A Graph Convolutional Approach for Joint Position-Based Gait Recognition

Md. Bakhtiar Hasan

Department of Computer Science and Engineering
Islamic University of Technology (IUT)
May, 2022.

A Graph Convolutional Approach for Joint Position-Based Gait Recognition

by

Md. Bakhtiar Hasan

Supervisor
Dr. Md. Hasanul Kabir
Professor
Department of Computer Science and Engineering
Islamic University of Technology

MASTER OF SCIENCE
IN
COMPUTER SCIENCE AND ENGINEERING

Department of Computer Science and Engineering
Islamic University of Technology (IUT)
Board Bazar, Gazipur-1704, Bangladesh.
May, 2022.

CERTIFICATE OF APPROVAL

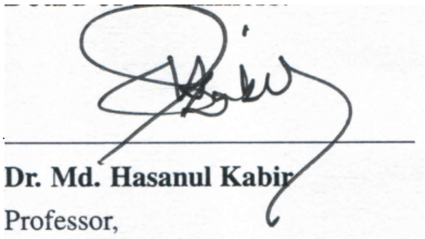
The thesis titled,“**A Graph Convolutional Approach for Joint Position-Based Gait Recognition**” submitted by Md. Bakhtiar Hasan, St. No. 181041013 of Academic Year 2018-19 has been found as satisfactory and accepted as partial fulfillment of the requirement for the degree Master of Science in Computer Science and Engineering on May 12, 2022.

Board of Examiners:

Dr. Md. Hasanul Kabir	Chairman
Professor,	(Supervisor)
Department of Computer Science and Engineering,	
Islamic University of Technology (IUT), Gazipur.	
Dr. Abu Raihan Mostafa Kamal	Member
Professor and Head,	(Ex-Officio)
Department of Computer Science and Engineering,	
Islamic University of Technology (IUT), Gazipur.	
Dr. Md. Moniruzzaman	Member
Assistant Professor,	
Department of Computer Science and Engineering,	
Islamic University of Technology (IUT), Gazipur.	
Dr. Md. Forhad Rabbi	Member
Professor,	(External)
Department of Computer Science and Engineering,	
Shahjalal University of Science and Technology (SUST), Sylhet.	

Declaration of Candidate

This is to certify that the work presented in this thesis is the outcome of the analysis and experiments carried out by **Md. Bakhtiar Hasan** under the supervision of **Dr. Md. Hasanul Kabir**, Professor, Department of Computer Science and Engineering (CSE), Islamic University of Technology (IUT), Dhaka, Bangladesh. It is also declared that neither this thesis nor any part of it has been submitted anywhere else for any degree or diploma. Information derived from the published and unpublished work of others have been acknowledged in the text and a list of references is given.



Dr. Md. Hasanul Kabir
Professor,
Department of Computer Science and Engineering
Islamic University of Technology (IUT)
Date: May 12, 2022.

Md. Bakhtiar Hasan
Student No.: 181041013
Date: May 12, 2022.

Dedicated to my parents

Table of Contents

List of Figures

Page

List of Tables

List of Abbreviations

CNN	Convolutional Neural Network
DAE	Deep AutoEncoders
DBN	Deep Belief Network
DNN	Deep Neural Network
DTW	Dynamic Time Warping
GAN	Generative Adversarial Network
GCN	Graph Convolutional Network
GRU	Gated Recurrent Unit
HMM	Hidden Markov Model
HPP	Horizontal Pyramid Pooling
HRN	Human Mesh Network
k**NN**	k-Nearest Neighbor
LSTM	Long Short-Term Memory
MLP	Multilayer Perceptron
RNN	Recurrent Neural Network
SRB	Set Residual Block
SVM	Support Vector Machine

Acknowledgment

My gratitude goes towards the almighty Allah, the supreme ruler of the universe, for enabling me to complete the thesis for the fulfillment of the degree of Master of Science in Engineering in due time by His grace and for granting me sound health and energy to carry out this research work successfully.

I express my deepest sense of gratitude, sincere appreciation, and immense indebtedness to Dr. Md. Hasanul Kabir, Professor, Department of Computer Science and Engineering, Islamic University of Technology for his support, careful supervision, scholastic guidance, instructions, encouragement, and constructive criticism during the period of the research work.

I convey my profound respect and heartiest gratitude to all the faculty members and staffs of the Department of Computer Science and Engineering, Islamic University of Technology for their active cooperation and sincere help in carrying out the research work.

Last but not the least, I would like to direct all my appreciation to my beloved parents and my spouse for their inspiration and endless encouragement.

Abstract

Gait recognition is becoming one of the promising methods for biometric authentication as a consequence of its self-effacing nature. Contemporary approaches of joint position-based gait recognition model gait features using spatio-temporal graphs. To incorporate long-range relationships among joints, these approaches utilize multi-scale operators. However, they fail to provide equal importance to all joint combinations resulting in an incomplete realization of long-range relationships. Further, only considering joint coordinates can fail to capture discriminatory information provided by the bone structures and motion. In this dissertation, a novel multi-scale graph convolution approach is proposed that utilizes an efficient hop-extraction technique to attenuate the issue. DropGraph regularization technique is employed to avoid overfitting the training samples. Utilizing these techniques, a multi-stream Graph Convolution Network is proposed that combines the joint, bone, and motion features. Finally, the architecture is further improved by introducing a Part-wise Attention technique that helps to identify the most important body parts over the gait sequence. On benchmark gait recognition dataset CASIA-B, the proposed system achieves 96.5%, 93.0%, and 90.1% in normal (NM), walking while carrying a bag (BG), walking while heavily clothed (CL) conditions outperforming the state-of-the-art joint position-based gait recognition methods in BG and CL conditions and achieving comparable performance in NM condition.

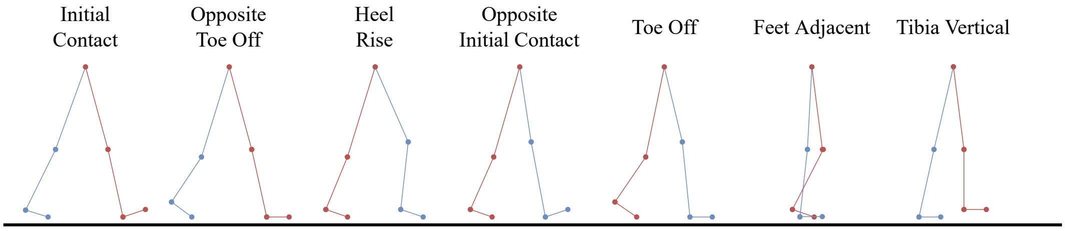
Chapter 1

Introduction

Biometric authentication pertains to the identification and re-identification of human individuals by analyzing their physical and behavioral characteristics. Due to the permanence and uniqueness of these characteristics, they can be used to differentiate one human being from another. Consequently, the use of systems that utilize complex modalities and features is on the rise. In this regard, the use of gait as a method of non-intrusive biometric authentication is getting more popular day by day.

Gait denotes the pattern generated by the movement of body parts of animals during locomotion over a plane. A variety of gaits is used by different animals based on the terrain, necessity, and energy cost. Gait is different among different animals due to their anatomy and habitat differences. As demonstrated in Figure 1.1, the human gait cycle can be seen as a sequence of repetitive steps involving the muscles and skeleton in coordination with the nervous, cardiac, and respiratory system [1]. Since able-bodied people tend to start walking from a very early age, this involuntary gait pattern becomes an intrinsic part of our life. For this reason, this pattern can be utilized to identify humans. This method of identifying humans using automatic extraction of gait characteristics is known as gait recognition.

Gait recognition does not obstruct the regular activities of the subject being authenticated, can be computed from a distance, can work without explicit human coopera-



■ ■	Loading Response	Mid-stance	Termial Stance	Pre-swing	Initial Swing	Mid Swing	Terminal Swing
■ ■	Stance Phase				Swing Phase		

Figure 1.1: The mechanics of gait. (Adapted from [2])

tion, is difficult to copy, and cannot be hidden easily [3, 4]. As a result, it has applications in biometric authentication [5, 6], clinical applications and healthcare [7–12], science of sports [13, 14], style and affect analysis [15, 16], etc.

1.1 Motivation and Scope

Non-wearable gait recognition systems are mostly vision-based. These systems utilize imaging sensors to capture the gait of the subject. As a result, they do not require subject cooperation and can identify human subjects from a distance. However, the performance of these systems is prone to i) variations in the appearance of the subject; ii) variations in the viewpoint or camera angle; iii) occlusion resulting from appearance or viewpoint, and iv) variations in the environment [17].

Recent joint position-based approaches to gait recognition depend on extracting the physical structure of a subject's body [18,19]. In the recent past, these approaches were mostly avoided due to their high computational requirement. Nowadays, the advances in robust pose estimation techniques have now made them feasible again. Moreover, the joint data extracted by pose estimators contain only joint positions, that can provide key information regarding the gait devoid of environmental noises. This allows the gait recognition systems to focus entirely on extracting robust spatial features that are necessary for gait recognition. To extract the temporal information, 3D-CNNs [18] or LSTMs [19, 20] are being used. These approaches can comprehend spatio-temporal information but require high computational resources.

Graph convolutional networks (GCN) have gained a lot of traction lately owing to their capability in harnessing the power of arbitrarily structured graphs using graph convolution filters [21]. The adjacency matrix of the underlying graph along with a feature vector representing each node can be used to model both the spatial information and temporal relationships available in the gait sequence to learn discriminative and robust features. Recent approaches [22, 23] extract gait features by forming a spatio-temporal graph from the available video sequences. These approaches mostly perform local convolutions. But since walking is accomplished using various body parts that are distant from each other, local movements conducted by a few adjacent joints could be ambiguous when differentiating between different gaits. To aggregate the effects of distant body parts, higher-order polynomials of the adjacency matrix have been proposed in the existing literature [24]. Unfortunately, this formulation is affected by the *biased weighting problem*, due to the cyclic walks present in the graph representing human gait. This results in a higher priority towards closer joints than the further ones.

Another useful trait of gait recognition systems is the propensity to leverage the discriminatory information provided by the joint coordinates, their bone structures, and motion [25]. This necessitates the use of a deep neural architecture that is able to capture the intricate information from the said features. Additionally, it should be able to identify and prioritize body parts that are useful in gait recognition. However, deeper architectures often tend to overfit, which means the generalization capability of the architecture should also be ensured.

1.2 Problem Statement

Based on the discussion above, this research aims to develop a multi-stream aggregation technique that can effectively model the relationship between closer and further limbs while also prioritizing important body parts for gait recognition using features that are independent of variations in appearance, camera viewpoint, occlusion, and background.

The specific objectives of this research are:

1. Creating a graph representation of the gait video sequences by extracting the joint positions using pose estimation techniques

2. Integrating multi-stream features extracted from joint, bone structure, and motion in a multi-scale feature aggregation scheme that can extract comprehensive understanding of gait considering both closer and further limb joints

3. Training a robust architecture capable of identifying specific body parts that are useful in gait recognition while also avoiding overfitting and learning generalized features to perform gait recognition

1.3 Research Challenges

Developing a joint position-based gait recognition system presents numerous challenges. First, the feature representation should be view and scale-invariant while also effectively representing human gait representing the subtle nuances of gait. It should be able to capture intra-class similarities and inter-class variances.

The network, used for gait recognition, should capture subtle gait patterns so that it can identify gait regardless of variation in walking direction, length of the walk, and/or speed. It should be robust to variations in the appearance, camera viewpoint, occlusion, lighting conditions, complex background, etc. It should be able to model the relationship between closer and further joints. To understand how human joints interact during gait, the bone structure and the movement of joints from frame to frame should be incorporated into the network. Finally, considering the misinterpretation of joints caused due to the pose-estimation network, the system should be able to identify and focus on joints that are helpful in gait recognition.

1.4 Research Contributions

The key contributions of this research can be summarized as follows:

1. A gait recognition system is proposed that utilizes joint, bone and joint-motion data that are independent of variances in background, camera angle and appearance of the subject while also providing a better understanding between different joints in a single frame and joints in consecutive frames.

2. A novel hop extraction technique is presented, using which the system can exploit the relationship between closer and further joints during gait while also avoiding redundant dependencies.

3. DropGraph technique is utilized to ensure that the system learns to avoid overfitting to the training set in order to learn generalized features that are helpful in recognizing unseen gait samples.

4. Part-wise attention module is introduced to identify and prioritize only specific body parts that are helpful in recognizing gait.

5. By combining all these techniques mentioned above, the system is able to outperform the state-of-the-art gait recognition methods in the existing literature.

1.5 Organization

The rest of the dissertation is organized as follows. Chapter 2 discusses the background and motivation for gait research. It also identifies the problems persistent in the existing literature. Chapter 3 presents a new gait recognition pipeline that is able to utilize multi-scale, multi-stream features while avoiding overfitting. Chapter 4 analyzes the performance of the proposed pipeline and compares it with other state-of-the-art systems. Chapter 5 concludes our discussion and provides direction for future research scope.

Chapter 2

Background Study

The history of observing gait dates back to Aristotle who described the locomotion of humans and other animals [26]. Later on, the father of Biomechanics, Giovanni Borelli took up his work to understand the mathematics behind animal gait [27]. Subsequently, further studies were conducted to identify the repetitive motion of legs during gait [28, 29]. Advancement in photography and videography facilitated the recording of human and animal gait [30]. After the second world war, research was conducted to understand gait biomechanics [31] and its implication in clinical studies [32, 33] resulted in a substantial understanding of how gait as we know it today. Subsequently, the advent of devices to capture motions and 3D positions further advanced the research on gait.

