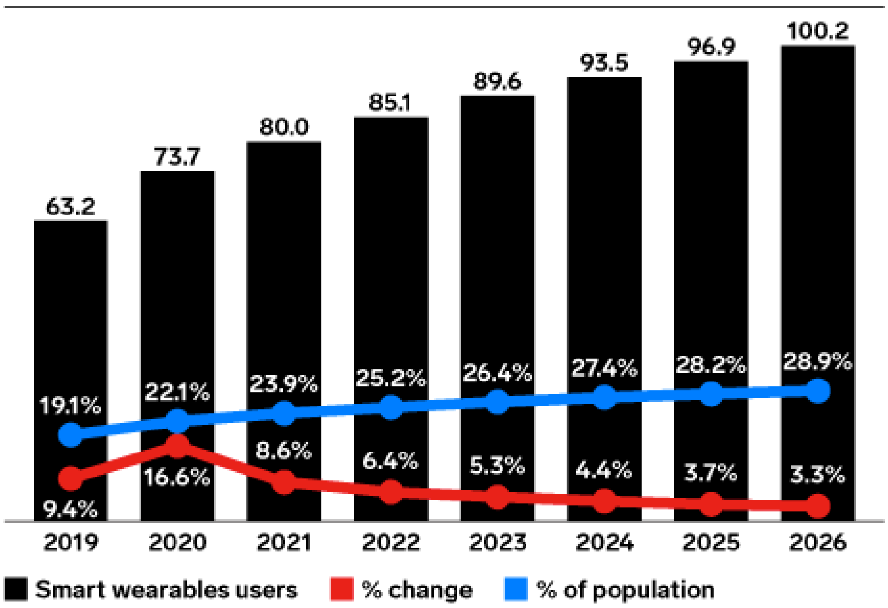


Figure 1.1: Smart Wearable Users(millions), 2019-2026

Figure 1.1 shows the number and percentages of the US population alone, who are smart wearable users.It shows that the numbers of smart wearable users increased by 21.9 million that is, from 63.2 million users in 2019 to 85.1 million users in 2022. However, it also predicted to increase to over 100 million users by the end of 2026[4]. The graph shows the exact data for 2019 to 2022, whiles it provides an estimation from 2023 to 2026. This depicts that users have begun to adopt smart wearable devices in their daily life to get useful information about their activities (e.g. heartbeat, speed, etc...)

As a result, smart wearable devices can be considered essential clothing accessories such as necklaces, bracelets, watches and glasses. Additionally, they can be worn on various parts of the body, complementing as clothing accessories without straining or making its users uncomfortable and usually replacing the existing clothing accessories. For example, smartwatches, fit bands and smart glasses are often used instead of standard wristwatches, fit bands and sunglasses.

The most essential attribute of smart wearable devices is that they incorporate sensor technologies that gather physiological and behavioural biometric data and possess wireless technologies like Bluetooth and WiFi technologies. These wireless technologies enable connectivity with smartphones, and laptops via Bluetooth or to the internet by using WiFi. As a result, smart wearable devices no longer require large amounts of data storage space, allowing for continuous, seamless sensor application processing and software updates.

Moreover, smart wearable devices often contain Inertial Measurement Units (IMUs) that can be used for motion tracking. IMUs built into smart wearable devices include an accelerometer, gyroscope, and magnetometer. They provide various types of sensing information which can be used in various applications such as indoor navigation, and gesture or gait recognition.

Smartwatches are becoming more and more popular in the world of wearable technology, and this trend is likely to continue as the technology gets better. Recent studies demonstrate that both smartphones[5, 6] and smartwatches[7] can implement gait recognition algorithms thanks to the information provided by their sensors.

However, smart glasses are a relatively new device and practical algorithms for their usage still need to be developed appropriately. For this reason, we explore the gait recognition problem with these devices in this thesis. Unfortunately, the algorithms used for smartwatches and

Android phones can only be directly applied if they are positioned on entirely different body parts, making this problem non-trivial. To tackle this issue, we randomly select a set of healthy people and try to assess their identity based on the gait information, which is extrapolated from the IMUs signals.

Additionally, in order to make the algorithm more adaptable, we also validate it in a continual learning scenario. This allows us to increase the difficulty of the task by introducing new users into the system without having to retrain the system.

1.1 Human Gait: The Fundamentals

Gait is simply defined as the way, pattern or style a person walks. It can also be defined as the coordinated, cyclic combination of the movements resulting in human locomotion [8]. Moreover, the gait of a person is a periodic activity with each gait cycle covering two strides (the left foot forward and right foot forward strides) [9].

Accordingly, **Human gait** is the periodic motion of body segments[10] or the pattern of locomotion along with posture [11].

Gait analysis, is, therefore, the study and detailed examination of human gait. Hence in [5], it is defined as the study of human locomotion. On the other hand, it is described as the study of human motion used in assessing and treating different conditions that impair the person's ability to walk properly [12].

A **Gait cycle** is made up of two(2) human steps. To facilitate an in-depth analysis of the human gait cycle, it is usually further divided into **gait phases**. These describe the complete walking process and allow for a better understanding of the processes of periodic walking.

There are primarily two(2) main phases of a gait cycle namely the swing and stance phases. The former corresponds to the part of the cycle where the foot is off the ground and accounts for roughly 40 per cent of the step, while the latter refers to the part of the period when the foot is in contact with the ground which covers the remaining time. The stance and swing phases might be further subdivided into eight segments (five Stance and three Swing), separating the gait cycle into eight sub-phases. The five sub-phases of the stance phase are Initial contact, Loading response, Mid Stance, Terminal Stance and pre-Swing whereas the swing phase has three sub-phases: Initial Swing, Mid Swing and Terminal Swing.

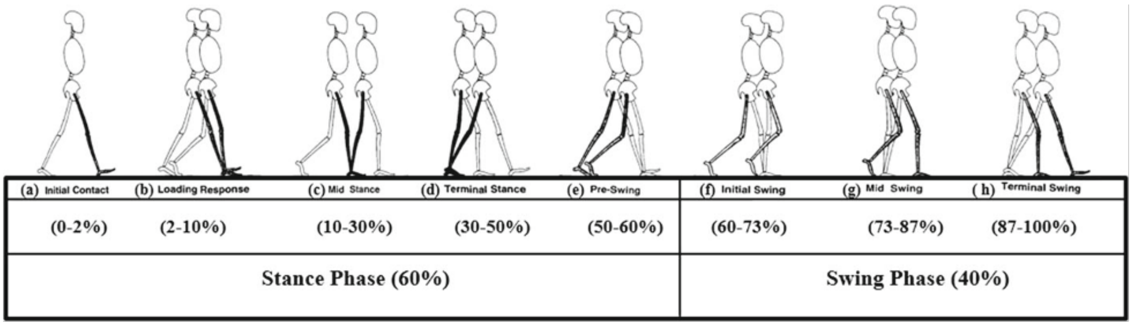


Figure 1.2: Gait phases of a normal gait.

In Figure 1.2, we present the normal gait cycle portraying the swing and stance phases, segregated into eight sub-phases as depicted in [11]. The right foot is shaded black and considered the reference foot for the cycle.

1. **Initial contact:** It is considered as the point at which the heel of the reference foot touches the ground and it is the beginning of the loading response.

2. **Loading response:** It begins with the initial contact of the reference foot and continues until the other foot is lifted for swing. The reference foot comes in full contact with the floor and body weight is fully transferred onto the stance limb.

3. **Mid-stance:** It begins with the toe opposite the reference foot's toe lifted off the ground and ends when the body's weight aligns on the reference foot.

4. **Mid-stance:** It begins with the toe opposite the reference foot's toe lifted off the ground and ends when the body's weight aligns on the reference foot.

5. **Terminal stance:** this phase begins with the heel rise of the reference foot and remains until the opposite foot hits the ground and the body's weight shifts ahead of the reference foot.

6. **Pre-swing:** is the second double stance interval of the gait. It begins with the initial contact of the opposite limb and ends with the toe-off. This positions the limbs for swing.

7. **Initial swing:** it begins with lifting the foot from the floor and ends when the stance foot is opposite the swinging foot.

8. **Terminal swing:** This is the phase when the tibia is perpendicular to the ground and ends with the strike of the foot on the floor.

The main challenge with the recognition and studies of the gait of a person is that it traditionally involves the use of complex instrumentation, systems and methods for gait recognition or detection. Some of these systems require several accessories to be attached to the people whose gait is being detected and analyzed. As such these methods and processes have to be undertaken in a confined space such as a laboratory. This is because these setups are very hard to deploy in outside environments since they are often hard to carry around. This makes it difficult to acquire gait data under different natural scenarios such as walking outdoors, indoors or walking down and up the stairs. However, the advancement in technology in recent years has made it possible to develop and use less complex gait recognition systems such as smart wearable devices, to easily detect and recognize the gait of a person.

Smart wearable devices such as smartwatches and smart glasses can be identified as the least complex and most user-friendly sensor-based systems for gait recognition. They are much more flexible than traditional vision-based systems. For this reason, the development and usage of wearable device-based gait recognition have simplified the process of data acquisition for this.

As we discussed earlier in this chapter, wearable devices are often designed with embedded inertial sensors (accelerometer, gyroscope and magnetometer). This allows them to detect and acquire amounts of data, required to train and test machine learning algorithms, which can be flexibly applied to different scenarios.

1.2 Motivations and Contributions

There is little research, on the usage of smart glasses-based gait recognition for motion tracking in humans, in spite of the numerous prior studies conducted on smart wearable devices-based gait recognition. As a result of this gap, this thesis focuses on the possibility of analyzing raw inertial signals extracted from a set of built-in inertial sensors of smart glasses, to extrapolate the gait characteristics of a few people. Meaningful data is further extracted from the inertial signals, to identify, distinguish and classify each person by their walking style as well as assessing their biological gender.

Additionally, a vital goal of this study is to create a model, that can continue to perform well on all previously recognised and classified gaits after learning new ones, without having to be retrained on their data.

1.3 Structure of Thesis

In this section, we will provide an overview of the thesis outline. Chapter 2 will focus on reviewing the state-of-the-art (SotA) in gait recognition and continual learning. The smart glasses, which are the focus of this study, will be presented in Chapter 3. Chapter 4 will discuss the proposed methodology. In Chapter 5, we will examine the continual learning strategies applied to gait analysis. The experimental results obtained will be explained in Chapter 6. Finally, Chapter 7 will present the conclusions drawn from the study and potential opportunities for further improvement.

2
State-of-the-Arts

2.1 STATE-OF-THE-ART ON GAIT RECOGNITION

Numerous research projects, studies, and analyses have been conducted on gait recognition. Some of these studies and investigations dating back to 1994, nearly two decades ago. In this chapter, we will discuss the main state-of-the-art (SotA) approaches for gait recognition. However, firstly we provide a summary of some traditional techniques to lay some foundations useful for understanding main works discussed later in the chapter.

Gait recognition is an effective human identification method that exploits an individual's walking style to recognize a person. In Sarkar's 2009 Evaluation, "gait recognition" is defined as automated vision methods that use video of human gait to recognize or identify a person.

Gait recognition is a difficult problem that has been tackled in many different ways using different types of data structures (e.g., silhouettes, 3D models, and wearable sensor signals) processed with standard and learned techniques. However, the resolve to tackle this problem has led to a significant increase in gait recognition research, resulting in several SotA approaches in recent years. Each of these approaches have their own peculiarities depending on the purpose and goal that the technique wants to tackle.

The two most common frameworks that try to address this problem are model based and feature based methods. In the case of the former a 3D model and an animation of the person's gait are obtained either thanks to a motion capture system or by analyzing the video or sensor signals. The latter directly extract the silhouette or the movement of different body parts from the video or the sensor data so it skips the 3D model generation step.