The use of gait to identify individuals started in the late 1960s [34, 35], who were the first to utilize intra-class similarities and inter-class differences among human gait. Later on, studies on the ability of humans in recognizing other people based on their gait solidified the base for gait recognition in biometric and forensic applications [36–38]. Since then, prominent use of gait can be seen utilizing video data [39], sensors [40], and mobile devices [41]. In this dissertation, we focus on vision-based gait recognition (Henceforth called gait recognition) that utilizes video data to develop machine vision.

2.1 Handcrafted Feature Extraction and Classification

Earlier machine learning-based gait recognition approaches focused on feature extraction and classification (Figure 2.1). First, handcrafted features were extracted from the input modality, such as video frame, silhouette, or skeleton joint. These features were then fed into a machine learning-based classifier, which produced the class label.

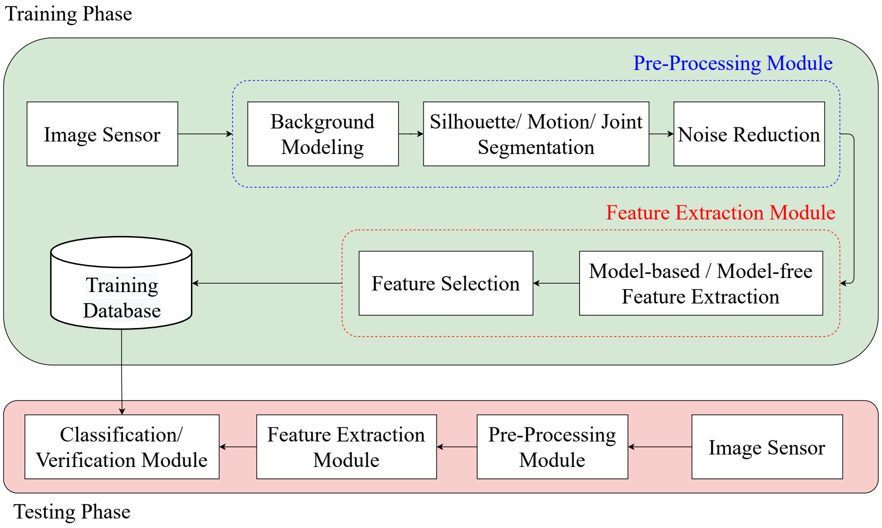
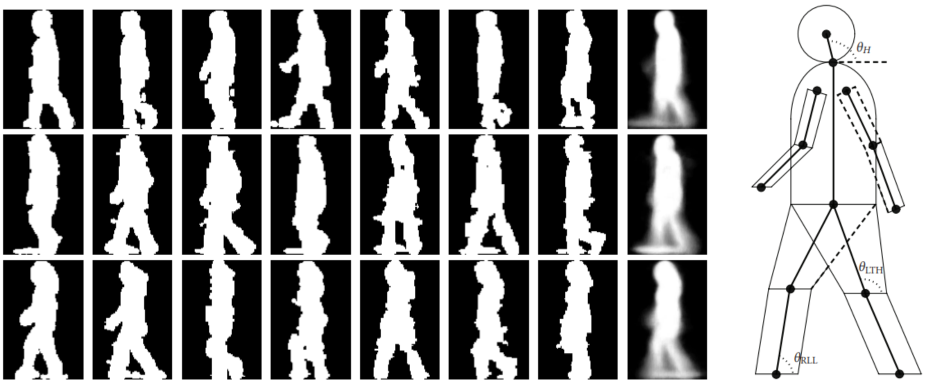

Training Phase
Pre-Processing Module
Image Sensor
Background Modeling
Silhouette/ Motion/ Joint Segmentation
Noise Reduction
Feature Extraction Module
Training Database
Feature Selection
Model-based / Model-free Feature Extraction
Classification/ Verification Module
Feature Extraction Module
Pre-Processing Module
Image Sensor
Testing Phase

Figure 2.1: Generalized handcrafted feature-based architectures

2.1.1 Feature Extraction

The feature extraction phase can be subdivided into two broad categories based on whether they try to fit a walking model onto the gait or not: model-free approaches (Figure 2.2a) and model-based approaches (Figure 2.2b). Unfortunately, due to extreme feature engineering, both of these approaches failed to generalize on large datasets and struggled in recognizing gait in a complex background.



(a) Representation of model-free gait using silhouettes

(b) Representation of model-based gait using distances and angles

Figure 2.2: Features extracted for model-free and model-based gait recognition. (Courtesy of [42] and [43])

One of the most common features for model-free gait recognition was based on silhouettes such as: moving silhouette [44–49], body shapes and size extracted from silhouette [50–63], and average silhouette [42, 48, 64–68]. To add another dimension to 2D silhouette features, depth images were generated using motion capture devices to perform gait recognition [69–71]. Other approaches include the use of Gait Entropy Images [72], Chrono-Gait Image [73], Histogram of Oriented Gradients [74], Gradient Histogram Energy Image [75], etc. Since these methods relied on silhouettes extracted from gait video sequences, they often failed to capture the temporal information that correlates with subsequent video frames.

Model-based gait recognition attempts to extract key anatomical information via body parts or skeletal joints from gait sequences. These approaches required a significant amount of computational resources to accurately model human gait. Pioneering research in this domain focused on fitting different shapes to gait sequences [39, 43, 76–82]. Additionally, skeletal joint-based compound features, such as joint distances, angles, height, stride length, cadence, etc. were used to perform gait recognition [83–91]. From the joint positions, gait trajectories were also mapped in Fourier space to increase class separability [92–95].

2.1.2 Classifiers

Classifiers are responsible for learning patterns from the training data in order to classify samples from the test data. In gait recognition, one of the most popular classifiers is k-Nearest Neighbor (kNN) [96] and its different variants [39, 42, 44–47, 49, 51, 55, 57–59, 61–63, 65, 67, 68, 70–72, 74, 75, 77–84, 88, 90, 92–95]. This is because the extracted features were amalgamated to create one representative sample for each class and stored in a database. To determine the class label of an unknown sample, it was matched with the representative samples, and based on the similarity, the label was decided. The value of k is usually set to 1. Higher values of k are also seen in the existing literature. However, it can cause problems in gait classification as illustrated in Figure 2.3.

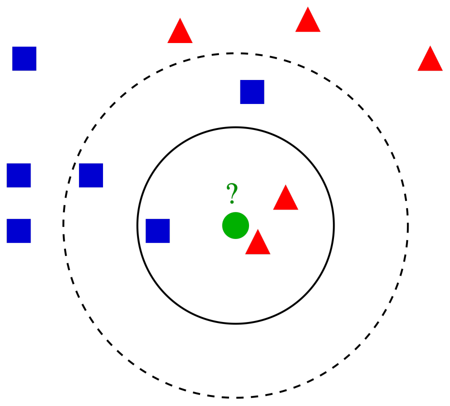


Figure 2.3: Example of kNN classifier. The test sample (shown as green dot) can be classified as red, if $k = 3$. However, it can be classified as blue, if $k = 5$. (Courtesy of [97])

Apart from kNN, Hidden Markov Model (HMM) [48, 52, 54], Dynamic Time Warping (DTW) [60,85,98], Multilayer Perceptron (MLP) [87,91], Linear Time Warping [44], Maximum Likelihood Classifier [50], Genetic Algorithm [69], Naïve Bayes [86], Support Vector Machine (SVM) [89].

2.2 Automated Feature Extraction and Classification

Deep Neural Networks apply a hierarchy of frameworks to extract high-level features using nonlinear functions and use them to classify human gait. These networks include 2D Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), Capsule Networks (CapsNet), Recurrent Neural Networks (RNN), 3D Convolutional Neural Networks (3D CNN), Graph Convolutional Networks (GCN), etc.

2.2.1 2D Convolutional Neural Networks

One of the most common architectures for gait recognition, Convolutional Neural Networks (CNN) employ a set of convolution and pooling layers, and activation functions to generate activation maps that encodes the skeleton joints and silhouettes in gait video sequences (Figure 2.4). The extracted activation maps are then passed through a set of flatten and fully-connected layers, which is then fed through a softmax function to classify based on the probability distribution of the individual classes.

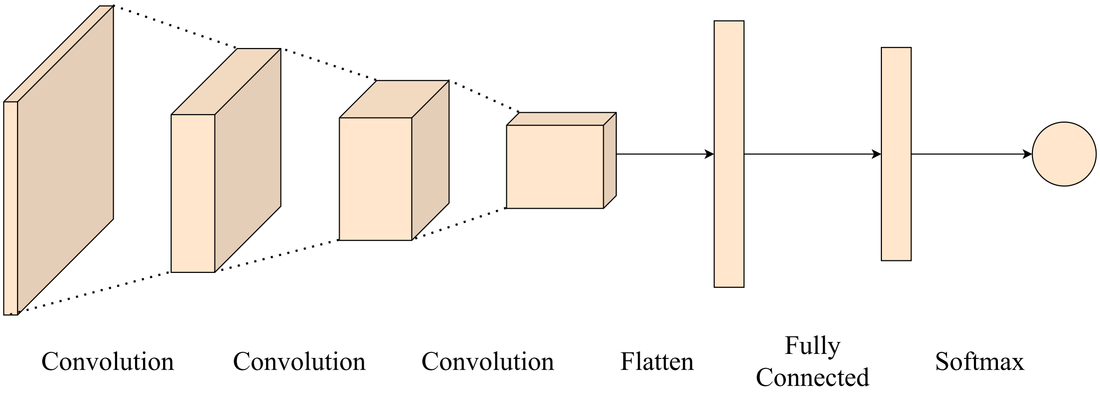


Figure 2.4: Generalized 2D CNN architecture. The architecture consists of a set of convolution, pooling, and fully-connected layers to generate features and perform classification.

Common CNN-based architectures that are utilized in gait recognition are GEINet [99], Ensemble CNNs [100], EV-Gait [101], GaitNet [102], GaitSet [103], Joint-CNN [104] GaitRNNPart [105], GaitPart [106], SMPL [107], CapsGait [108]. All CNN architectures mentioned here, except for GaitNet, require less than 10 layers to extract relevant features. The layers consist of 2-6 convolution layers, 0-2 pooling layers, and 1-3 fully-connected layers. This less number of layers can be attributed to the fact that

CNNs are capable of extracting informative texture information with a large number of layers, which is mostly absent in skeleton joints or silhouettes.

2.2.2 Generative Adversarial Networks

Generative Adversarial Networks (GAN) [109] combine generator and discriminator to address the issue of viewing angle, clothing, and carrying condition invariance in gait recognition (Figure 2.5).

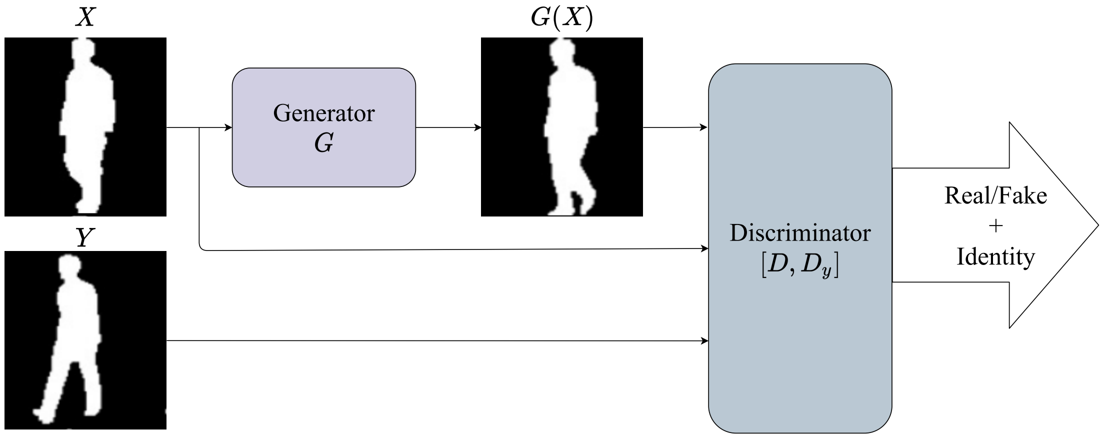


Figure 2.5: Generalized GAN architecture

GANs can be used to change viewing angles, change clothing type, or remove carried objects. To preserve the identifying information while also modifying the appearance, two discriminator networks are often employed by GAN-based architectures: one for distinguishing real and fake samples, and one for preserving identifying information. Different GAN-based architectures such as MGAN [110], DiGAN [111], and TS-GAN [112] are used in gait recognition. These approaches require huge computational resources since multiple architectures need to be trained at the same time to generate, discriminate, and identify gait.

2.2.3 Capsule Networks

Capsule Networks (CapsNet) [113] have been used in gait recognition to model the structural relationships between different body parts by preserving different positional information. The information is encoded using a set of capsule blocks (Figure 2.6).

The network is utilized in gait recognition due to its capability in understanding intrinsic view-invariant features that helps recognize gait from different camera angles. It has been used separately [114] and in combination with other networks [108, 115].

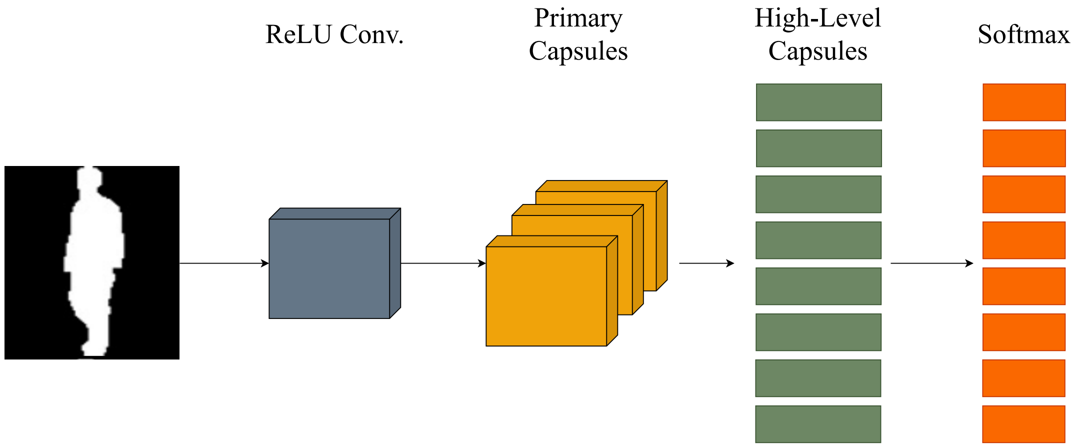


Figure 2.6: Generalized CapsNet architecture

2.2.4 Recurrent Neural Networks

To exploit the temporal relationship among the consecutive frames of a gait sequence, Recurrent Neural Networks (RNN), such as Long Short-Term Memory (LSTM) [116] and Gated Recurrent Unit (GRU) [117], have been applied in gait recognition (Figure 2.7a).

The proposed gait recognition architectures feed either skeleton joints [118] (Figure 2.7b), or CNN-extracted features [105, 108, 119, 120] (Figure 2.7c) to the RNN which then generates the class label for the provided sequence.

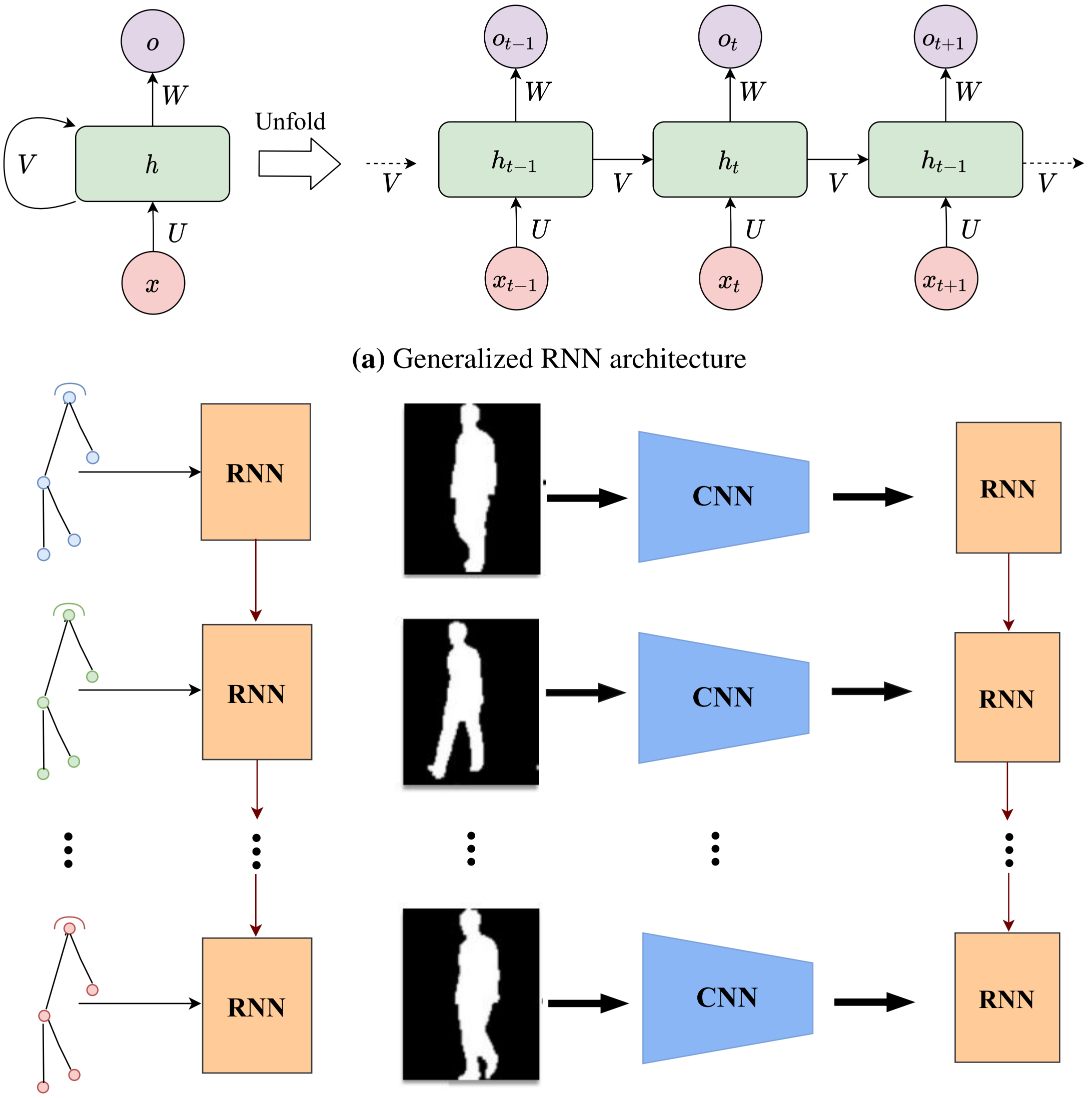
2.2.5 3D Convolutional Neural Network

To combine the spatial and temporal information of gait sequence, 3D Convolutional Neural Networks (3D CNN) have been used to extract view- and appearance-invariant features (Figure 2.8). However, due to the variability in the number of frames in the gait sequence, applying 3D CNNs directly is not possible. As a result, multiple 3D CNNs of varying scales and filter sizes are used in the existing gait recognition literature [121–124].

2.2.6 Graph Convolutional Networks

Graph Convolutional Networks (GCN) [21] have been recently developed as an extension of CNNs that utilizes higher dimensional graph structures and adjacency matrix-based convolution filters (Figure 2.9).

GCNs exploit the inherent graph-like nature of human gait. The advantage of using this approach is that it can combine both structural information from a single frame and temporal relationships among consecutive frames. As a result, the extracted features



(a) Generalized RNN architecture

(b) RNN architectures fed with joint position information

(c) RNN architectures fed with features extracted from CNNs

Figure 2.7: Generalized RNN architecture used in gait recognition.

can be view- and appearance-invariant. [22] was the pioneer in using GCN in combination with Joint Relationship Pyramid Mapping considering the joint positions as vertices and the bones connecting them as edges. Recently, [125] used a combination of ST-GCN [126] and Canonical Polyadic Decomposition [127] to improve the performance. [128] introduced the concept of residual connection in gait recognition models. However, these approaches only considered the joint-stream data for gait recognition. Powered by the multi-stream feature extraction of ResGCN [129], [23, 130] further enhanced this idea by combining bone and motion data with joints. A similar approach was followed by [25].

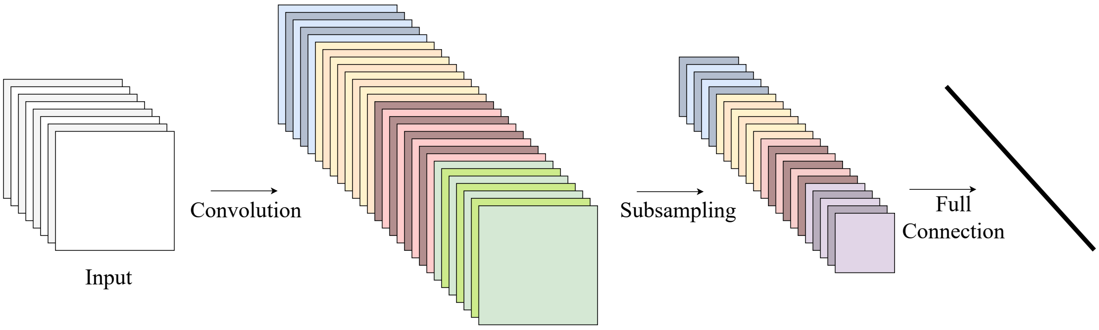


Figure 2.8: Generalized 3D CNN architecture

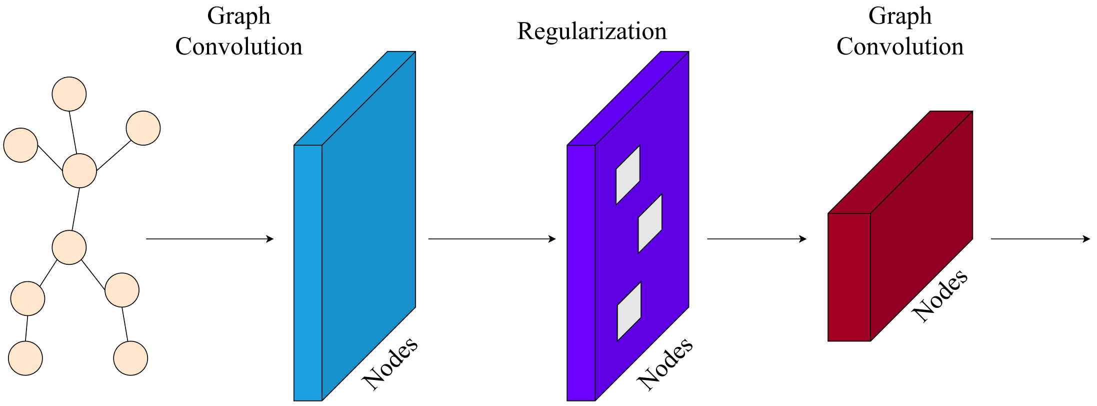


Figure 2.9: Generalized GCN architecture

2.2.7 Other DNNs

Other deep learning based approaches for gait recognition utilize Deep AutoEncoders (DAE) [131–133], Deep Belief Network (DBN) [134,135], Lateral Network [136], Human Mesh Network (HRN) [137], Set Residual Block (SRB) [138], Horizontal Pyramid Pooling (HPP) [139], etc.

Hybrid deep learning architectures combine multiple architectures such as CNN + RNN, RNN + CapsNet, DAE + RNN, DAE + GAN, etc. to harness the power of both architectures. A combination of CNN and RNN can be seen in most of the hybrid networks that combine spatial encoding of CNN and temporal relationship of RNN [105,140–142]. On the other hand, in GANs, DAEs have been used as generators and/or discriminators in GaitGAN [143], GaitGANv2 [144], Alpha-blending GAN [145], and CA-GAN [146]. With a view to disentangle representation learning for gait recognition, a combination of DAEs and RNNs has been used [147, 148]. Finally, to extract robust appearance and view-invariant features, RNNs are combined with CapsNet. Here, CapsNet can act as an attention mechanism by applying higher priority to important features helpful in gait recognition [108, 115].