Related works on these SoTA approaches are further discussed as follows:

2.1.1 silhouette-based gait Recognition

The majority of the early approaches to human gait detection are based on reference videos or images and one of the most effective ways to gather useful information from this type of signals is to extract and analyze the silhouette of the individual.

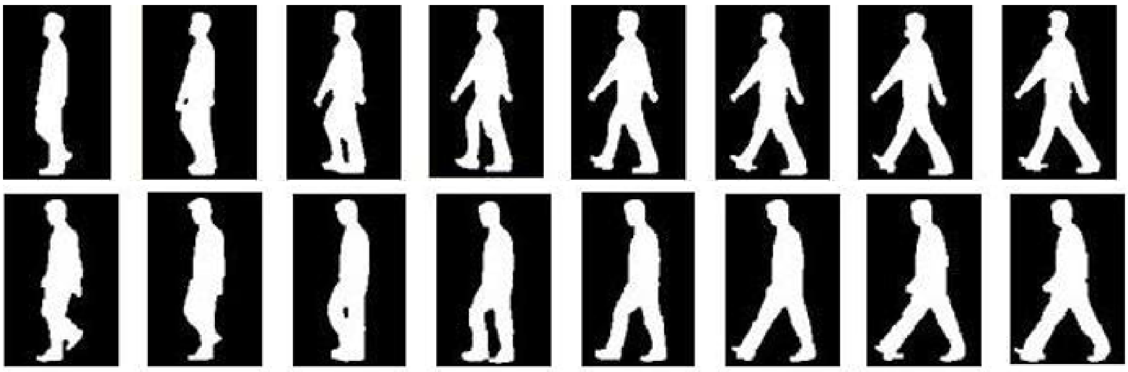


Figure 2.1: Examples of extracted, silhouette frames in different human walking sequences.

Silhouette-based gait recognition is a computer vision-based technique that involves extracting the shape of a person's body from a video sequence from which it is possible to identify the individual's walking pattern. Most algorithms implementing it involve three main steps i.e. background subtraction, silhouette extraction and features calculation.

Background subtraction is the process of separating the foreground (the moving object) from the background. This step is necessary for extracting a silhouette image of a person's walking motion.

Silhouette extraction entails separating the foreground object from the background using segmentation techniques such as thresholding, edge detection, or model-based segmentation.

The silhouette image is then shown as a binary image, with pixels in the foreground set to 1 and pixels in the background set to 0.

Feature extraction entails representing a silhouette image as a set of features that capture an individual's walking pattern. Various methods, such as Fourier descriptors, wavelets, and Gabor filters, can be used to extract the features.

The approach described in [1] is a gait recognition method that uses silhouette template matching to identify individuals from their walking patterns. The method has been evaluated on a moderately large set of subjects (122) under controlled lighting and camera conditions.

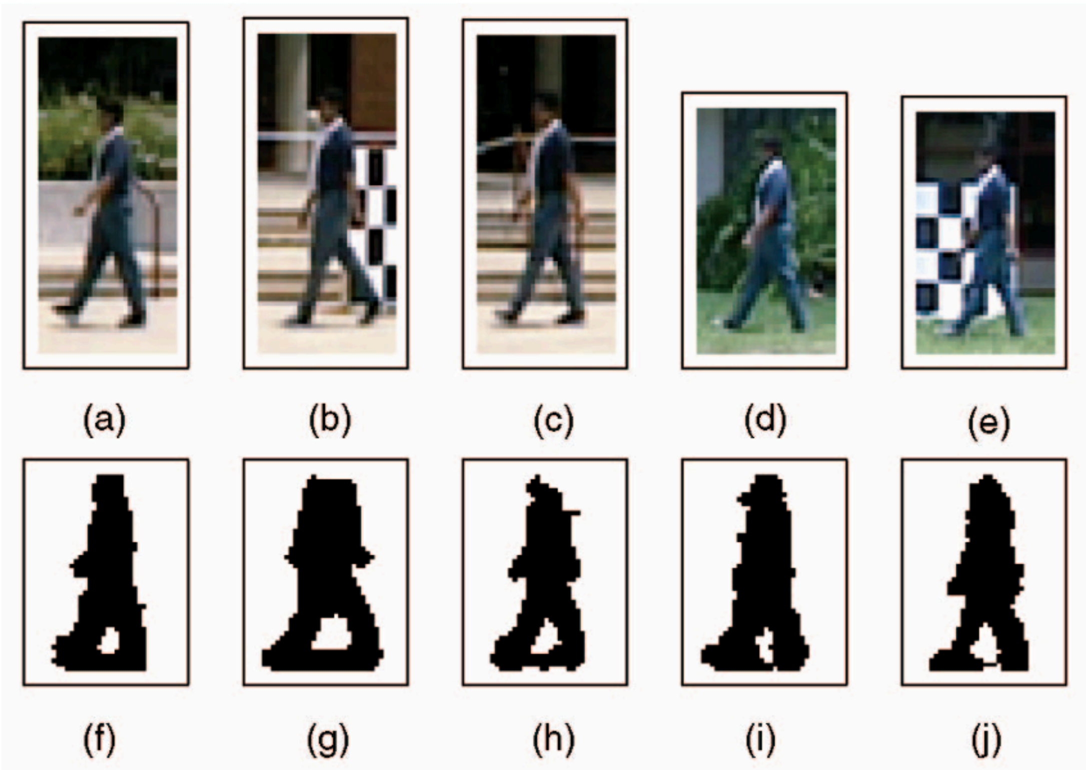


Figure 2.2: The bottom row ((f)-(j)) shows sample silhouette frames with a variety of segmentation errors. The raw image corresponding to each silhouette is shown on the top row ((a)-(e)[1].

The authors developed a four-part algorithm that involves defining bounding boxes around the moving person, extracting the silhouettes, computing the gait period, and performing spatial-temporal correlation to compare the gait sequences.

The algorithm splits each video frame into two parts to extract the silhouettes. Mahalanobis facing distance between each pixel and the background pixel is used to compute the area of the silhouette. Then expectation maximization (EM) algorithm estimates the posterior of each pixel based on its Mahalanobis distance.

For gait period detection, the authors counted the number of foreground pixels in the silhouette in each frame over time to estimate gait periodicity. They computed the median of the distances between minima, skipping every other minimum to acquire two estimates of the gait cycle. Finally, the gait period was estimated by the average of these two medians.

The authors also highlighted the challenges and limitations of their methods, such as variations in gait due to clothing, carrying objects, changes in surface and shoes, and variations in walking speed.

In another silhouette-based gait recognition study [13], the authors made use of wavelet descriptors (WD) as a feature extraction method. Features were derived by computing the wavelet transform of a silhouette contour at multiple resolutions. This enabled the authors to obtain features that are well-suited for representing digitized silhouettes. They also pointed out that wavelet descriptors are a significantly better shape representation for silhouettes than a complete Fourier descriptor set of the same dimensionality.

The authors applied this method to the Large Gait Database, compiled at the University of South Florida (NIST/USF) and found that the wavelet descriptors obtained for each person appeared to be unique, enabling reliable and successful gait recognition with recognition rates above 90 percent with a the k-nearest neighbour classifier.

2.1.2 3D-based gait recognition

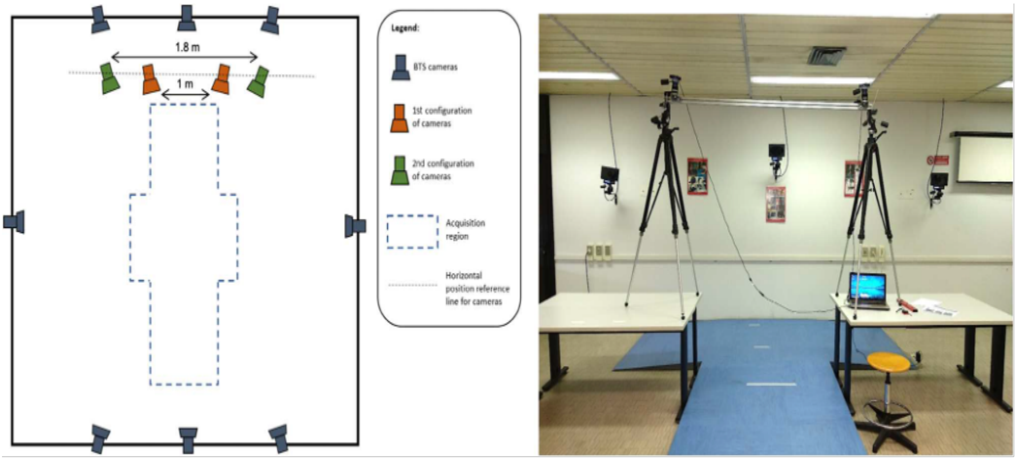
One of the popular approaches for gait recognition analysis is based on the 3D modelling of a person's gait. It is also a computer vision-based technique that involves reconstructing the 3D model of a person's walking motion. The technique involves several steps, including camera calibration, 3D reconstruction, and feature extraction;

Camera calibration is the process of determining the intrinsic and extrinsic parameters of a camera. The intrinsic parameters include the focal length, principal point, and distortion coefficients. The extrinsic parameters include the position and orientation of the camera relative to the world coordinate system.

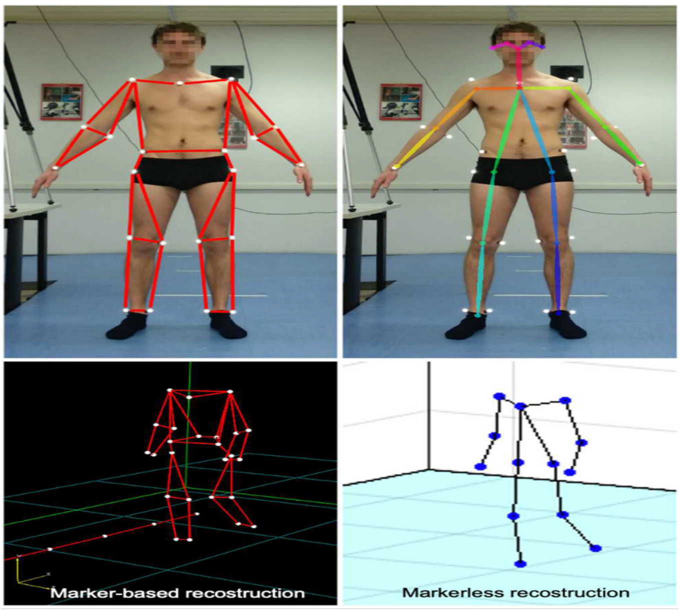
3D reconstruction involves using multiple calibrated cameras to reconstruct the 3D model of a person's walking motion. The 3D model can be reconstructed using different techniques, including stereo vision, structured light, and time-of-flight cameras.

Feature extraction involves representing the 3D model as a set of features that capture the walking pattern of an individual. The features can be extracted using different techniques, including Fourier descriptors, shape context, and shape distribution.

The study done in [14], describes a 3D tracking of human motion for gait analysis, utilizing a two-camera OpenPose-based markerless system.



(a) Laboratory setup, schematic (left) and pictorial (right) view[14].



(b) Stick diagrams as returned by the marker-based optical system (top, left) and OpenPose model (top, right); corresponding 3D reconstruction of the skeletal structures during walking (bottom)[14]

It involved, two healthy volunteers who performed a walking test at a comfortable walking speed within an instrumented gait analysis laboratory. The two participants were both male adults of the same age, with similar physical features such as height and body mass. They also wore minimal, close-fitting clothes.

In the experimental design, the authors considered three factors to attain the proper accuracy, namely the relative distance of the cameras, gait direction, and video resolution. It also involved procedures such as calibration of the stereoscopic system; acquisition of video recordings simultaneously with the reference marker-based system; video processing within OpenPose to extract the subject's skeleton; videos synchronization; triangulation of the skeletons in the two videos to obtain the 3D coordinates of the joints.