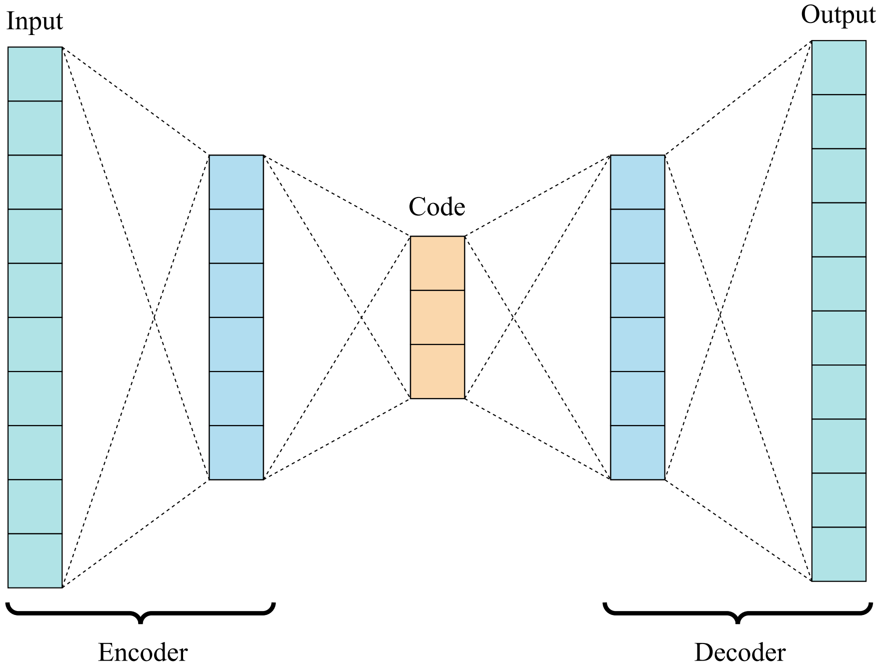


Figure 2.10: Generalized Deep AutoEncoder architecture

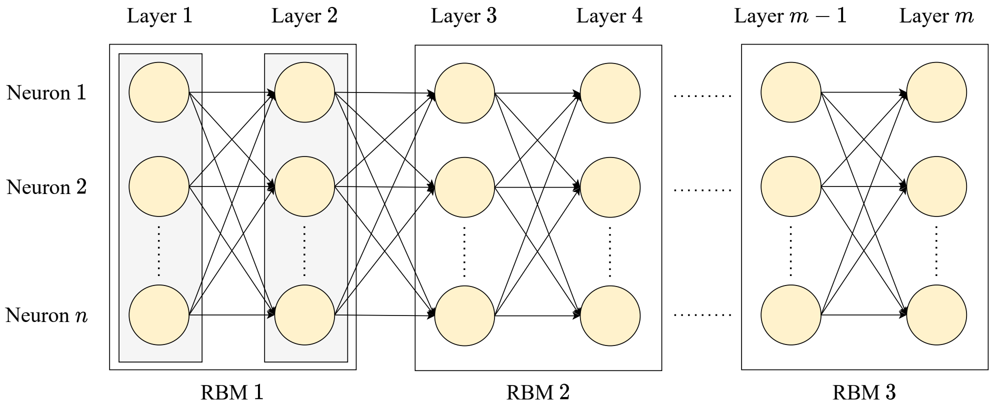


Figure 2.11: Generalized Deep Belief Network architecture

Chapter 3

Proposed Methodology

3.1 Overview

The proposed pipeline for skeleton joint position-based gait recognition system comprises of multiple stages. First, each frame of a gait video sequence is passed through a pose estimation network to determine the joint positions representing the human pose. These joint positions are then utilized to generate the graph representation of the gait sequences by considering the joint positions as vertices and the bones connecting them as edges. The sequence is then preprocessed to remove low confidence predictions. The preprocessed joint sequence is then passed through a graph convolutional network where the bone structure and joint-motion features are generated along with the joint positions. Each data stream is then fed to a set of basic convolution and residual bottleneck blocks consisting of graph convolution and temporal 2D convolution layers. The graph convolution layers are enhanced using hop extraction technique for multi-scale feature aggregation that can extract a comprehensive understanding of gait considering the relationship between closer and further joints. The residual bottleneck blocks in the deeper layers are combined with part-wise attention module to identify and prioritize specific body parts that are useful in gait recognition. DropGraph regularization technique is employed in the training phase of the network to avoid overfitting while also learning generalized features that are useful in identifying unseen gait samples. Finally, based on the activation maps generated from gait sequences, class labels are generated identifying the subjects to whom the sequence pertain to. A pictorial view of the proposed pipeline can be seen in Figure 3.1.

3.2 Pose Estimation Network

A gait video sequence consists of N RGB images $f_1, f_2, \ldots, f_N$. Each image is fed to a pose estimation network to extract M keypoints. In this work, we employ Higher

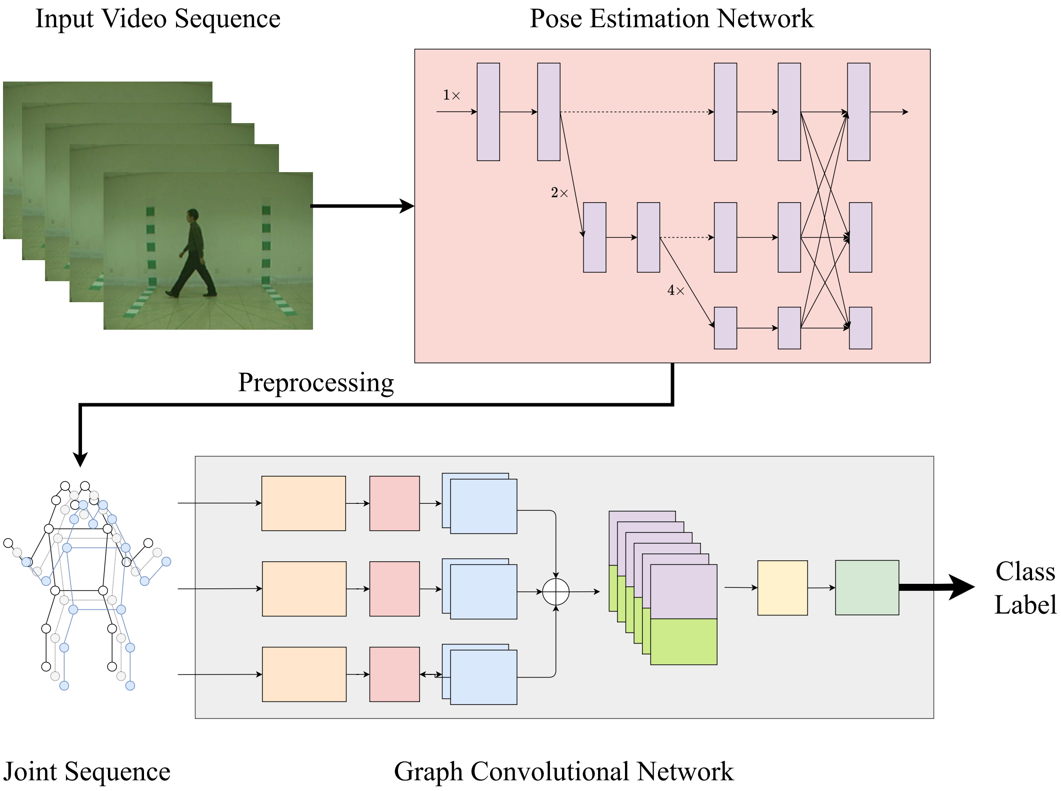


Figure 3.1: Overview of the proposed pipeline for gait recognition

Resolution Net (HRNet) [149] to extract the key points from each frame. The architecture is pretrained on the COCO dataset [150].

Consider each RGB image has a size of $W \times H \times C$, where W, H, and C denotes the width, height, and the number of channels in the image, respectively. To generate M keypoints from each RGB image $f_i\,(1 \le i \le N)$, HRNet generates M heatmaps with size $W' \times H'$ and a set $\{H_1, H_2, \ldots, H_M\}$. Here, $H_i\,(1 \le i \le M)$ denotes the confidence of the network in predicting the i^{th} keypoint.

As shown in Figure 3.2, the network employs a sequential high-resolution subnetwork to maintain the high resolution of the image throughout the pose estimation process. It connects high-to-low and low-to-high subnetworks in parallel to exchange information via a set of downsampling and upsampling process. This results in an accurate and efficient high-resolution representation of key points from the original frame.

The extracted M keypoints denote the joint positions representing the vertices of the graph. In our work, $M = 17$. The joint connections representing the edge information is created using the configuration provided in [23] (Figure 3.3).

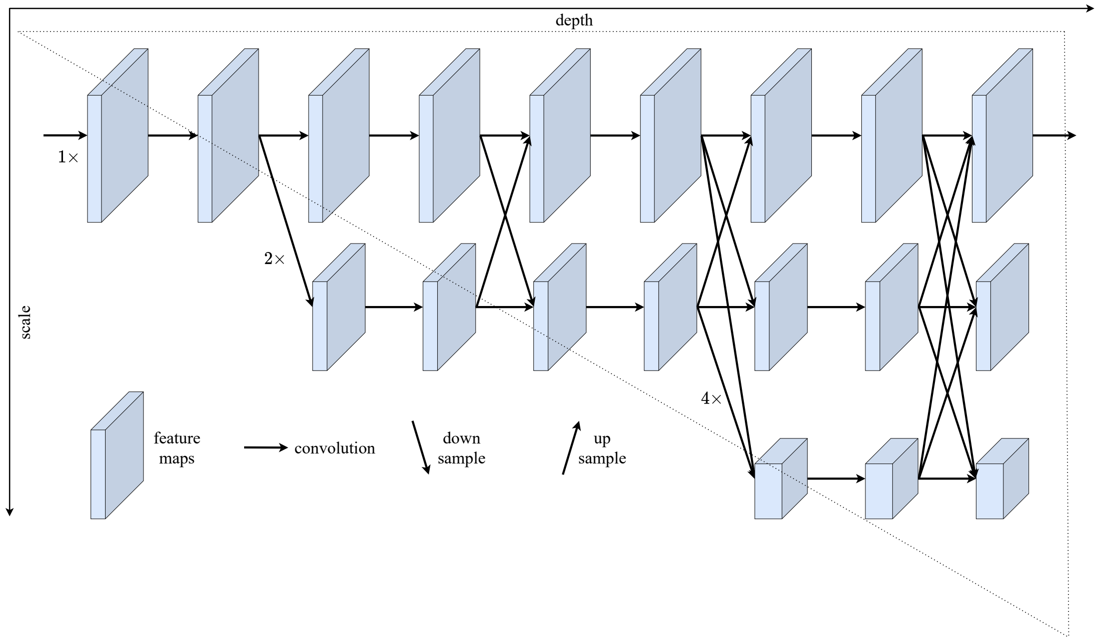


Figure 3.2: Pose extraction architecture, HRNet. (Adapted from [149])

3.2.1 Preprocessing

On top of the 2D coordinates of the joints, HRNet also generates a confidence score for each of its predictions. The higher the confidence score, the better the prediction. In occluded conditions, such as walking while carrying a bag and walking while heavily clothed, the confidence for the generate 2D coordinates can be significantly low.

The 2D coordinates, being a feature and the source for generating other data streams, plays an important role in gait recognition. However, as illustrated in Figure 3.4, the predictions can have low confidence resulting in poor estimation of the joint position.

Our hypothesis is that any joint coordinate with low confidence can capture random

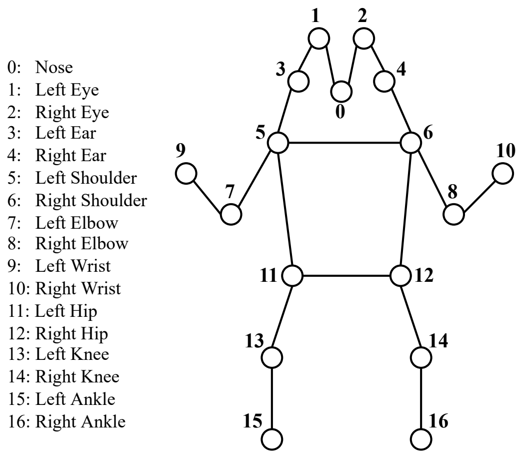
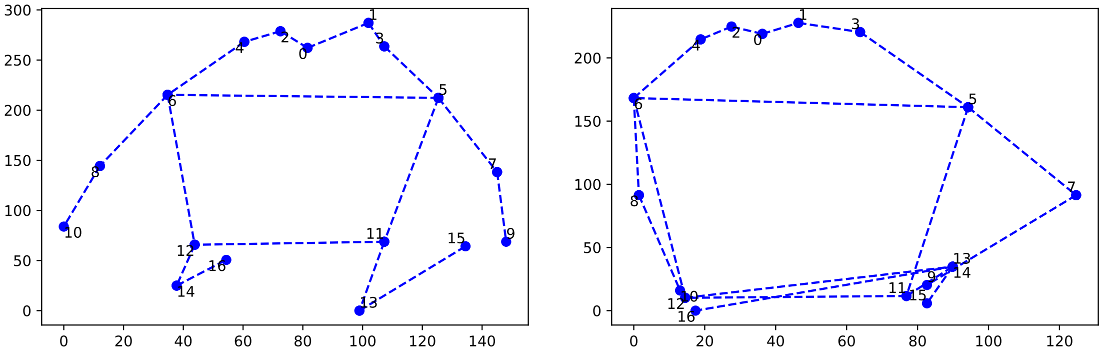


Figure 3.3: Joint data extracted from HRNet and bone configuration



(a) Normal walking condition, captured from 0° angle, 55% confidence

(b) Heavily clothed condition, captured from 0° angle, 58% confidence

Figure 3.4: Illustration showing poorly predicted joint positions

noise hampering both the training process and recognition performance of the graph convolutional network. To understand how confident HRNet is in generating predictions for a certain frame, the average confidence for that frame is calculated:

$$\text{Confidence} = \frac{\sum_{i=1}^{M} c_i}{M} \times 100\% \tag{3.1}$$

Here,
c_i = Confidence score for the i^{th} joint,
M = The number of keypoints.

After that, if the average confidence of the frame is less than threshold C, the frame is removed to alleviate the effect of poor predictions.

In addition to that, since the graph convolutional network learns to recognize gait based on the relationship between different joints in both spatial and temporal dimension, missing information, due to any error in calculation of the coordinates, can have detrimental effect on the training and recognition process. To address the issue, if any frame did not not contain the 2D coordinates and the corresponding confidence score of all 17 joints, that frame was also removed.

3.3 Graph Convolutional Network

Residual Graph Convolutional Network (ResGCN) architecture [129], pretrained on NTU RGB+D [151] and NTU RGB+D 120 [152] to perform action recognition, has been adapted for our task.

In ResGCN (shown in Figure 3.5), initially, the batch normalized multi-stream input is passed through spatial and temporal 'Basic' blocks. The 'Basic' block is a

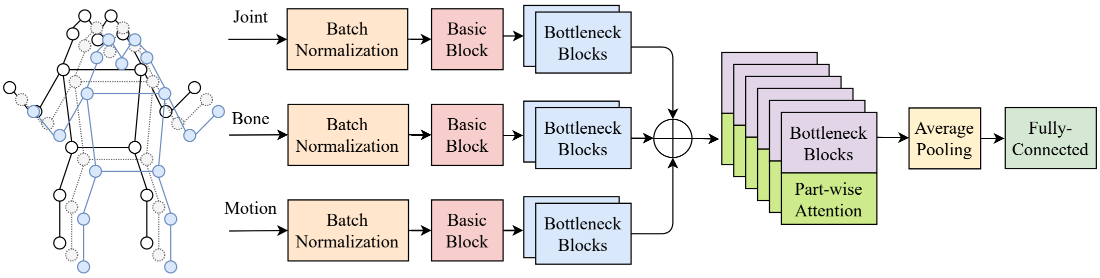


Figure 3.5: Overview of the ResGCN architecture

sequential set of convolution and batch normalization layers that conditionally has a residual connection. The output of the 'Basic' block is produced by passing the output obtained from convolution blocks through a ReLU activation function.

The 'Bottleneck' blocks are used to reconstruct the input using a sequence of up-convolution and down-convolution where each convolution layer is followed by a batch normalization layer. In case of the block being residual, a convoluted and batch normalized form of the input is added with the output. Final output is produced by passing the output through ReLU activation function.

Afterwards, the features are extracted smoothly instead of retaining more pronounced features like edges by applying an average pooling on the output feature map from the last bottleneck block. Finally, the feature vector is mapped to the output units through a fully-connected layer.

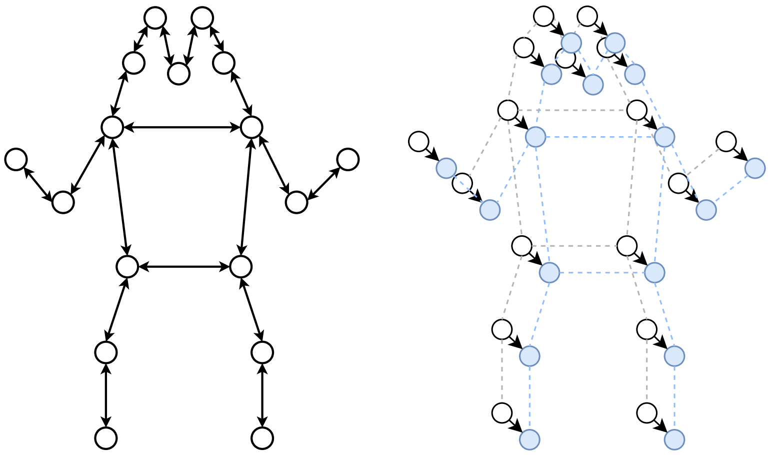
3.3.1 Multi-stream Input

In addition to joint information generated using HRNet, bone and joint-motion information is fed to the ResGCN architecture. These information can provide supplementary insights to the model that might not be readily available. Additionally, evaluation of such multi-stream features early in the network can help reduce the overall complexity of the model by reducing the number of layers that would have been required otherwise to infer such features deep in the network.

A bone is considered as a link between two joints. We define a bone as a vector pointing from one joint to another. Let us assume that a joint in frame $p\,(0 \leq p < N)$ is denoted as $v_{i,p} = (x_{i,p}, y_{i,p})$ and another joint is denoted as $v_{j,p} = (x_{j,p}, y_{j,p})$. Then the bone vector is calculated from $v_{i,p}$ to $v_{j,p}$ as:

$$\begin{aligned} b_{i,p} &= v_{j,p} - v_{i,p} \\ &= (x_{j,p} - x_{i,p}, y_{j,p} - y_{i,p}) \end{aligned} \tag{3.2}$$

The bone information is calculated for each possible pair of joints that are con-



(a) Generated bone data stream. Here, the directions denote that the bone data was generated considering both directions.

(b) Generated motion data stream. Here, the black nodes are in frame f_p and the blue nodes are in frame f_{p+1}.

Figure 3.6: Data streams generated from joint positions

nected by an edge as defined in the earlier section. The introduction of these bone features, as shown in Figure 3.6a, can encode rich structural information that may be helpful for the network to understand how the joints are connected and provide further insight into their interaction.

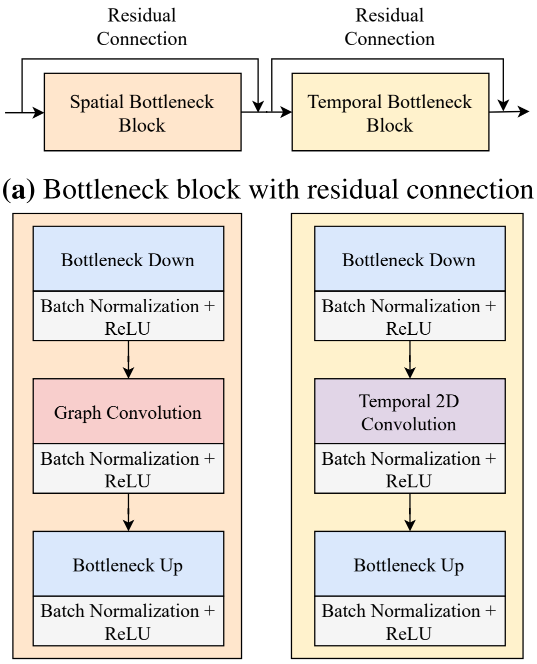
The motion data indicates the change of coordinates for same joint in subsequent frames (Figure 3.6b). The joint-motion for joint $v_{i,p}$ on frame $p\,(0 \leq p < N-1)$ is calculated as:

$$\begin{aligned} jm_{i,p} &= v_{i,p+1} - v_{i,p} \\ &= (x_{i,p+1} - x_{i,p}, y_{i,p+1} - y_{i,p}) \end{aligned} \tag{3.3}$$

The introduction of motion features in this manner can encode the temporal information present in the graph structure of the gait sequence.

3.3.2 Bottleneck and Residual Connection

For faster optimization and to reduce the performance tuning cost, ResGCN utilizes a bottleneck structure with residual connection, pioneered by ResNet [153]. The bottleneck architecture, as shown in Figure 3.7, is composed of a sequence of spatial graph convolution module and temporal 2D convolution module. The graph convolution module and the temporal convolution module can help aggregate information from a single frame of a data stream and multiple frame of a data stream, respectively.


Residual
Connection
Residual
Connection
Spatial Bottleneck Block
Temporal Bottleneck Block
Bottleneck Down
Batch Normalization + ReLU
Graph Convolution
Batch Normalization + ReLU
Bottleneck Up
Batch Normalization + ReLU
Bottleneck Down
Batch Normalization + ReLU
Temporal 2D Convolution
Batch Normalization + ReLU
Bottleneck Up
Batch Normalization + ReLU

(a) Bottleneck block with residual connection

(b) Spatial Bottleneck Block

(c) Temporal Bottleneck Block

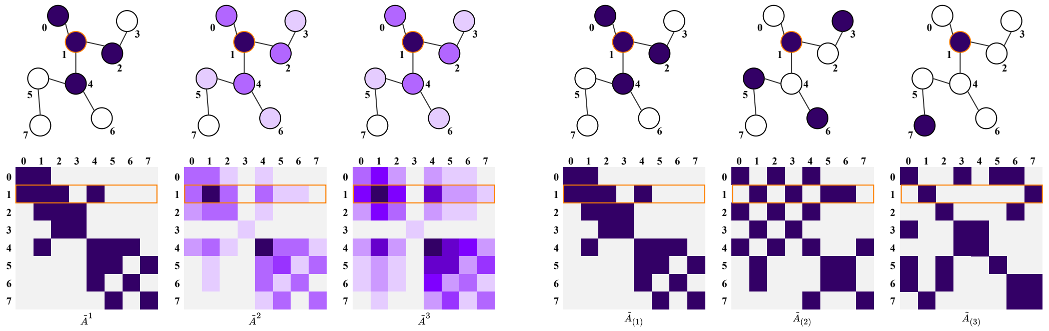
Figure 3.7: Structure of ResGCN bottleneck with residual connection

The bottleneck structure enables the reduction of feature channels by using two 1×1 convolution layers right before and after regular convolution layers. Before each spatial and temporal block, the bottleneck structure is used to reduce the number of parameters in the overall architecture. For example, in a block where the input and output channel is 256, channel reduction rate is 4 and the temporal window size is 9, bottleneck architecture can reduce the number of parameters by more than 88% [129]. This results in faster optimization of the model.

With the increase in the number of layers in GCN, one unstable behavior can be noticed. Since gradients in the deeper layers are calculated as the product of several gradient values, smaller gradients tend to have minimal effect on the weight update, resulting in longer convergence time of the model. This is known as the vanishing gradient problem [129]. The residual connection is used as a skip connection between the spatial and temporal bottleneck blocks. It can help propagate larger gradients to earlier layers resulting in faster convergence of the model [153].

3.3.3 Addressing the Biased Weighting Problem using Hop Extraction Technique

To understand and alleviate the Biased Weighting problem, let us concentrate on the the Graph Convolution module of the Spatial Bottleneck Block.



(a) Biased Weighting Problem: Vertices that are close to the central vertex get higher weights from higher-order polynomial of adjacency matrix, hampering the effectiveness of the long-range relationship.

(b) Proposed Solution: Hop-extracted adjacency matrix ensures equal contribution for each neighbor in a certain distance while also keeping the identify features intact.

Figure 3.8: Demonstration of biased weighting problem and the proposed solution on a simple graph. A lighter color denotes lower weight and vice-versa. Self-loops are not shown to ensure the clarity of the image. Here, 1 is considered as the center vertex.

Let, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the human skeleton graph. Here, $\mathcal{V} = \{v_1, v_2, \ldots, v_P\}$ denotes the set of k vertices corresponding to each keypoint in a data stream. For joint and bone data stream, $P = M$ and for joint-motion data stream, $P = M - 1$. $\mathcal{E}$ denotes the set of edges corresponding to the connections between each keypoint. Each vertex consists of a pair of values and the edge information is represented using an adjacency matrix $\mathbf{A} \in \mathbb{R}^{k \times k}$ where

$$A_{i,j} = \begin{cases} 1, & \text{denotes the existence of an edge between } v_i \text{ and } v_j \\ 0, & \text{otherwise} \end{cases} \tag{3.4}$$

Note that, $\mathbf{A}$ encodes the structural information of the skeleton. Also, $\mathbf{A}$ is symmetric due to the undirected nature of $\mathcal{G}$.

For each data stream, the human gait can be represented using a set of vertex features $\mathbf{X} = \{x_{i,j} \in \mathbb{R}^C \,|\, i, j \in \mathbb{Z}; 1 \leq i \leq N; 1 \leq j \leq P\}$ where $x_{i,j} \in \mathbb{R}^C$ for vertex v_j at frame i.

That means, $\mathbf{X} \in \mathbb{R}^{N \times M \times C}$. And since HRNet extracts 2D coordinates for each joint, $C = 2$. Again, $\mathbf{X}$ encodes the feature information for the skeleton.

The layer-wise update function that is applied on frame f is:

$$X_f^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X}_f^{(l)} \Theta^{(l)}\right) \tag{3.5}$$

Here,

$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, that conserves the identify features of the skeleton by introducing self-

loops,
$\tilde{\mathbf{D}}$ = A matrix containing the degrees of $\tilde{\mathbf{A}}$ in its main diagonal,
$\Theta^{(l)}$ = The learnable weight matrix for layer l, and
$\sigma(\cdot)$ = The activation function.

The purpose of the diagonal degree matrix is to normalize the features to prevent vanishing/exploding gradients [24].

The spatial aggregation framework utilizes higher-order polynomials of the adjacency matrix to aggregate multi-scale structural information.

$$X_f^{(l+1)} = \sigma\left(\sum_{k=0}^{K} \hat{\mathbf{A}}^k \mathbf{X}_f^{(l)} \Theta_{(k)}^{(l)}\right) \tag{3.6}$$

Here,
K = The scale of aggregation,
$\hat{A} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}$.

Note that, $A_{i,j}^k = A_{j,i}^k$ is the total number of length k walks between v_i and v_j. That means, $\hat{\mathbf{A}}^k \mathbf{X}_f^{(l)}$ can be used to perform feature aggregation weighted by the number of such walks.

As illustrated in Figure 3.9, since walks consist of hops between vertex i to vertex j, where i can be equal to j, there can be cyclic walks concentrated in the originating vertex. Additionally, the existence of self-loops to preserve identity features can further increase such walks. As a result, the adjacency matrix contains higher values for the vertices close to the originating vertex and lower values for that further. This creates a bias in feature aggregation, rendering the process less effective in capturing the long-range relationship between joints [17, 154].

We address the above issue by defining a k-adjacency matrix as:

$$\left[\tilde{\mathbf{A}}_{(k)}\right]_{i,j} = \begin{cases} 1, & \text{if } d(v_i, v_j) = k \\ 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases} \tag{3.7}$$

Here,
$d(v_i, v_j)$ is the shortest distance between v_i and v_j considering the number of hops

Since $\mathcal{G}$ is undirected, the value is calculated using Breadth-First Search (BFS) algorithm to find the k-hop neighbors. Note that, $\tilde{\mathbf{A}}_{(1)} = \tilde{\mathbf{A}}$ and $\tilde{\mathbf{A}}_{(1)} = \mathbf{I}$.