Finally, the authors in this study showed that the maximization of camera distance and video resolution allowed them to achieve the highest overall performances in tracking the kinematics and gait parameters of a single subject in a 3D space using two low-cost cameras and the OpenPose engine.

Additionally, they observed that the accuracy of markerless motion tracking depends on three factors that are the occlusions between body parts, the camera's position/orientation and video resolution. They consider the best combination of the considered factors being, a camera distance of 1.8m, maximum resolution, and no occlusions due to straight walking. The lowest error in 3D trajectory reconstruction was about 20mm, the lowest error in swing/stance time was 0.03 s and 1.23 cm in step length.

In another study[2], the authors use a multi-camera system consisting of eight cameras and necessary software tools to capture images. They assemble the captured images to develop a 3D model with full motion of the human gait pattern. Additionally, a group of 23 markers was connected to the different specified parts of the lower body to get the exact kinematics of six joints (hips, knees and ankles) of the subject.

The authors then obtained the markers' static position and dynamic movement and initialised a 3D model. They also used a force plate placed in the walking platform to procure kinetics data for movement and locomotion. They then tracked the six joints of the lower limb. A inverse kinematics and inverse dynamics library for human gait was developed and validated with analytical geometrical results.

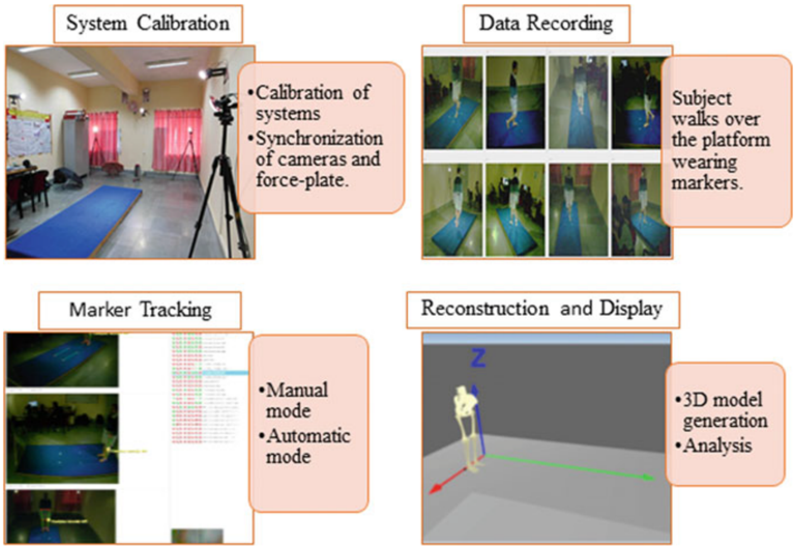


Figure 2.4: block diagram depicting the steps taken to develop the library for the human gait in[2].

They observed using the open-source sotware OpenSim3.3, that all the analytically calculated values of the joint moments closely resemble the optical marker vision-based system. Hence, the study concluded that the optical marker-based multi-camera system efficiently and accurately finds appropriate human gait patterns for biomechanics and biomedical studies.

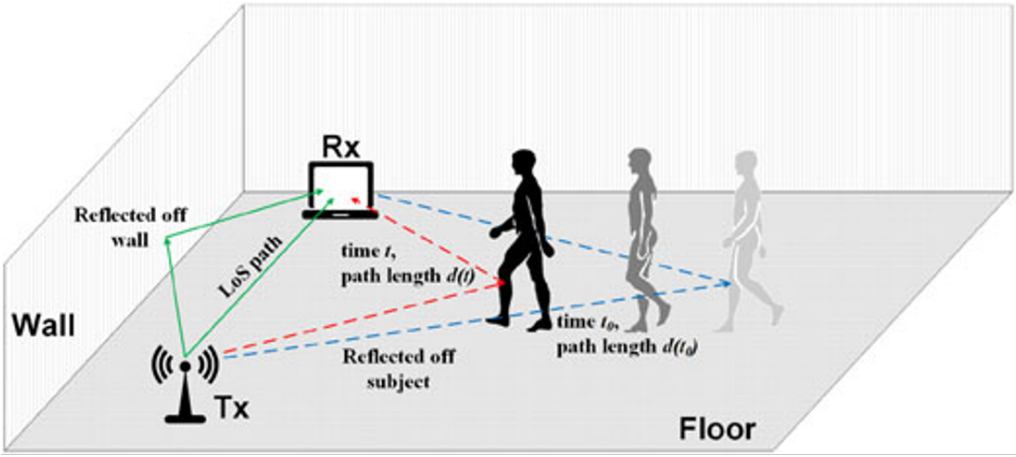
2.1.3 DEEP LEARNING-BASED GAIT RECOGNITION

Deep learning-based methods have become popular in gait recognition research, due to their ability to learn complex features from raw data. These approaches can typically be broken down into three main steps: data preprocessing, feature extraction, and classification. In data preprocessing, the gait sequences are normalized and segmented to extract the individual steps. In feature extraction, deep neural networks, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), are used to learn discriminative features from the gait sequences. Finally, the learned features are used to classify the gait sequences into different classes.

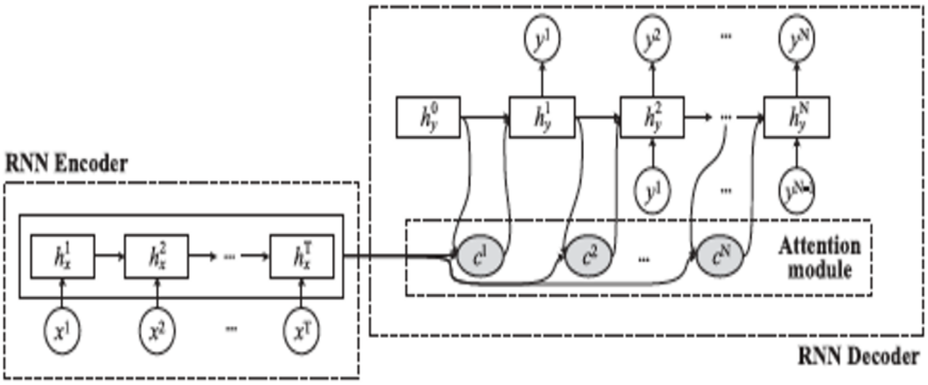
In the paper[15], the authors developed a specialized deep CNN architecture, four convolutional and subsampling layers. The model was then trained on the CASIA-B large gait database.Each convolutional layer has 8 filters leading to feature maps with eight channels.

The experimental results showed that the deep CNN model that was developed outperformed the other state-of-the-art gait recognition techniques in several cases. However, their CNN model could not learn very well under the scenario where the size of the gallery sets were very small and did not contain sufficient data.

In a more recent deep learning based-gait recognition, the authors[16] adopted an attention-based Recurrent Neural Network (RNN) encoder-decoder and created a cycle-independent



(a) Multipath effect of Wi-Fi signals in indoor environment.[16].



(b) Architecture of attention-based RNN encoder-decoder[16].

human gait recognition and walking direction estimation system, termed AGait, Using a Wifi system as a radar. AGait consists of four essential modules: CSI(Channel State Information) Collection, Raw CSI Processing, Walking Profile Generation, and Gait and Direction Sensing.

The wireless setup was made of two receivers and one transmitter placed in different positions, which were used to get more information about how people walk.

The research sought to understand how human walking activity exerts impact on Wi-Fi signals, and the multipath effect of wireless signals. Therefore, the attention scheme was used to allow AGait to learn to adaptively align with different critical clips of the CSI data. The approach was implemented in order to try to solve two different problems: identifying human gait appended with walking direction and estimating the speed of walking. This enabled the authors to design much more practical and interesting applications that could indicate whether an individual was approaching and going to access a service.

In AGait the CSI measurements from the two receivers are gathered together and refined to form an integrated walking profile. Then, the RNN encoder reads the walking profile and turns it into a hidden state sequence, which can be thought of as a representation of the profile's primary features. Then, given a specific (gait or direction) sensing task, the decoder computes a corresponding attention vector, which is a weighted sum of the hidden states with different attentions, and uses the vector to predict the target.

They implemented AGait on commercial Wi-Fi devices and evaluated its performance by conducting the walking experiment in three indoor scenarios, where 22 subjects were required to walk on 12 different paths in 8 different directions. Their experimental results demonstrated that AGait achieved average F1 scores of 97.32 to 89.77 percent for gait recognition from a group of 4 to 10 subjects, and an average F1 score of 97.41 percent for direction estimation among 8 directions.

2.1.4 WEARABLE SENSOR-BASED GAIT RECOGNITION

In this section, we will discuss some of the recent studies in wearable sensor-based gait recognition. Wearable sensor-based gait recognition has gained attention in recent years due to the widespread use of wearable devices and their potential in healthcare, rehabilitation, and sports. These methods typically involve four main steps: data acquisition, data preprocessing, feature extraction, and classification. In **data acquisition**, wearable sensors, such as accelerometers, gyroscopes, and magnetometers, are used to collect gait data from the subject. In **data preprocessing**, the raw sensor data is processed to remove noise and artifacts and to segment the gait cycle. In **feature extraction**, statistical and time-frequency features are extracted from the segmented gait data. Finally, the learned features are used to classify the gait patterns.

In this paper[17], the authors presents a gait-based authentication method that relies on accelerometric information collected at the user's wrist using a smartwatch. They decided not to use other sensors commonly available in smartwatches, such as gyroscopes or magnetic sensors, because their power consumption is significantly larger than the one of an accelerometer (in some cases by an order of magnitude). The proposed method distinguished the genuine user (i.e., the owner) from unauthorized users. The typical gait pattern of the genuine user was learned during an initial period of use; subsequently, anomalies in gait were automatically detected and used to infer if the current user was an impostor.

The approach uses genuine instances from unauthorized users' data based on semi-supervised anomaly detection. It considers a set of the genuine user's instances available to form a training set. Euclidean distance and Nearest-Neighbor (NN) analysis assign an anomaly score to each gait instance. This anomaly score, in turn, is compared against a threshold to classify a gait instance as normal (genuine user) or abnormal (unauthorized user). In particular, normalization exploits the average and standard deviation of the NN distances among instances in the training set, to produce the anomaly score (AS).

For example, if the average distance in the training set is 0.8 and the standard deviation is 0.2, then a gait instance with NN distance equal to 1.2 will have AS = 2. Higher anomaly score values indicate that the instance is more distant from the user's training data, thus, more likely to belong to an unauthorized user. The threshold to distinguish normal and abnormal instances (ASth) is selected by evaluating the trade-off between detecting anomalies and generating false positives. An experimental evaluation of the method was carried out with the help of 15 volunteers.Results show that the method can achieve an Equal Error Rate (EER) as low as 2.9

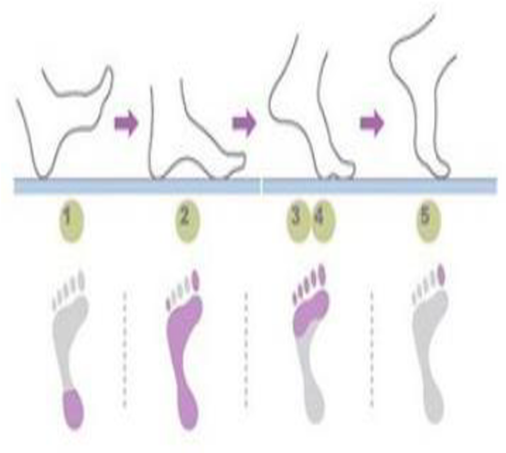
Also, they compared their study to a similar method executed on a pocket-worn device (POCKETmhv). They realized that hands are subjected to a much more significant amount of acceleration when compared to parts of the body that are near the center of mass. Thus, a smartphone in a user's pocket is generally exposed to significantly fewer movements than a smartwatch. However, the problems introduced by spurious hand movements were significantly reduced by adopting autocorrelation-based filters in the walking detection technique. In particular, they found that such filters reduce the EER by 7.8%.

The authors deliberately used information produced by a single sensor, the accelerometer. The results further confirmed that when the local reference frame is used, the wrist-worn device achieves similar results to a pocket-worn device. The average result obtained by POCKETmhv (97.3%) and WRISTxyz (97.4%) is almost identical.

In another study [3], the authors developed a shoe that analyzed the gait shape using a strain sensor. When walking wearing the developed shoes, the signal of the three sensors mounted on the shoes appeared differently. The authors included a Bluetooth-based communication system allowing gait to be automatically saved wirelessly based on the application. Users received feedback on various walking information in real-time, such as gait status and posture distortion through the recorded information.

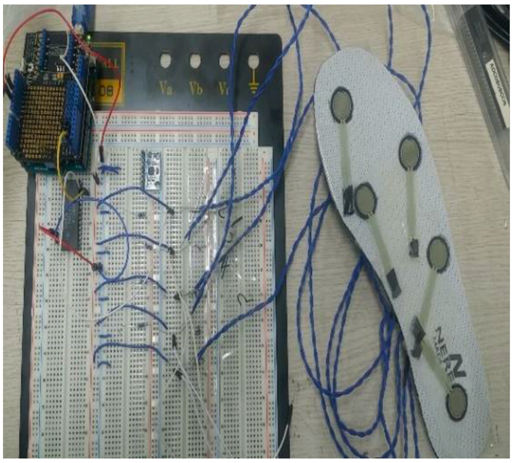
The developed wearable device (shoe) measured the wrong gait using the pressure sensor and a 3-axis acceleration sensor so that users could walk correctly. As a result, the correct order was checked when the sole with the pressure sensor touched the ground, and the gait shape was checked using the 3-axis acceleration sensor.

The authors distinguished a normal gait from an out-toeing or in-toeing gait by analyzing the user's gait status. The change value of the acceleration sensor was monitored through a waveform. It was confirmed as an output value of 1 and 0 by converting the analog value of the pressure sensor into a digital value. Users checked whether the soles of their feet correctly touched the ground via the designated color of neopixel through programming. This feature enabled the users to determine when they had taken the right step and when the pressure sensor was pressed, thereby increasing the accuracy of gait correction.

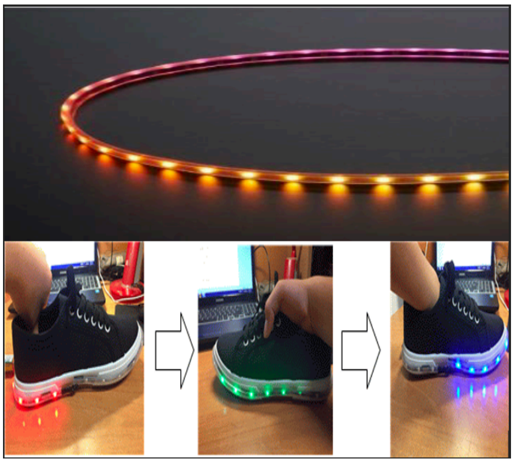


(a) Distribution of Load on the Floor.

(b) Pressure sensor arrangement.



(c) System Connection



(d) Neo-pixel Strip & operation of indicating out-toeing, normal stance, in-toeing.

Figure 2.6: System design & Operation of IoT-based sensor shoe [3].

The authors experimented by setting the normal gait at a 15 to 20-degree angle using the 3-axis acceleration sensor. When a user's gait angle was more significant than the standard angle, it was deemed out-toeing gait and deemed as in-toeing gait when narrower than the standard angle. The authors achieved a distinguishing recognition rate through the experiment with in-toeing gait at 56.25% and out-toeing gait at 81.25%.

2.2 STATE-OF-THE-ART ON CONTINUAL LEARNING

Continual learning, also known as lifelong learning or incremental learning, is a research area in machine learning that focuses on developing models that can continually learn from a stream of data, adapting to new tasks and changes in the environment over time without forgetting previously learned knowledge.

The main aim of continual learning is to build machine learning models that can learn incrementally, updating their parameters as they receive new data while retaining knowledge from previous tasks and minimizing the harmful effects of catastrophic forgetting i.e. models tend to forget how to solve old tasks when trained on new ones. Hence, it enables models created to continue to perform well on all previously seen tasks without being explicitly trained on them again, even after learning new tasks.

In addition, current research has addressed continual learning with longer task sequences and more examples, resulting in the proposal of methods for continual learning. Thus we will now examine significant works related to the proposed methods. In general, these methods can be categorized as follows[18]:

1. Replay methods
2. Regularization-based methods
3. Parameter isolation methods

2.2.1 Replay methods

This approach either maintains samples in raw format or produces pseudo-samples using a generative model. To prevent forgetting, these past task samples are replayed while learning a new task. These are either reused as model inputs for rehearsal or utilized to restrict the optimization of the current task loss to prevent interference from the prior task.

Replay methods have limited scalability over the number of classes, necessitating extra computation and storage of raw input samples. While restoring memory reduces memory use, it diminishes exemplar sets' ability to represent the original distribution. Also, keeping these unprocessed input samples may raise privacy concerns.

In this study[19], the authors formalized the scenario of continual learning by defining training and evaluation protocols to assess the quality of models in terms of their accuracy, as well as their ability to transfer knowledge forward and backward between tasks. Their aim is to tackle the problem of continual learning, where the model observes, once and one by one, examples concerning a sequence of tasks. They also introduced a model for continual learning, called Gradient Episodic Memory (GEM) that leverages episodic memory to avoid forgetting and favor positive backward transfer. Their experiments demonstrated the competitive performance of GEM against the state-of-the-art. GEM however misses three key features.

Firstly, GEM does not leverage structured task descriptors, which may be exploited to obtain positive forward transfer (zero-shot learning). Second, advanced memory management was not investigated (such as building core sets of tasks). Third, each GEM iteration requires one backward pass per task, increasing computation time.

In another work utilizing replay method, the authors[20] introduced iCaRL (incremental classifier and representation learning), a practical strategy for simultaneously learning classifiers and a feature representation in the class-incremental setting as a result of the shortcomings of existing approaches. iCaRL's three main components are:

1. a nearest-mean-of-exemplars classifier that is robust against changes in the data representation while needing to store only a small number of exemplars per class,
2. a herding-based step for prioritized exemplar selection, and
3. a representation learning step that uses the exemplars in combination with distillation to avoid catastrophic forgetting.

Their experiments on CIFAR-100 and ImageNet ILSVRC2012 data showed that iCaRL is able to learn incrementally over a long period of time whereas other methods fail quickly.

The main reason for iCaRL's strong classification results was its use of exemplar images. The authors hypothesized that also other architectures will benefit from using a combination of network parameters and exemplars, especially given the fact that many thousands of images can be stored (in compressed form) with memory requirements comparable to the sizes of current deep networks.

2.2.2 Regularization-based methods

The line of work employing this method avoids keeping raw inputs, prioritizes privacy, and reduces memory needs. Instead, an additional regularization term is added to the loss function, consolidating prior knowledge while learning new data. These methods can be further classified as data-focused and prior-focused methods.

In this prior-focused study, the authors [21] proved that task-specific synaptic consolidation provides a novel answer to the challenge of artificial intelligence's ongoing learning. They created a method similar to synaptic consolidation for artificial neural networks, known as elastic weight consolidation (EWC). This approach slows down learning on specific weights based on how significant they are to previously observed tasks. In stark contrast to earlier deep-learning algorithms, their work demonstrated how EWC may be utilized in supervised learning and reinforcement learning problems to teach several tasks sequentially without forgetting older ones.

In addition, their research underlined that the consolidation of high-precision weights permits continuous learning over extended periods. At each synapse, EWC requires the storage of three values: the weight, its variance, and its mean. Surprisingly, synapses in the brain also hold many pieces of information. They suggested that, for instance, the state of short-term plasticity may contain information about the variance. The weight associated with the early phase of plasticity would encode the present synaptic strength, whereas the weight associated with the late phase or the consolidated phase could convey the average weight.

In another paper utilizing data-focused approach[22], the authors expounded on their previous work, Learning without Forgetting (LwF). Using only examples for the new task, they optimized both to obtain high accuracy on the new task and to preserve the response on the existing tasks. The main advantages of LwF are:

Classification performance: Learning without Forgetting outperforms feature extraction and, more surprisingly, fine-tuning on the new task while doing better than models fine-tuned on the old task.

Computational efficiency: The training time for the proposed approach is faster than the joint training approach. Moreover, the test time for a single sample across all tasks is considerably faster when compared to using multiple fine-tuned networks for individual tasks.

Simplicity in deployment: Once a task is learned, the training data does not need to be retained or reapplied to preserve performance in the adapting network.

They further discussed in their research that: LwF operates on distinct tasks. Like many multitask learning methods, it cannot properly deal with domains that are continually changing on a spectrum (e.g., old tasks being classified from a top-down view, and the new task being classified from views of unknown angles); the tasks must be enumerated. In addition, LwF requires each sample to be accompanied by information about which task it belongs to, and this information is needed for both training and testing.

Secondly, in contrast to methods such as Never Ending Learning[23], LwF requires all new task training data to be present before computing their old task responses. This can't be applied if the data comes in a stream or if the model needs to be trained incrementally.

Thirdly, the ability of LwF to incrementally learn new tasks is limited, as the performance of old tasks gradually drops.

2.2.3 Parameter isolation methods

This category assigns distinct model parameters to each task to prevent forgetting. Since the architecture has no size limits, one can create new branches for the new task while freezing existing task parameters.

The authors of the research article[24] developed an approach that employs weight-based pruning approaches to free up redundant parameters across all layers of a deep network after it has been trained for a task, with minimal loss in accuracy. Consequently, the freed-up parameters are updated to learn a new task while the remaining parameters remain unchanged. Utilizing the task-specific parameter masks created by pruning, their models retained the same level of accuracy even after adding several tasks. They had very little storage expense per additional task.

They proved the effectiveness of their strategy on numerous tasks for which high-level feature transfer performs poorly, suggesting the necessity to change network parameters at all layers. It exceeds previous work in robustness against catastrophic forgetting and the quantity and complexity of the new task. It was effective for the comparatively "roomy" VGG-16 architecture and more compact, parameter-efficient networks such as ResNets and DenseNets.

Subsequently, the authors of the research paper[25] introduced HAT, a task-based hard attention mechanism that, by focusing on task embedding, can protect the information of previous tasks while learning new tasks. The hard attention mechanism is lightweight because it adds a small fraction of weights to the base network. It was trained with the primary model, with negligible overhead backpropagation and minibatch stochastic gradient descent (SGD).

They demonstrated the approach's effectiveness in controlling catastrophic forgetting in the image classification context by running a series of experiments with multiple data sets and state-of-the-art approaches. Thus they evaluated HAT in the context of image classification, using a high-standard evaluation protocol by considering random sequences of 8 publicly-available data sets. They demonstrated favorable results in 4 different experimental setups, cutting current rates by 45 to 80%. They also showed robustness concerning hyperparameters and illustrated several monitoring capabilities.

3
Smart Glasses

This chapter discusses our approach to gait analysis using smart glasses and the preprocessing mechanisms used to obtain primary data. Chapter 5 will cover the experimental data acquisition and results. While previous studies have used a multi-modal approach that combines inertial data from sensors and image data from cameras, we only used inertial sensor data from smart glasses for gait analysis.