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

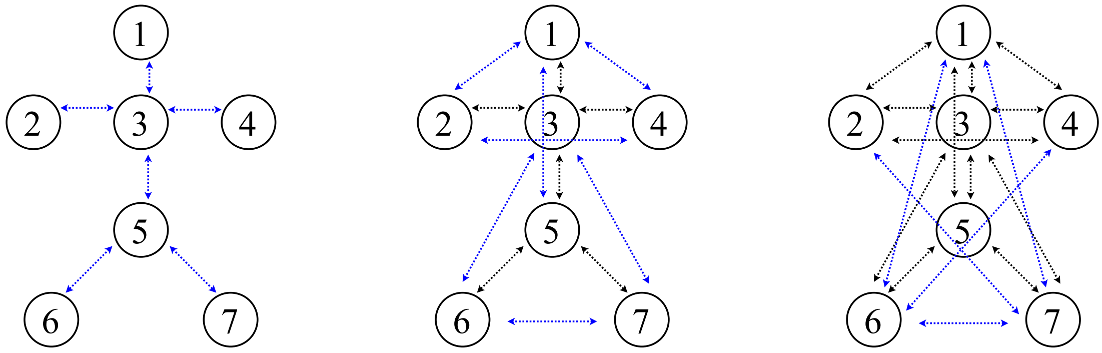
(a) Adjacency matrix for A^1

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

(b) Adjacency matrix for A^2

$$\begin{bmatrix} 0 & 0 & 4 & 0 & 0 & 1 & 1 \\ 0 & 0 & 4 & 0 & 0 & 1 & 1 \\ 4 & 4 & 0 & 4 & 6 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 1 & 1 \\ 0 & 0 & 6 & 0 & 0 & 3 & 3 \\ 1 & 1 & 0 & 1 & 3 & 0 & 0 \\ 1 & 1 & 0 & 1 & 3 & 0 & 0 \end{bmatrix}$$

(c) Adjacency matrix for A^3



(d) Corresponding joint connections for A^1

(e) Corresponding joint connections for A^2

(f) Corresponding joint connections for A^3

Figure 3.9: Illustration of the Biased Weighting problem considering 7 joints with length 3 walks. Here, self-loops are avoided for simplicity. The blue connections are the only joints that need to be considered. The black connections are redundant but still considered due to the property of walks in a graph.

Thereafter, incorporating Equation 3.7 with Equation 3.6, we get:

$$X_f^{(l+1)} = \sigma\left(\sum_{k=0}^{K} \tilde{\mathbf{D}}_{(k)}^{-\frac{1}{2}} \tilde{\mathbf{A}}_{(k)} \tilde{\mathbf{D}}_{(k)}^{-\frac{1}{2}} \mathbf{X}_f^{(l)} \Theta_{(k)}^{(l)}\right) \tag{3.8}$$

In contrast with Equation 3.6 where the total number of length k walks is dependent on length $k-1$ walks, Equation 3.8 provides equal importance to the vertices in closer and further neighborhoods alleviating the biased weighting problem. Consequently, this results in effective consideration of long-range relationships. A sample illustration considering 7 joints can be seen in Figure 3.10.

3.3.4 Addressing the Overfitting Issue using DropGraph Technique

Considering ResGCN has to handle multiple data streams in a large network, it is very much likely that overfitting issues can occur in the model. As a result, the model might try to learn the structural noise in the training data which in-turn can have negative consequences in the performance of the model in unseen test data. This is known as

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

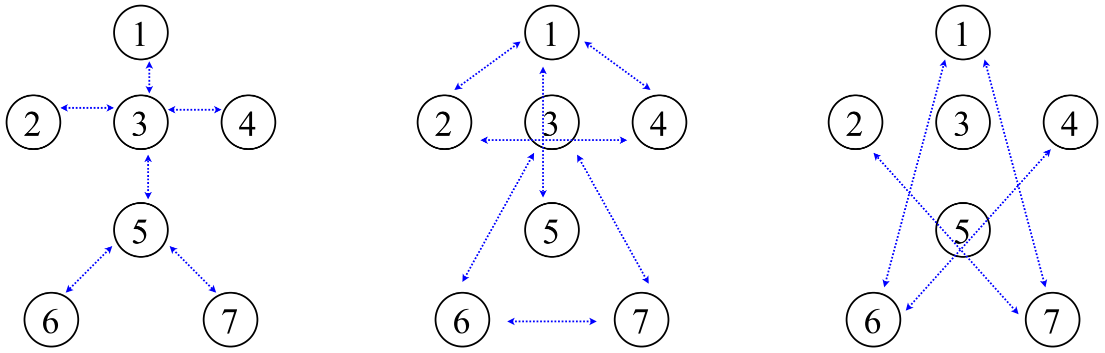
(a) Adjacency matrix for A_1

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

(b) Adjacency matrix for A_2

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(c) Adjacency matrix for A_3



(d) Corresponding joint connections for A_1

(e) Corresponding joint connections for A_2

(f) Corresponding joint connections for A_3

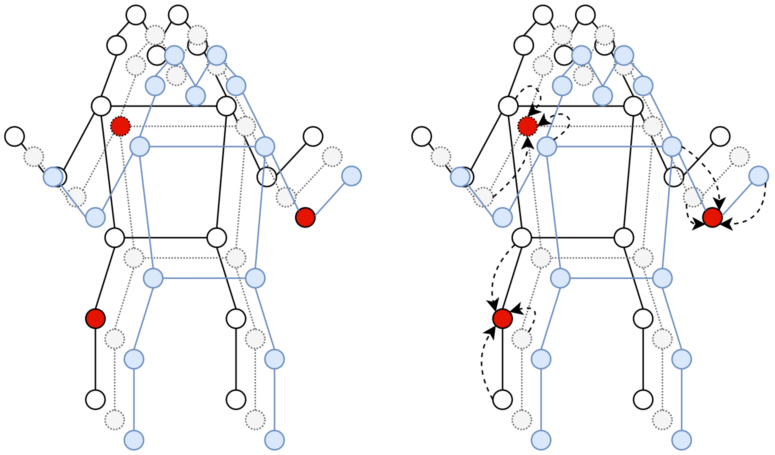
Figure 3.10: Solution to the Biased Weighting problem considering 7 joints and hop distance 3. Here, self-loops are avoided for simplicity.

overfitting.

Our initial experiments with different baseline architectures of the existing state-of-the-art model showed that, even after careful consideration via augmentation of training samples and various measures taken to control the convergence, the difference between the training and test accuracies were huge. For example, one of the baseline architectures had a remarkable 88.60% training accuracy. However, the test accuracy was only 69.90%. These results further enforced the notion of overfitting in the graph convolutional network.

To address the overfitting issue, one simple solution could be the use of dropout layers [155]. During training, the dropout layer is used to ignore a set of randomly selected key points. In graph convolutional network, this would translate to randomly selecting a set of vertices and setting their activation to 0 (Figure 3.11a). This minimizes the complex co-adaptation of the network layers forcing the network to learn sparse representation that are more helpful in identifying unseen samples.

However, as illustrated in Figure 3.11b, due to the closely related nature of the key points, the features of a keypoint can be estimated from the features of the neighboring key points. As a result, if a certain node is dropped, the information for that node can be extracted from the neighboring nodes leading to overfitting [156]. To solve this



(a) Dropout applied to graph convolutional network. Vertices are randomly selected (red colored) and their activation is set to 0.

(b) Inference of a keypoint based on neighboring key points resulting in the nullification of the effect of dropout. (A few of the possible connections are shown for simplicity)

Figure 3.11: Dropout regularization technique and its problem in graph convolutional network

issue, DropGraph is utilized by dropping the entire node set in a neighborhood.

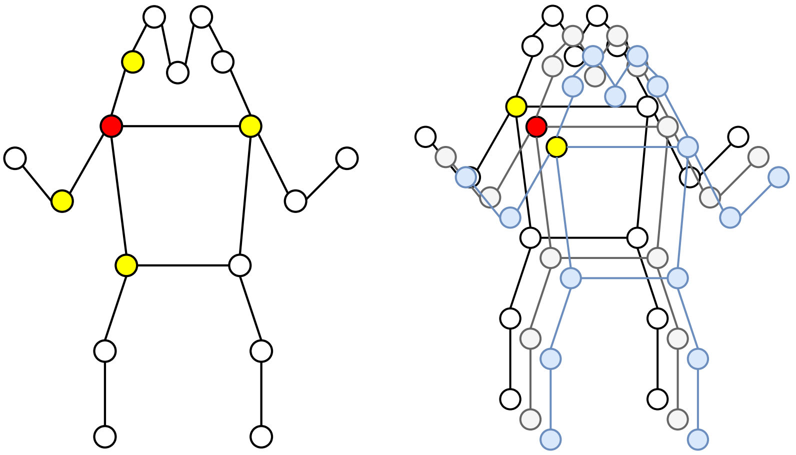
DropGraph technique selects a vertex v_{base} from the graph with a small probability λ. Then, the nodes that are at most K steps away from the v_{base} are dropped by setting their activation to 0. This introduces a regularization effect in the network by randomly dropping a set of values while also ensuring that the values can be reconstructed from the neighboring nodes.

As illustrated in Figure 3.12, this approach is applied randomly to a joint and its neighborhood in a single frame (spatial dimension), or a joint and its neighborhood in previous and subsequent frames (temporal dimension).

3.3.5 Part-wise Attention

Not all body parts contribute equally in gait. For example, since walking consists of mostly hand and leg movement, it might be beneficial to focus on the changes in joint positions that are related to hands and feet to better to better recognize gait. That means, identifying and prioritizing important body parts that are helpful in gait recognition can improve the performance of the system. This can be achieved using attention mechanism [157].

Motivated by the Split Attention of ResNeSt model [158], a Part-wise Attention block is used to understand the importance of different body parts over the entire gait sequence. This block is used to evaluate the attention weight of different body parts.



(a) DropGraph applied in spatial dimension

(b) DropGraph applied in temporal dimension

Figure 3.12: Demonstration of DropGraph applied in both spatial and temporal dimensions. Here, the red node denotes the V_{base} node and the yellow nodes are its neighbors, where $K = 1$.

This facilitates the prioritization of body parts that are important in gait recognition.

Another benefit of prioritizing important body parts is that it can also learn to ignore noisy joints. Since we use off the shelf pose estimation network trained in a different dataset, there is a possibility that it might not be able predict the joint positions accurately. If we assume that the network might provide wrong predictions for a specific set of joints, the attention block can learn to ignore those joints from all the training samples, resulting in better performance in the unseen test samples.

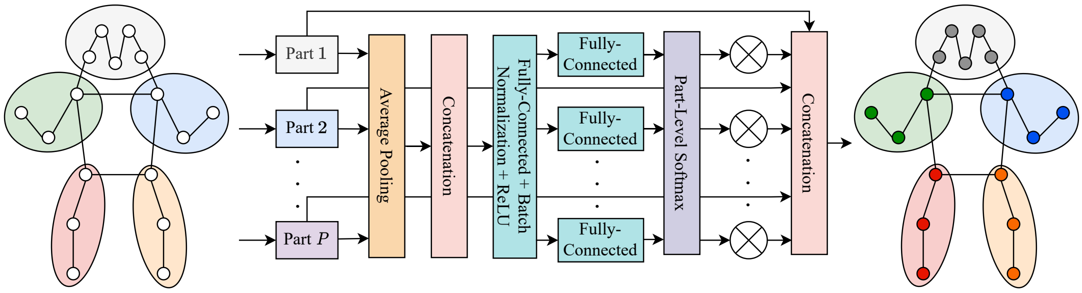


Figure 3.13: Part-wise Attention Module

As depicted in Figure 3.13, the joints are divided into five body parts to apply Part-wise Attention. The features of each part are then concatenated and passed through a sequence of average pooling in the temporal dimension, fully-connected, batch normalization, and ReLU layers. Finally, another fully-connected layer is used to calculate the attention of each part and multiplied with the feature values.

3.4 Class Label Generation

Following the convention recommended by [103], a set of gait sequences are preserved as gallery set and another set of gait sequences are preserved as probe set. The activation map generated by a sequence in the probe set is matched with the activation map generated by all the sequences of the gallery set. Then considering the closest normalized distance between the probe sequence with the sequences from the gallery set, the class label of the probe sequence is determined. The process is shown in Figure 3.14.

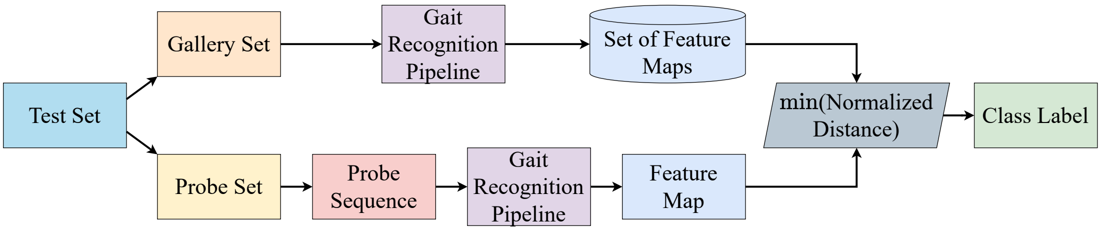


Figure 3.14: Generating class label for probe sequence from the test set

Chapter 4

Results and Discussion

4.1 Dataset

To evaluate the performance of the proposed pipeline, we utilize one of the largest and most popular gait recognition datasets, CASIA-B [159]. The dataset provides RGB videos of 124 subjects (93 male, 31 female). For each subject, 10 videos are provided considering three walking conditions: 6 of them captured the subject walking normally (NM), 2 of them captured the subject walking while carrying a bag (BG), and 2 of them captured the subject walking while heavily clothed (CL). Each video sequence is captured simultaneously from 11 angles, namely: $0°$, $18°$, $36°$, $54°$, $72°$, $90°$, $108°$, $126°$, $144°$, $162°$, $180°$. The variation in appearance and camera angle are depicted in Figure 4.1.