3.1 Smart Glasses

To begin, we provide a detailed description of the smart glasses we used in our study. The smart glasses have inertial and other embedded sensors and Bluetooth capability for gait activity detection and data acquisition. Unlike other devices, this glasses did not have built-in storage for data acquisition. Therefore, we paired the smart glasses with a PC via Bluetooth to collect and record data on detected gait activity. Figure 2.1 illustrates the smart glasses and their features relevant to this study, which we will further discuss.

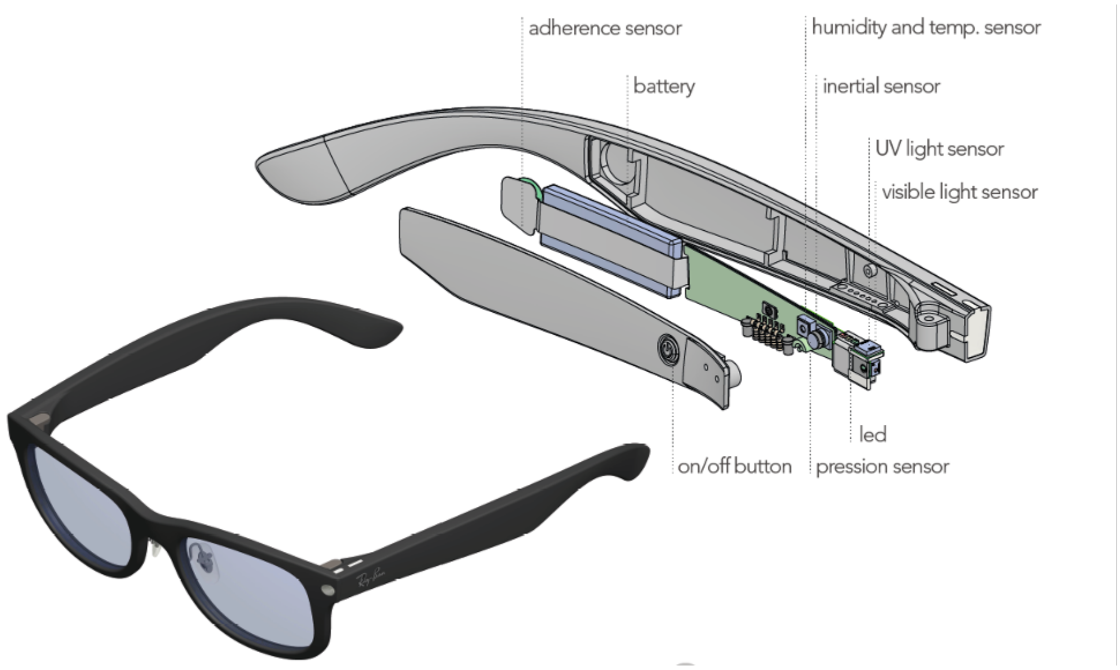


Figure 3.1: Labeled component of the Smart Glasses.

The smart glasses used has specific operational features that were prevalent during the research. These features include:

1. **Battery** - when fully charged, the battery lasts approximately one hour.
2. **LED** - The LED was color-coded and signaled other features' activation.
3. **Switch Button** - The on/off switch button activates or deactivates the smart glasses.
4. **Bluetooth** - connect smart glasses to other smart devices, such as PCs, to collect data.

FEATURE	EVENT	TRIGGER	LED SIGNAL
Battery	charging (not full)	charger plugged	RED, continuous
	charging (full)	charger plugged	GREEN, continuous
	full	100% level	GREEN, continuous long blink
	running low	< 20%	RED, continuous long blink for 3 secs
	dead	0%	RED, short blink
Switch Button	turned ON	long pressed	RED, for few secs then turns GREEN
	turned OFF	long pressed	RED, for few secs then OFF
Bluetooth	connected	device paired	BLUE continuous long blink
	disconnected	device unpaired	GREEN or RED, depending battery's level

Table 3.1: Features functionality & LED Behavior.

Additionally, table 3.1, displays the LED behavior when these features are in operation:

The LED blinks red when the smart glasses are turned on or off. It then blinks green when it has a full battery level after being turned on. When Bluetooth is connected to a smart device, the LED blinks blue. During gait activity detection and data transfer, the LED blinks blue.

3.2 Data Gathering

This section describes the data-gathering process for gait analysis in this study. The smart glasses used in this study is equipped with inertial and other embedded sensors, allowing automated gait data collection during walking activity sessions. Thus we introduce this part with discussions on embedded sensors for gait detection and data acquisition.

3.2.1 Embedded Sensors

Embedded sensors are devices integrated into a system or object to detect, measure, and transmit data. In the context of gait analysis, embedded sensors can detect and collect data on the body's movements during walking. These sensors usually include accelerometers, gyroscopes, and magnetometers, which can measure acceleration, angular velocity, and magnetic field strength. The measurements from these sensors can provide information on the kinematics and kinetics of human motion for gait analysis.

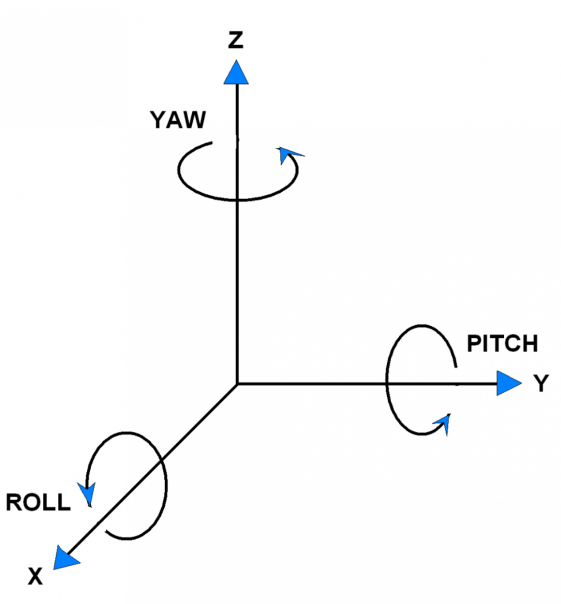


Figure 3.2: Roll, pitch, and yaw angles

Accelerometers present in inertial sensors for gait identification measure the correct acceleration, or g-force, which is the acceleration relative to freefall. Individuals and objects undergo acceleration. The piezoelectric effect is the most common form of accelerometer and uses microscopic crystal structures that become stressed due to accelerative forces. These crystals create a voltage from the stress, and the accelerometer interprets the voltage to determine velocity and orientation.

At a particular moment, data consists of the acceleration values recorded for the x, y, and z axes. Like the majority of sensors, accelerometers are susceptible to calibration issues. Linearity is one of the primary features of an accelerometer, which means that its response is precisely proportionate to the actual acceleration being measured. This trait may be utilized in various ways, including data normalization, which is explored in further detail in the following part of this chapter.

Gyroscopes monitor the triplet of rotation angles along three axes, making them another class of inertial sensors. Accelerometers and gyroscopes are the most often utilized sensors for gait detection due to their integration into a single micro electro-mechanical systems (MEMS) component.

Magnetometers are sometimes included with MEMS. This sensor monitors the magnetic field strength. It is not utilized in gait identification since it can only detect the walk's direction. This sensor is often employed for geolocation instead.

Barometer is a type of pressure sensor that measures atmospheric pressure. In wearable sensor-based gait recognition, a barometer sensor is used to measure changes in altitude as a person walks or runs. By detecting changes in atmospheric pressure, the barometer provides information about the vertical displacement of the body during movement.

The data from the barometer can be combined with data from other sensors, such as accelerometers and gyroscopes, to build a more comprehensive picture of a person's gait and is useful for applications such as health monitoring, fitness tracking, and security systems.

For example, when a person takes a step, their body moves up and down as they transfer their weight from one foot to another. This movement causes a change in atmospheric pressure, which the barometer sensor can detect. By analyzing these pressure changes over time, algorithms can be developed to recognize a person's gait pattern and identify them based on their unique walking style.

In addition to the sensors discussed above, several other types of sensors can be used in gait analysis and recognition. These sensors can be integrated into wearable devices or implanted in shoes or other body parts.

Force sensors like capacitive, resistive piezoelectric and piezoresistive sensors are other types of sensors that measure the ground reaction force under the foot and return a current or voltage proportional to the pressure measured. They can be used to determine how a person distributes their weight while walking or running, which can be an important indicator of the gait pattern.

Pressure sensors are another type of sensor that measures the force applied to the sensor without considering its spatial components. They can be used to measure the pressure distribution on foot during movement, which can be useful for detecting abnormal gait patterns or identifying individuals based on their unique pressure patterns.

Goniometers measure joint angles for ankles, knees, hips, and metatarsals. They can be used to track the movement of joints during gait and can provide valuable information about gait patterns.

Ultrasonic sensors analyze short steps, stride lengths, and the distance between the feet. They can detect changes in gait patterns, such as those associated with certain injuries or conditions.

Finally, **electromyography (EMG)** measures the electrical manifestation of muscle contraction. EMG signals can be obtained either non-invasively with surface electrodes or invasively with wire/needle electrodes. EMG can measure different gait features, such as kinematic plots of joint angular motion, and has recently been used for gait recognition.

Using embedded sensors in gait analysis offers several advantages over traditional methods, such as motion capture systems requiring a laboratory setting and specialized equipment. They utilized measurements in natural environments, allowing for more realistic data collection. They are also lightweight, portable, and non-invasive, making them suitable for long-term monitoring of gait patterns in real-world settings.

In this study, the smart glasses' inertial and other embedded sensors automatically collect gait data during walking activity sessions. Using these sensors allows for continuous data collection without needing a laboratory setting or specialized equipment, making it easier and more convenient for subjects to participate in the study. This study combines raw data from the accelerometer, gyroscope, and barometer to recognize subjects' gait patterns accurately.

3.2.2 Raw Dataset

The study includes data from eight healthy subjects, four males and four females aged between 25 and 29. The heights of male subjects ranged from 1.72 to 1.87 meters, and female subjects ranged from 1.50 to 1.73 meters. Unlike other gait recognition studies that collect data in a lab or confined environment, we recorded the IMU data of each subject wearing smart glasses while walking around the city center of Padova for a minimum of 30 minutes. This approach was intentionally taken to acquire data that reflect the subjects' daily normal walking patterns. The IMU data from the smart glasses' inertial sensor was transmitted in real-time to a Bluetooth-paired PC, where the raw (unprocessed) data was stored in CSV file format at the end of each walking session with each subject.

A sample of raw data acquired with the smart glasses from a subject is presented in Table 3.2. Based on the data provided, we can see that the smart glasses sensors measured different aspects of the wearer's movements, orientation, and environment. Specifically, the measurements include:

ACCx	ACCy	ACCz	GYRx	GYRy	GYRz	ROLL	PITCH	YAW	PRESSURE
-141	-954	-3	23.9	-23.8	-62.0	4.5	15.5	144.0	101819
-155	-920	-36	21.2	-26.3	-70.4	3.7	14.1	143.6	101816
-165	-885	-79	18.0	-29.1	-77.3	3.2	12.9	143.2	101818

Table 3.2: Sample data of raw gait signal from a subject.

1. Acceleration in the X, Y, and Z axes (accx, accy, accz), where acceleration is the rate of velocity change with respect to time.
2. Angular velocity around the X, Y, and Z axes (gyrx, gyry, gyrz), where angular velocity is the rate of change of angular displacement with respect to time.
3. Rotation around the X, Y, and Z axes (roll, pitch, yaw) are measures of the orientation of the sensor.
4. Atmospheric pressure (pressure), where atmospheric pressure is the force exerted by the weight of the air in the Earth's atmosphere.

Correct, preprocessing is necessary to clean and filter the raw data to remove noise and artifacts and prepare it for further analysis. It can include filtering, normalization, resampling, segmentation, and feature extraction. Preprocessing helps to improve the accuracy and reliability of the subsequent analysis and modeling tasks.

4
Methods

In this chapter, we discuss the preprocessing methods for processing the raw gait signals and extracting their gait features. More also, in order to identify each subject based on the extracted features, we trained a machine learning classifier to this effect.

4.1 Data Preprocessing

Data preprocessing is essential in any gait recognition system since it eliminates noise and normalizes the data for precise feature extraction and analysis. This part will present the data preparation techniques generally used as good practice for sensor-based gait detection, including smart glasses.

Data filtering. The initial step in data preparation is to filter the raw data to eliminate noise. The filtering method utilized will depend on the sensor type and noise characteristics. For instance, a low-pass filter can eliminate the high-frequency components if the data is impacted by high-frequency noise.

Segmentation. Segmentation is the process of breaking down gait data into distinct steps. Each step requires the calculation of gait characteristics such as step length and stride time. Many segmentation techniques exist, including peak detection and thresholding. This technique may use the threshold-based method to specify an acceleration or gyroscope data threshold. A step is detected when the acceleration or gyroscope data passes the threshold.

Feature extraction. Following segmentation, the gait characteristics are extracted from the preprocessed data. These characteristics encompass both temporal and spatial domain characteristics. The formal include step length, stride length, and rhythm, whereas the latter include joint angles and foot contact locations.

Normalization. Normalization is utilized to eliminate the impacts of inter-participant variation in gait data. Individual changes in walking style might impair the categorization accuracy, making this step necessary. Min-max normalization and z-score normalization are two typical normalizing techniques. Min-max normalization scales the data to a specific range, often [0,1], whereas Z-score normalization scales the data to a distribution with zero mean and unit variance.

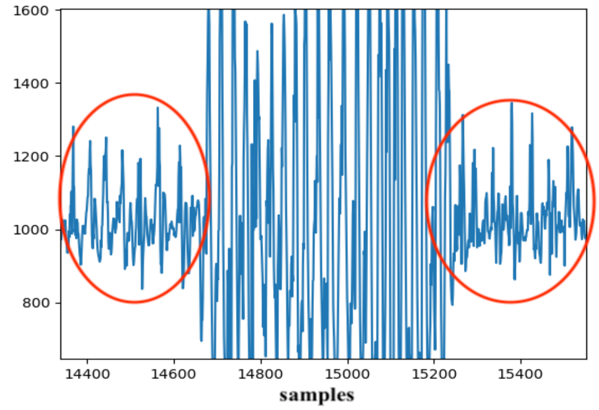
Feature selection. Feature selection is the process of choosing a subset of the most informative characteristics for classification. This step can increase classification accuracy by reducing the dimension of the feature space. Various approaches can be used to select features, including mutual information, correlation-based feature selection, and sequential forward/backward selection. Practically, these operations are performed by exploiting the powerful API scikit-learn.

4.2 Processing Structure

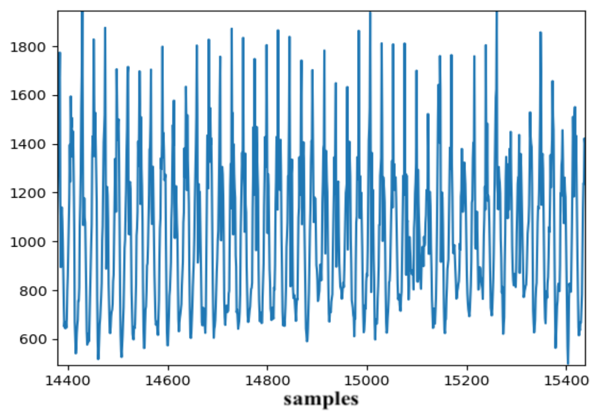
The above-discussed preprocessing methods are the general process raw inertial signal undergoes. However, the preprocessing performed for this study includes (1) filtering to remove unwanted motion artifacts, (2) Walking cycle extraction, and (3) Normalization of the walking cycle (accelerometer and gyroscope data).

4.2.1 Filtering

As discussed in the previous chapter, we considered only the accelerometer and gyroscope signals to preprocess the gait signal. Most signal strength is centered at low frequencies, primarily below 40 Hz. The raw inertial signals were acquired at a 100 to 200 Hz sampling rate, sufficient for capturing most of the walking signal. We use a cubic Spline interpolation to represent the input data with evenly spaced points (200 points/second) due to the non-uniform sampling of raw inertial signals. Hence, a low-pass Finite Impulse Response (FIR) filter with a cutoff frequency of 40 Hz is applied for de-noising and minimizing motion artifacts that may appear at undesired frequencies. Thus, the filter removes only noise while retaining the user's movement (discriminative) information. We represent the filtered and interpolated acceleration and gyroscope time series along axis x, as $a_x(i)$ and, $g_x(i)$ respectively, where i = 1, 2, . . . is the sample number. This notation also holds for axes y and z in further discussions. In Figure 4.1, we shown a sample of the gait signal before and after de-noising during data preprocessing



(a) Raw gait signal with noise.



(b) Signal after denoising.

Figure 4.1: Raw gait before and after preprocessing.

4.2.2 Walking cycle extraction

Here, we adopt a template-based and iterative algorithm that extracts walking cycles using only the accelerometer's magnitude signal. The computation for each sample's acceleration magnitude is

$$a_{mag}(i) = (a_x(i)^2 + a_y(i)^2 + a_z(i)^2)^{1/2} \tag{4.1}$$

A reference point must be located in $a_{mag}(i)$, therefore we begin by passing $a_{mag}(i)$ through a low-pass filter with a cutoff frequency of 3 Hz to determine a gait template, in accordance to [26]. Consequently, we detect the initial minimum of this filtered signal, corresponding to the heel strike [27], and we refer to the related index as i. Then, this minimum is refined by examining the original signal,, $a_{mag}(i)$ an interval centered on i covering one second of data, and picking the minimum value of $a_{mag}(i)$ in this interval. This determines a new index i for which $a_{mag}(i)$ is the local minimum. This minimum is shown by a red vertical (dashed-dotted) line. As a second step, we select a one-second frame centered on i, depicted 4.2, symbolized by two blue (dashed) vertical lines. Now, the samples of $a_{mag}(i)$ falling between the two blue lines create the first gait template (T), where $|T| = N_s$ samples, where N_s is the number of samples measured in one second.

The retrieved template is refined repeatedly and utilized concurrently to identify the subjects' subsequent walking cycles. In [26], the template is kept unaltered until minima can no longer be identified, at which point a new template is obtained. As this is required by the subsequent feature extraction and classification algorithms, a normalization phase is necessary to represent all cycles through the same number of points N.

4.2.3 Normalization

Each gait cycle varies depending on walking speed and stride length, resulting in variable-sized acceleration and gyroscope vectors in the orientation-invariant coordinate system. However, to perform feature extraction and classification, these vectors need to be of a fixed size, denoted by N. We use Spline interpolation to represent all walking cycles as N = 200-sample vectors. This value of N was chosen to avoid aliasing, assuming a maximum cycle period of 2 seconds and a signal bandwidth of 40 Hz. Amplitude normalization ensures that the resulting vectors have a mean of zero and a variance of one, improving training and classification performance. This yields eight N-dimensional vectors for each walking cycle, which are then used in the subsequent feature extraction and classification methods.

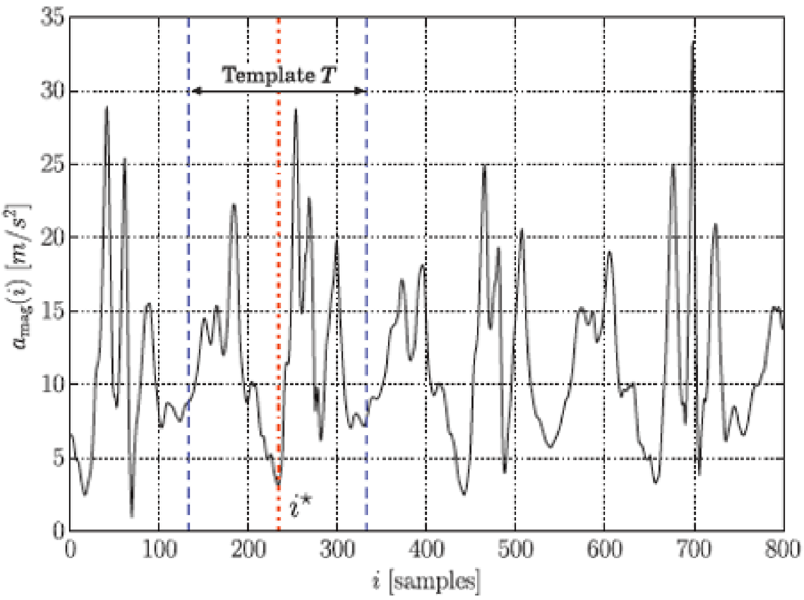


Figure 4.2: Template extraction using the accelerometer magnitude $a_{mag}(i)$.

Finally, the preprocessed data is fed into a machine-learning algorithm for training and classification. The classification's accuracy depends on the quality of the preprocessed data and the effectiveness of the feature extraction process.

4.3 Convolutional neural network

Convolutional neural networks (CNNs) are deep neural networks that are designed for feed-forward processing. CNNs contain one or more convolutional layers, each of which has a number of kernels. These kernels consist of weights that are convolved with the input data, and the same set of weights is applied to all input data, with the convolution process being shifted throughout the input span. The connectivity structure of CNNs is sparse, with shared weights being used repeatedly and each kernel operating on a small amount of the input signal. This significantly reduces computing complexity compared to fully linked feed-forward neural networks.

4.3.1 Dataset

The data for this study was collected using smart glasses sensors worn by eight subjects. The gait data acquired after preprocessing for training and testing are shown in Table 4.1

MALE	FEMALE
Subject A-Number of steps: 3709	Subject B-Number of steps: 1013
Subject C-Number of steps: 3605	Subject D-Number of steps: 4904
Subject E-Number of steps: 2135	Subject F-Number of steps: 4154
Subject G-Number of steps: 3732	Subject H-Number of steps: 3650

Table 4.1: Preprocessed Gait data from Subjects.

Though all subjects walked for 30 minutes, some had several noise in their collected IMU signal data. This resulted from several breaks or pauses during the 30-minute walking period. Hence, after cleaning the data during the preprocessing stage, Subject C acquired the least gait cycle (steps), followed by Subject E.

Therefore, our dataset is composed of a large number of standard gait cycles. The dataset comprises eight acquisitions divided into a training and validation set and a testing set, amounting to about 26,900 cycles. The division of the samples between the training and validation data set and test data set is performed by selecting the acquisitions with a (75%/25%) split.

4.3.2 CNN architecture

CNNs have been demonstrated to be good image feature extractors. Here we discuss their efficacy for motion data. The convolutional layers of CNN are responsible for dimensionality reduction (or feature extraction), whereas the fully connected layer serves as a classifier. Each walking cycle's accelerometer and gyroscope data is handled using the techniques described in the previous section. All vectors are normalized to N samples, the input matrix for a generic walking cycle.

In our study, we deployed a multi-class CNN for training to classify and identify the eight individuals based on their extracted gait features. We also separately trained the network to classify the individuals into male and female.

Our CNN architecture consists of multiple one-dimensional (1D) convolutional layers. It has explicitly six layers: four 1D convolutional layers, a Flatten layer, a fully connected (FC) layer, and an instance variable LRelu which is the LeakyReLU activation function. We randomly initialized four convolutional filters in every convolutional layer. The output of each convolutional layer is passed through a dropout layer before being fed to the next layer.

This 1D CNN architecture is usually used for various types of 1-D signals, such as time series (e.g., inertial signals), audio, and speech signals. It uses the 1-D convolutional layers, decreases kernel size, and strides through the layers, which can help extract more complex features and reduce the spatial dimensions of the data.