4.2 Experimental Setup

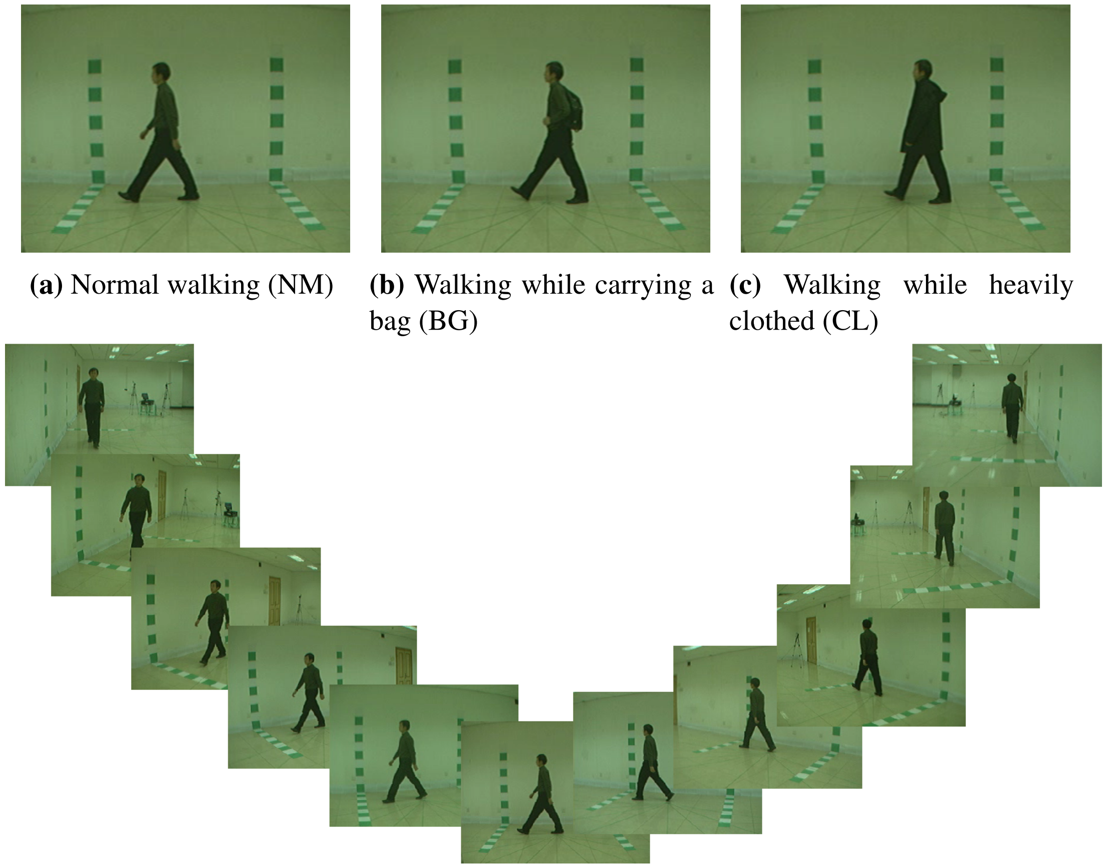
In this section, we discuss the experimental environment and different hyperparameters related to the training and testing process of our gait recognition pipeline.

4.2.1 Environment

The proposed system was implemented in Google Colab using PyTorch framework. The environment provides an NVIDIA Tesla T4 GPU with a VRAM of 11 GB and an Intel Xeon CPU with a base clock speed of 2.3 GHz. The total usable memory of the system was 13 GB.

4.2.2 Dataset Split

Due to the lack of any official split, the dataset split recommended by [17] was used. The first 74 subjects were kept in the training set and the rest in the test set. Further,



(a) Normal walking (NM)

(b) Walking while carrying a bag (BG)

(c) Walking while heavily clothed (CL)

(d) Eleven different viewing angles (0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, 180°)

Figure 4.1: Variations in appearance and camera angles in the gait sequences of CASIA-B dataset.

the test set was divided into gallery set and probe set. The gallery set contains the first 4 clips in normal walking conditions (NM01 - NM04). The remaining clips (NM05 - NM06, CL01 - CL02, BG01 - BG02) were kept in the probe set. The GCN architecture extracted features from both gallery sets and probe sets. Then the features of the probe set were matched with the features from the gallery set to determine the most similar sample as the class label. The test results were reported considering the accuracy averaging over 11 gallery views excluding the identical views.

4.2.3 Augmentation

Data augmentation techniques were used to increase the number of samples and make the model robust against various noises. The video frames were inverted to synthesize the effect of walking in reverse. Mirroring with respect to the vertical axis mimicked the situation of the subject walking in the opposite direction. To make the model resistant to pose estimation inaccuracies, small Gaussian Noise was added to each joint. These augmentations were performed only in the training set during runtime.

4.2.4 Batch Size

The batch size was chosen as 128 following the mini-batch gradient descent technique [160]. This often works as a regularizer helping to reduce the generalization error due to the noise introduced by a small number of samples in each batch. Additionally, due to the small size of the skeleton joint data, we were able to easily fit the training samples into memory.

4.2.5 Epoch and Learning Rate

The system was trained in 4 cycles each consisting of 100 epochs. The learning rate (LR) was set to 0.01 at the beginning to ensure rapid learning and divergence from local minima. However, it is often recommended to reduce the LR over time to ensure that the learning does not stagnate and regularize the learning process [161]. For this reason, once each cycle was completed, LR was reduced down to 10% of the previous cycle.

4.2.6 Optimizer and Loss Function

To reduce the effect of noise generated due to the estimation of joints, Adam optimizer [162] is used. It can help models converge faster.

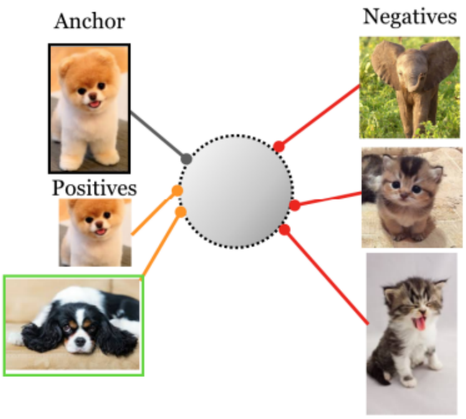


Figure 4.2: Demonstration of the effect of Supervised Contrastive Loss. (Adapted from [163])

Supervised Contrastive Loss [163], which is an extension of self-supervised batch contrastive loss, was used to calculate the loss. As shown in Figure 4.2, it can introduce a normalization effect by reducing the intra-class distance and increasing the inter-class distance in the embedding space. The loss is given by:

$$L^{sup} = \sum_{i=1} 2N L_i^{sup} \tag{4.1}$$

$$L_i^{sup} = \frac{-1}{2N_{\tilde{y}_i} - 1} \sum_{j=1}^{2N} 1_{i \neq j} \cdot 1_{\tilde{y}_i = \tilde{y}_j} \cdot \log \frac{\exp(z_i \cdot z_j / \tau)}{\sum_{k=1}^{2N} 1_{i \neq k} \cdot \exp(z_i \cdot z_k / \tau)} \tag{4.2}$$

Here,
$N_{\tilde{y}_i}$ = The number of images in the minibatch from the same class
$\tilde{y}_i$ = The base anchor
τ = A scalar temperature parameter
z_i = Projection of feature vector representing sample i in embedding space

4.2.7 Baseline Architecture

To compare the performance of our model, we implemented Gaitgraph [23] as our baseline since it is the most similar to our model in terms of the backbone network: ResGCN. Their implementation and pretrained weights are available in GitHub [1].

4.3 Hyperparameter Tuning

In this section, we discuss the experiments that were performed to determine specific hyperparameter values for our pipeline.

4.3.1 Confidence Threshold, C

To determine the threshold C for removing low confidence frame, we evaluated the average accuracy of the baseline architecture based on different thresholds.

Table 4.1: Performance of the baseline architecture for different values of C. The values under the Δ column denotes the increase/decrease in accuracy compared to the baseline.

Threshold	Frame	Accuracy (%)	Δ
0 (Baseline)	508211	69.89	0
50	507942	71.57	1.68
60	505410	**71.59**	**1.70**
70	500476	71.54	1.65
80	463773	66.72	−3.17

[1] https://github.com/tteepe/GaitGraph

According to Table 4.1, having a 60% confidence threshold increased the accuracy by the highest amount. Similar accuracy can be obtained by keeping 50% threshold, however it increased the number of frames that are used for making inference. To achieve the highest accuracy, while also keeping the number of frames minimum, we removed any frame that had an average confidence below 60%. In this regard, it is worth mentioning that preprocessing in this manner removed only 0.9% frames from the dataset.

4.3.2 DropGraph Neighborhood Threshold, K

To determine the number of neighbors to be dropped along with the selected base node, we evaluated the average accuracy of the baseline architecture based on different values of K in both spatial and temporal dimensions. It is worth-mentioning that a combination of both spatial and temporal DropGraph was used as regularizer during the training phase of ResGCN.

Spatial DropGraph

For each frame, HRNet can extract 17 joint positions. Hence, we considered the values of K to be 0, 1, 2, 3, and 4 in the spatial dimension.

Table 4.2: Performance of the architecture for different values of K in the spatial dimension. Here, the values under the Δ column denote the increase/decrease in accuracy compared to the network with no dropout.

K	Accuracy (%)	Δ
No dropout	79.67	0
0	79.86	0.19
1	80.79	1.12
2	80.31	0.64
3	80.04	0.37
4	79.89	0.22

As seen in Table 4.2, for $K = 0$ in the spatial dimension, the DropGraph technique is equivalent to Dropout as only a single node is dropped, which was not much effective. Dropping 1 neighbor resulted in effective regularization, whereas setting $K > 1$ may have caused too strong regularization. For this reason, the value of K was set to 1 for the spatial dimension. That means, in each frame, with a small probability, a root vertex and its 1-hop neighbors were dropped.

Temporal DropGraph

The gait sequences in the training set of CASIA-B dataset consists of 93 frames on average. However, the smallest video sequence consists of 30 frames. That means, on the temporal dimension, we have at least 30 nodes. Considering the small size in the temporal dimension, we considered the values of K to be 0, 1, 2, 3, and 4.

Table 4.3: Performance of the architecture for different values of K in the spatial dimension. Here, the values under the Δ column denote the increase/decrease in accuracy compared to the network with no dropout.

K	Accuracy (%)	Δ
No dropout	79.67	0
0	79.86	0.19
1	80.09	0.48
2	80.35	0.68
3	80.98	1.31
4	80.55	0.88

As shown in Table 4.3, for $K = 0$, the temporal DropGraph is equivalent to Dropout as only single node is dropped, which was not very effective. The effectiveness of DropGraph increased up to $K = 3$, which then decreases. For this reason, we set the value of K to 3 for the temporal dimension. That means, for each gait sequence, with a small probability, a root vertex and its 3-hop neighbors in earlier and later frames were dropped.

4.4 Ablation Study

In this section, we evaluate the performance of each individual module of our pipeline to understand their effect on the overall recognition performance.

4.4.1 Effect of Preprocessing

To evaluate the performance of our overall preprocessing technique, we compared the performance of the baseline architecture with vanilla dataset consisting of poses extracted using HRNet and our preprocessed dataset. Compared to the baseline shown in Table 4.4, our preprocessing techniques were able to improve the performance of the network in all aspects. Specifically, considering the BG and CL walking conditions are more prone to inaccurate estimation of poses, the improvement of performance was more prominent compared to that of NM.

Table 4.4: Effect of Preprocessing on CASIA-B Dataset

Dataset	Accuracy (%)					
	NM	Δ	BG	Δ	CL	Δ
Vanilla	87.5	0	75.3	0	67.1	0
Preprocessed	88.9	1.4	77.6	2.3	69.7	2.6

The reduction of noisy and inaccurate frames resulted in an increase in accuracy by 1.67% on average. The highest increase in accuracy is seen in CL conditions denoting, in most of the cases, HRNet struggled to estimate poses correctly when the subject is heavily clothed.

4.4.2 Effect of Hop Extraction

Replacing the higher-order polynomials with the hop extraction technique to capture the long-distance relationships among the joints resulted in an overall increase in accuracy on average. As seen in Table 4.5, in all walking conditions, the significant increase goes to prove the superiority of this technique.