The implementation of this network is performed by exploiting the Pytorch, a machine learning framework based on the Torch library, used for applications such as computer vision and natural language processing and Python programming language.

Training

Some additional parameters must be established prior to the training phase. One of these hyper-parameters is initializing every neuron's weights and biases. For this goal, we used two distinct approaches. We initialized the biases in each convolutional layer and the fully connected layer with different values. In contrast, initial weights are generated by randomly selecting samples from a truncated normal distribution with zero-mean standard deviation.This decision is made to preserve the variance of network activations, hence accelerating the convergence of the weights to the optimum value that minimizes the loss function.

The Adam optimizer is the training procedure optimizer [28], which provides the best performance. The batch size and the number of training epochs are the last parameters specified, with batch size = 500 samples and epochs = 30, respectively.

Classification

After training, the model is used to classify the test set. The trained model was used to classify the gaits from the eight individuals. The gaits of these individuals were registered and labeled during the training phase. Furthermore, we labeled the gait data as either 0 or 1 for binary classification. The gaits of the individuals were classified as 0 for female gaits and 1 for male gaits. The predicted labels are compared to the true labels to compute the confusion matrix and other performance metrics. The accuracy of the classification prediction performed by the model is investigated for a complete analysis. The results are discussed in Chapter 6.

5

Continual Learning on Gait Recognition

In this chapter, we will explore the use of rehearsal strategies for incremental task learning in gait analysis, using gait data extracted from smart glasses' inertial sensors. Continual learning techniques such rehearsal method can be applied to gait analysis to develop models that can learn incrementally and adapt to changes in the environment over time. It enables the learning model to adapt to new data and tasks without forgetting the previously learned knowledge.

5.1 Rehearsal Method

The rehearsal method is a continual learning technique that mitigates the catastrophic forgetting problem, which occurs when a learning model is trained on new data and forgets the previously learned knowledge. The rehearsal method involves storing a small subset of previously learned data and retraining the model on this data alongside new data to help retain the past knowledge as seen in 5.1.

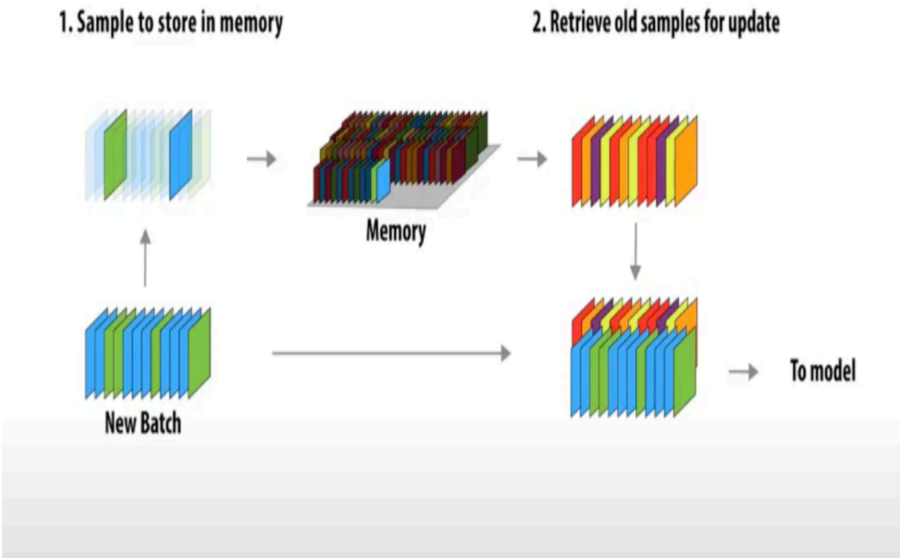


Figure 5.1: The Rehearsal Method: Continual learning Process

5.1.1 Rehearsal Process

The essential part of implementing rehearsal method is the memory buffer that stores a small subset of the previously learned data. When new data is collected, a portion of it replaces the oldest data in the buffer, ensuring that the buffer always contains a representative sample of the most recent data. The buffer's size is a crucial factor, as it should be large enough to contain a diverse set of samples, but not too large to cause storage and computational overheads.

There are several strategies for managing the rehearsal buffer, such as:

1. Random sampling. Selecting random samples from the new data to replace the oldest samples in the buffer.
2. Importance sampling. Selecting samples based on their importance, such as the prediction error, ensuring that the most informative samples are retained in the buffer.
3. Curriculum sampling. Organizing the data in a curriculum, where the model is initially trained on easier samples and gradually exposed to more challenging samples

In our study, we used importance sampling, in that we selected samples from training data and labels of these data after training each task training. The first task contained data samples of 3 subjects and each of these subject data had different labels(0,1,2). The first task data is used to the train the learning model, which in our case is 1D CNN model.

When design the memory buffer to select a percentage of trained data and it labels and keep in memory. To enable easy and effective computational process we saved the few samples.

This memory is then loaded and then added the new task containing new dataset with new labels. Thus aside the old labels this new task data has labels(0,1,2,3,4,5). Before training of the task data the memory data of the previous task is then added to current task data this enables model to learning and classify the new data as learning remember the labels or features of the old data for better classification.

The same process done for the final task that data which contained to new classes, hence the final task contained data with label (0,1,2,3,4,5,6,7).

5.1.2 Evaluation

In our study, we utilized memory buffer library ,continnum by [29] in implementing the rehearsal process discussed in this chapter. Evaluation is essential to measure the model's performance after its training process. Therefore, we evaluated the performance as new gait data is added by adjusting the buffer size between 10 to 20% during the training process to test the model's recognition accuracy.

The result of rehearsal method employed in our study is discussed in the next chapter 6.

6
Results

In this chapter, the gait recognition results, both in the standard and continual setups, will be presented. These include the analysis of the algorithms, evaluation of results, and the visualization of the networks' outputs.

6.1 Gait Recognition

The preprocessed gait data was then used for training and testing the CNN model for recognition and classification. The gait recognition and classification results using the CNN model were evaluated using the accuracy and confusion matrix.

To evaluate the classification prediction performance of the deep learning algorithm, we trained and tested the CNN model in two different classification modes:

1. Person Identification: Training, test, and validation sets are composed of individual labeling of each subject.Thus we assigned different labels (0 - 7) to each of the eight subjects. In other words, data from all subjects are represented in a multi-classification training and test phase.

2. Sex Identification: Training, test, and validation sets are composed of either 0 or 1 labels assigned to each subject, in a binary classification. Male subjects were assigned 0 as their label and female subjects were assigned 1.

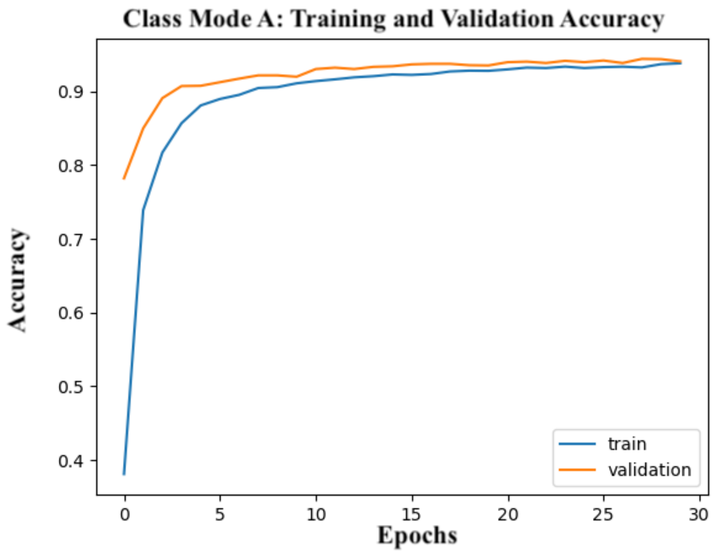


Figure 6.1: Training and Validation Accuracy Person Identification

The results of the classification prediction performed by the model are presented in the next paragraphs. Due to dropout techniques in the training phase and the low number of validation samples, we see in Person Identification (Figure 6.1) that the validation accuracy is higher than the training accuracy. This behavior can also be seen in Sex Identification (Figure 6.2).

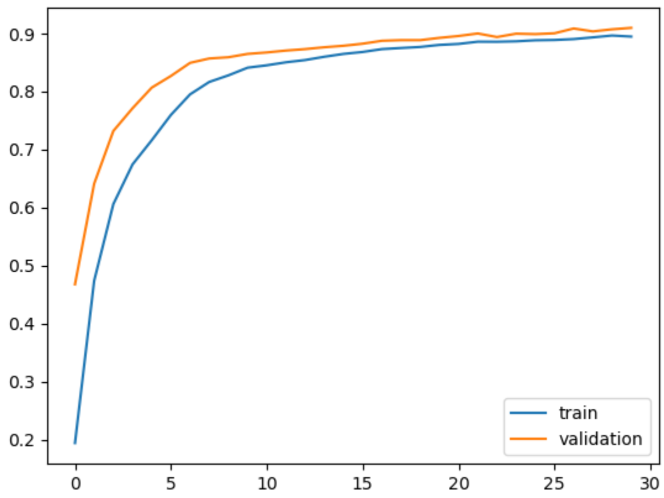


Figure 6.2: Training and Validation Accuracy for Sex Identification

MALE	FEMALE
Subject A (Label 0): 0.98	Subject B (Label 2): 0.88
Subject C (Label 1): 0.98	Subject D (Label 3): 0.93
Subject E (Label 4): 0.82	Subject F (Label 5): 0.95
Subject G (Label 6): 0.97	Subject B (Label 7): 0.95

Table 6.1: Tabular Representation of Person Identification.

The model achieved an overall average accuracy of 93% in classifying the gait data of the 8 subjects (class mode A). The confusion matrix showed that the model performed well in classifying the gait data for each subject, with only a few misclassifications. This is illustrated in Figure 6.3 and better outlined, showing the individual classification accuracies in Table 6.1. More also, the model achieved an overall average accuracy of 91% in classifying the gait data of the subjects as male and female (class mode B). This is also illustrated in Figure 6.4 and better outlined, showing the individual classification accuracies in Table 6.2.

MALE(0)	FEMALE(1)
0.97	0.97

Table 6.2: Tabular Representation of Sex Identification.

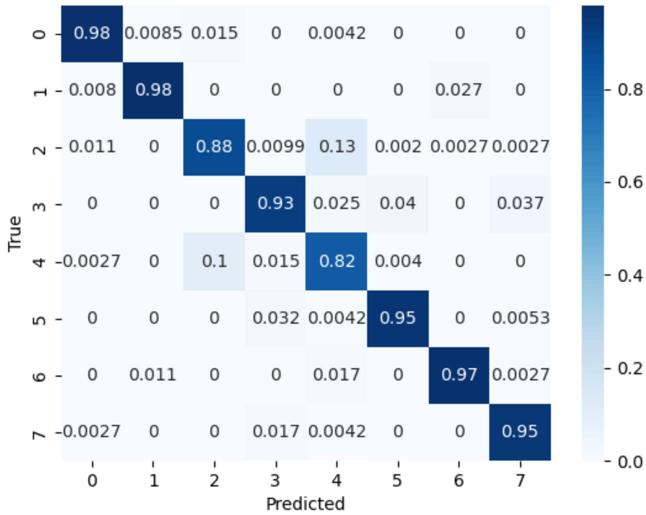


Figure 6.3: Classification Accuracies for Person Identification

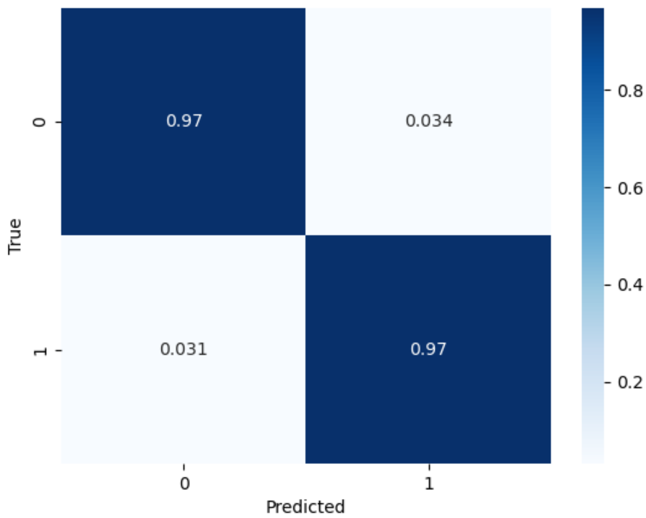


Figure 6.4: Classification Accuracies for Sex Identification

6.2 Continual Learning

To evaluate the performance of the CNN model in a continual learning setting, the study used the rehearsal strategy in an incremental task method. The CNN model was first trained on the gait data from the first 3 subjects, then incrementally trained on the data from the next 3 subjects, and finally, the remaining 2 subjects. The model was trained on each incremental task for 30 epochs.

The results showed that the CNN model could learn and adapt to the new gait data from the remaining sets of 3 subjects and the last 2 subjects using rehearsal method precise memory buffer process. The model's accuracy increased and also decreased as it was trained on each incremental task.

The accuracy decreased in the senario where memory data did not contain samples or even sample of particular subject hence the model suffered forgetting.

However accuracy increased in the scenario where memory data samples contain previous data of the subject. The confusion matrix showed that the model could correctly classify the gait data for each subject where memory had enough previous data of the subject, even after being trained on new data. This seen in figures 6.5, 6.6 and 6.7 below.

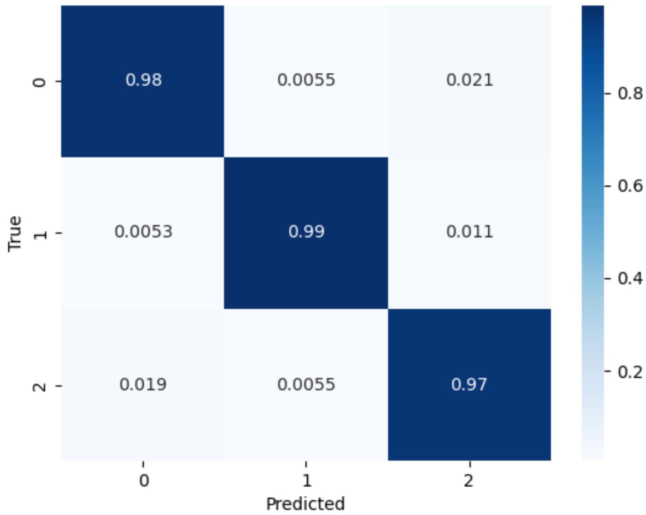


Figure 6.5: Accuracies for first task

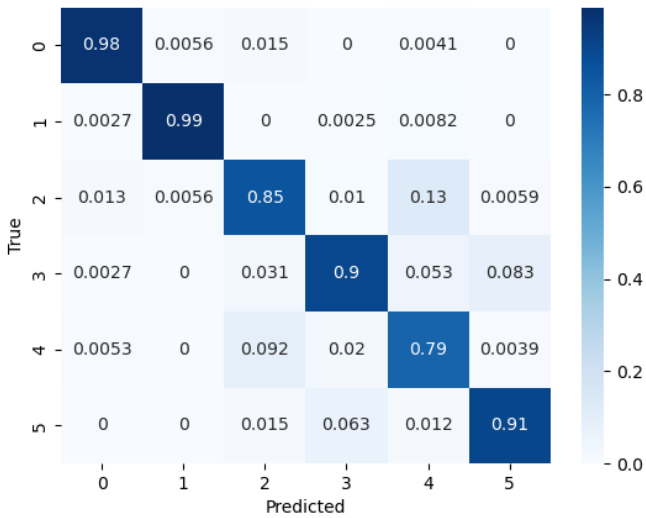


Figure 6.6: Accuracies for incremental second task

Figure 6.7: Accuracies for incremental third task

7
Conclusion

In this study, we presented an analysis of gait recognition using smart glasses sensor data. We trained a CNN model on gait data collected from eight subjects and evaluated the model's performance using two different classification modes: multi-class and binary classification. We also investigated the use of continual learning in an incremental task method to improve the model's performance.

Our results showed that the CNN model achieved an overall average accuracy of 93% in classifying the gait data of the eight subjects in multi-class classification mode and 97% in binary classification mode. The confusion matrix indicated that the model performed well in classifying the gait data for each subject, with only a few misclassifications.

Furthermore, after adapting the model to learn new data in a continual learning setting, the confusion matrices showed that the continual setting had worse performance than the standard setting with an overall accuracy of 89.88%.

Our study demonstrates the effectiveness of using sensor data from smart glasses for gait recognition and classification. The results also highlight the challenge of correctly applying the rehearsal method to improve the adaptation of CNN models for continual learning in gait analysis.

Our study provides a foundation for future gait recognition analysis using smart glasses sensor data. Therefore, we recommend subsequent research on using other deep learning architectures, such as RNN and LSTM, to improve gait recognition and classification performance. We also suggest using transfer learning techniques[30] to improve the model's accuracy, particularly when working with limited data. Other continual learning methods should be utilized in gait recognition studies, and their performance should be compared with the method used in our study.

Future work could investigate the potential of using smart glasses-based gait recognition as a biometric authentication system, particularly for people with mobility impairments, to monitor their progress. This approach could also be useful for investigating fall detection in elderly people. Although, the magnetometer is usually not leveraged in wearable sensor-based gait analysis, further research could investigate using the magnetometer's geolocalization capabilities in tracking and improving gait patterns in patients with Parkinson's or Alzheimer's since they may suffer from dementia and memory loss.

In conclusion, our study demonstrates the effectiveness of using smart glasses sensor data and the CNN model for gait recognition. Additionally, the study shows that continual learning using rehearsal strategies in the incremental task method could be a viable approach. However further research is needed to better understand the implementation of the rehearsal method for improving the model's performance in a dynamic setting.

References

[1] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "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.

[2] G. Roy, T. Jacob, D. Bhatia, and S. Bhaumik, "Optical marker-and vision-based human gait biomechanical analysis," *Hybrid Machine Intelligence for Medical Image Analysis*, pp. 275–291, 2020.

[3] S. Y. Lee, "Iot-based sensor shoes system for gait correction," 2022.

[4] Y. Wurmser, "Us smart wearables users, 2019-2026," 2023.

[5] M. Soldan, "Deep recurrent neural network based multi-modal gait anomaly detection system," 2017.

[6] R. Bonetto, M. Soldan, A. Lanaro, S. Milani, and M. Rossi, "Seq2seq rnn based gait anomaly detection from smartphone acquired multimodal motion data," *arXiv preprint arXiv:1911.08608*, 2019.

[7] Z. Song, Z. Cao, Z. Li, J. Wang, and Y. Liu, "Inertial motion tracking on mobile and wearable devices: Recent advancements and challenges," *Tsinghua Science and Technology*, vol. 26, no. 5, pp. 692–705, 2021.

[8] I. Rida, N. Almaadeed, and S. Almaadeed, "Robust gait recognition: a comprehensive survey," *IET Biometrics*, vol. 8, no. 1, pp. 14–28, 2019.

[9] S. Sarkar and Z. Liu, "Gait recognition," in *Handbook of Biometrics*. Springer, 2008, pp. 109–129.

[10] D. Sethi, S. Bharti, and C. Prakash, "A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and future work," *Artificial Intelligence in Medicine*, p. 102314, 2022.

[11] C. Prakash, R. Kumar, and N. Mittal, "Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges," *Artificial Intelligence Review*, vol. 49, no. 1, pp. 1–40, 2018.

[12] L. K. Drewes, P. O. McKeon, G. Paolini, P. Riley, D. C. Kerrigan, C. D. Ingersoll, and J. Hertel, "Altered ankle kinematics and shank-rear-foot coupling in those with chronic ankle instability," *Journal of sport rehabilitation*, vol. 18, no. 3, pp. 375–388, 2009.

[13] S. Rahati, R. Moravejian, and F. M. Kazemi, "Gait recognition using wavelet transform," in *Proceedings of the Fifth International Conference on Information Technology: New Generations*, 2008, pp. 932–936.

[14] M. Zago, M. Luzzago, T. Marangoni, M. De Cecco, M. Tarabini, and M. Galli, "3d tracking of human motion using visual skeletonization and stereoscopic vision," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 181, 2020.

[15] M. Alotaibi and A. Mahmood, "Improved gait recognition based on specialized deep convolutional neural network," *Computer Vision and Image Understanding*, vol. 164, pp. 103–110, 2017.

[16] Y. Xu, W. Yang, M. Chen, S. Chen, and L. Huang, "Attention-based gait recognition and walking direction estimation in wi-fi networks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 2, pp. 465–479, 2020.

[17] G. Cola, M. Avvenuti, F. Musso, and A. Vecchio, "Gait-based authentication using a wrist-worn device," in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2016, pp. 208–217.

[18] H. Qu, H. Rahmani, L. Xu, B. Williams, and J. Liu, "Recent advances of continual learning in computer vision: An overview," *arXiv preprint arXiv:2109.11369*, 2021.

[19] D. Lopez-Paz and M. Ranzato, "Gradient episodic memory for continual learning," *Advances in neural information processing systems*, vol. 30, 2017.

[20] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "icarl: Incremental classifier and representation learning," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 2001–2010.

[21] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[22] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2935–2947, 2017.

[23] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel *et al.*, "Never-ending learning," *Communications of the ACM*, vol. 61, no. 5, pp. 103–115, 2018.

[24] A. Mallya and S. Lazebnik, "Packnet: Adding multiple tasks to a single network by iterative pruning," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 7765–7773.

[25] J. Serra, D. Suris, M. Miron, and A. Karatzoglou, "Overcoming catastrophic forgetting with hard attention to the task," in *International Conference on Machine Learning*. PMLR, 2018, pp. 4548–4557.

[26] Y. Ren, Y. Chen, M. C. Chuah, and J. Yang, "User verification leveraging gait recognition for smartphone enabled mobile healthcare systems," *IEEE Transactions on Mobile Computing*, vol. 14, no. 9, pp. 1961–1974, 2014.

[27] T. Teixeira, D. Jung, G. Dublon, and A. Savvides, "Pem-id: Identifying people by gait-matching using cameras and wearable accelerometers," in *2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*. IEEE, 2009, pp. 1–8.

[28] T. Dozat, "Incorporating nesterov momentum into adam technical report," 2015.

[29] A. Douillard and T. Lesort, "Continuum: Simple management of complex continual learning scenarios," 2021.

[30] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Advances in neural information processing systems*, vol. 27, 2014.

Acknowledgments

During the course of the research, there were great deal of challenges and obstacles. I wish to express my deepest gratitude to my supervisor, Professor Simone Milani who provided the opportunity to learn and apply deep learning to a real-world and practical problem. I am indebted to Daniele, my co-supervisor, for his patient, support and invaluable help without which the project and research would not have been realized.

I whole-heartedly appreciate the support of my friends and colleagues throughout the course of this project. Especially Chris, Doreen, Gauri, Maxwell, Ting, Nas and Alois who all dedicated their time to help with their gait data which is a crucial part to this thesis success.

To the Gabellis, and my housemates at Cattaro, I very much appreciate the unwavering support, the laughs and encouragement not only during the duration of the project but the entire master's degree.

I'd also like to extend my gratitude to my friends and everyone who helped and supported me along the way.