Table 4.5: Effect of Hop-Extraction on CASIA-B Dataset

Strategy	Accuracy (%)					
	NM	Δ	BG	Δ	CL	Δ
Preprocessed	88.9	0	78.2	0	69.8	0
Hop Extraction	93.3	4.4	87.5	9.3	82.3	12.5

In fact, among all the techniques applied, Hop Extraction demonstrated the maximum increase in terms of accuracy. This goes to show the importance of providing equal priority to the relationship among further and closer joints in gait recognition.

4.4.3 Effect of DropGraph

To understand the overfitting issue, we considered the training and testing accuracy of the model in the above-mentioned scenarios (Table 4.6).

Even after careful consideration in the training process via dynamic learning rate adjustment through each individual cycles, a 6% difference between the training accuracy and test accuracy was noticed. Addition of DropGraph to the model decreased the training accuracy by 0.3%, but it increased the test accuracy by 1.9%. That means, due to the introduction of the DropGraph, the model was able to learn more generalized features to perform better on unseen dataset.

Table 4.6: Comparison of training and test accuracy after integrating DropGraph into our model on CASIA-B dataset

Strategy	Accuracy (%)			
	Train	Δ	Test	Δ
Hop Extraction	85.7	0	79.7	0
DropGraph	85.4	-0.3	81.6	1.9

DropGraph works as a regularizer by reducing overfitting and increasing the generalization capability of the model. The test accuracy is further expressed in detail in Table 4.7. It is evident from the table that DropGraph increased the accuracy of the model by allowing it to learn general features to better classify unseen samples.

Table 4.7: Effect of DropGraph on CASIA-B Dataset

Strategy	Accuracy (%)					
	NM	Δ	BG	Δ	CL	Δ
Hop Extraction	93.3	0	87.5	0	82.3	0
DropGraph	94.1	0.8	89.9	2.4	85.2	2.9

4.4.4 Effect of Part-wise Attention

Part-wise attention block was added to identify specific body parts that are helpful in identifying people. It does so by learning to generate a higher weight for important parts of the body.

Table 4.8: Effect of Part-wise Attention on CASIA-B Dataset

Strategy	Accuracy (%)					
	NM	Δ	BG	Δ	CL	Δ
DropGraph	94.1	0	89.9	0	85.2	0
Part-wise Attention	96.5	2.4	93.0	3.1	90.1	4.6

As seen in Table 4.8, introduction of part-wise attention block resulted in higher accuracy in gait recognition. Compared to the NM walking condition, the accuracy improvement was higher in BG and CL conditions. This can be attributed to the fact that the model already has a high accuracy in NM considering the joint pose estimated by the pose estimation model. However, since there are certain noises introduced due to the occlusion and appearance in BG and CL conditions, the attention block helps the network to determine a specific portion of the body that can still be used to identify gait in a better way.

4.5 Comparison with State-of-the-Art Models

4.5.1 Joint position-based Approaches

Table 4.9 compares the performance of the proposed system with existing state-of-the-art models in gait recognition. Our system improved the accuracy by a commendable margin in varying viewing angles and walking conditions. Even in the case where the performance of our model does not exceed the state-of-the-art (NM walking condition), the performance is comparable with the state-of-the-art.

Table 4.9: Performance comparison with the state-of-the-art models for joint position-based gait recognition.

Gallery NM#1-4		Viewing angles											Mean
Probe	Ref.	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM#5-6	PTSN [18]	34.5	45.6	49.6	51.3	52.7	52.3	53	50.8	52.2	48.3	31.4	47.4
	PTSN-3D [141]	38.7	50.2	55.9	56	56.7	54.6	54.8	56	54.1	52.4	40.2	51.8
	PoseGait [19]	55.3	69.6	73.9	75	68	68.2	71.1	72.9	76.1	70.4	55.4	68.7
	JointsGait [22]	68.1	73.6	77.9	76.4	77.5	79.1	78.4	76	69.5	71.9	70.1	74.4
	GaitGraph2 [130]	78.5	82.9	85.8	85.6	83.1	81.5	84.3	83.2	84.2	81.6	71.8	82.0
	Gaitgraph [23]	85.3	88.5	91	92.5	87	86.5	88.4	89.2	87.9	85.9	81.9	87.6
	Gait-D [125]	87.7	92.5	93.6	95.7	93.3	92.4	92.8	93.4	90.6	88.6	87.3	91.6
	MS-Gait [25]	89.4	91.7	91.6	90.2	90.6	90.6	90.4	90.9	90.4	88.5	85.6	90.0
	RGCNN [128]	94.8	**98.5**	**96.9**	**98.3**	**96.8**	**98.9**	96.9	**98.8**	**97.9**	93.9	**95.9**	**97.0**
	Ours	**95.2**	95.7	96.6	96.8	96.5	97.6	**97.2**	97.2	96.0	**97.1**	95.3	96.5
BG#1-2	PTSN [18]	22.4	29.8	29.6	29.2	32.5	31.5	32.1	31	27.3	28.1	18.2	28.3
	PTSN-3D [141]	27.7	32.7	37.4	35	37.1	37.5	37.7	36.9	33.8	31.8	27	34.1
	PoseGait [19]	35.3	47.2	52.4	46.9	45.5	43.9	46.1	48.1	49.4	43.6	31.1	44.5
	JointsGait [22]	54.3	59.1	60.6	59.7	63	65.7	62.4	59	58.1	58.6	50.1	59.1
	GaitGraph2 [130]	69.9	75.9	78.1	79.3	71.4	71.7	74.3	76.2	73.2	73.4	61.7	73.2
	Gaitgraph [23]	75.8	76.7	75.9	76.1	71.4	73.9	78	74.7	75.4	75.4	69.2	74.8
	Gait-D [125]	78.2	80.1	79.3	80.2	78.4	77.6	80.4	78.6	79.1	80.2	76.5	79.0
	MS-Gait [25]	75.7	84.8	83.7	83.2	80.6	80.1	82.2	79.8	79.1	75.9	71.1	79.7
	RGCNN [128]	88.8	**93.8**	91.5	88.5	91.6	90.0	91.9	**91.4**	92.4	89.7	89.0	90.8
	Ours	**92.4**	92.4	**95.1**	**93.7**	**92.5**	**94.1**	**93.9**	90.3	**93.1**	**93.1**	**92.0**	**93.0**
CL#1-2	PTSN [18]	14.2	17.1	17.6	19.3	19.5	20.0	20.1	17.3	16.5	18.1	14	17.6
	PTSN-3D [141]	15.8	17.2	19.9	20	22.3	24.3	28.1	23.8	20.9	23	17	21.1
	PoseGait [19]	24.3	29.7	41.3	38.8	38.2	38.5	41.6	44.9	42.2	33.4	22.5	35.9
	JointsGait [22]	48.1	46.9	49.6	50.5	51	52.3	49.0	46.0	48.7	53.6	52	49.8
	GaitGraph2 [130]	57.1	61.1	68.9	66.0	67.8	65.4	68.1	67.2	63.7	63.6	50.4	63.6
	Gaitgraph [23]	69.6	66.1	68.8	67.2	64.5	62	69.5	65.6	65.7	66.1	64.3	66.3
	Gait-D [125]	73.2	71.7	75.4	73.2	74.6	72.3	74.1	70.5	69.4	71.2	66.7	72.0
	MS-Gait [25]	75.1	79.7	80.5	84.7	84.0	82.4	79.8	80.4	78.3	78.0	70.9	79.4
	RGCNN [128]	87.9	**91.1**	90.9	89.8	88.8	89.8	89.2	**90.0**	**91.0**	**91.0**	**89.4**	**89.9**
	Ours	**88.9**	88.3	**93.2**	**90.9**	**91.6**	**92.0**	**90.2**	89.8	89.2	89.4	87.1	**90.1**

4.5.2 Appearance-based Approaches

Even though appearance-based methods still achieve the better result in gait recognition compared to model-based approaches, our system took a huge step to close this gap. Additionally, we achieved a higher average accuracy in CASIA-B dataset compared to the appearance-based methods. A comparison can be seen in Table 4.10.

Table 4.10: Performance comparison with state-of-the-art models for appearance-based gait recognition

Ref.	Accuracy (%)		
	NM	BG	CL
GaitNet [102]	91.6	85.7	58.9
GaitSet [103]	95.0	87.2	70.4
GaitRNNPart [105]	95.2	89.7	74.7
GaitPart [106]	96.2	91.5	78.7
GLN [136]	96.9	94.0	77.5
SRN + CBlock [138]	97.5	**94.3**	77.7
HMRNet [137]	**97.9**	93.1	77.6
3DCNN [122]	96.7	93.0	81.5
GaitGL [123]	96.4	92.7	83.0
SRN [138]	97.1	94.0	81.8
Multi3D [124]	97.6	94.1	81.2
Vi-GaitGL [139]	96.2	92.9	87.2
Ours	96.5	93.0	**90.1**

One key advantage of joint position-based methods is that they are invariant to walking condition. As evident from the table, our method performed significantly better than appearance-based approaches when the subject is walking wearing heavy clothes. In other walking conditions, our model provided comparable performance to the state-of-the-art appearance-based methods.

Chapter 5

Conclusion

In this dissertation, we presented a gait recognition pipeline consisting of effective preprocessing, and a robust and generalized feature extractor based on a graph convolutional network. The joint-position based features extracted from the video sequences of gait are robust to variations in background, camera angle, and appearance of the subject. The preprocessing technique can enhance the performance of the overall pipeline by removing noisy and inaccurate frames. The Graph Convolutional Network (GCN) is able to extract multi-stream features considering joint, bone, and motion information extracted from gait sequences. The bone and motion information can provide the network with better understanding of the correlation between joints in the same frame and joints in consecutive frames, respectively. The extracted features succinctly capture the relationship among close and further joints giving them equal priorities via the hop extraction technique. The technique also alleviates the problem of considering redundant relationships. The network can determine important body parts and prioritize them in identifying gait using part-wise attention module. The module can also help the GCN to learn to ignore wrongly estimated pose. The network can avoid overfitting to training samples due to DropGraph technique used during the training process and extract meaningful generalized features to detect unseen gait samples easily. It allows the model to have a higher test accuracy without sacrificing much of the training accuracy. By combining all these techniques, we are able to achieve state-of-the-art performance among joint-position based and appearance-based approaches in challenging conditions.

Despite that, there are several scopes for improvement. Since the video sequences in the CASIA-B dataset are captured in an indoor environment, the efficacy of the network in real-life scenarios still remains an open question. Future studies can focus on creating real-life datasets large enough to train graph convolutional networks considering a variety in background and appearances. Further, considering the recent advances in pose estimation techniques, the performance of other techniques considering different number of joint positions can be evaluated to understand whether increasing/decreasing the number of joints considered can affect the overall recognition performance or not. Better pose estimation techniques can also provide noise-free, accurate joint position coordinates. Again, different body parts consisting of different sets of joints can be considered to identify the optimal configuration for part-wise attention module. To tackle the effect of poor pose estimation and variance in appearance simultaneously, an ensemble of joint position-based and appearance-based techniques can be explored. Finally, a full-fledged system can developed utilizing the pipeline to evaluate the performance of the system in real-life scenarios.

REFERENCES

[1] Physiopedia, “Gait — Physiopedia,” 2022, [Online; accessed 25-April-2022]. [Online]. Available: https://www.physio-pedia.com/index.php?title=Gait&oldid=295011

[2] P. Connor and A. Ross, “Biometric recognition by gait: A survey of modalities and features,” *Computer Vision and Image Understanding*, vol. 167, pp. 1–27, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314218300079

[3] C. Wan, L. Wang, and V. V. Phoha, “A Survey on Gait Recognition,” *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–35, 2018. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3230633

[4] A. Nambiar, A. Bernardino, and J. C. Nascimento, “Gait-based Person Re-identification: A Survey,” *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–34, 2019. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3243043

[5] M. Taiana, D. Figueira, A. Nambiar, J. Nascimento, and A. Bernardino, “Towards fully automated person re-identification,” in *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 3. IEEE, 2014, pp. 140–147. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7295073

[6] A. M. Nambiar, A. Bernardino, J. C. Nascimento, and A. L. Fred, “Towards View-point Invariant Person Re-identification via Fusion of Anthropometric and Gait Features from Kinect Measurements.” in *Visigrapp (5: Visapp)*, 2017, pp. 108–119. [Online]. Available: https://www.scitepress.org/Link.aspx?doi=10.5220/0006165301080119

[7] A. Muro-De-La-Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, “Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications,” *Sensors*, vol. 14, no. 2, pp. 3362–3394, 2014. [Online]. Available: https://www.mdpi.com/1424-8220/14/2/3362

[8] T. T. Verlekar, P. L. Correia, and L. D. Soares, “Using transfer learning for classification of gait pathologies,” in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 2376–2381. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8621302

[9] D. Jarchi, J. Pope, T. K. Lee, L. Tamjidi, A. Mirzaei, and S. Sanei, “A Review on Accelerometry-Based Gait Analysis and Emerging Clinical Applications,” *IEEE reviews in biomedical engineering*, vol. 11, pp. 177–194, 2018. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8293814

[10] S. Aghanavesi, J. Westin, F. Bergquist, D. Nyholm, H. Askmark, S. M. Aquilonius, R. Constantinescu, A. Medvedev, J. Spira, F. Ohlsson *et al.*, “A multiple motion sensors index for motor state quantification in Parkinson’s disease,” *Computer Methods and Programs in Biomedicine*, vol. 189, p. 105309, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260719319418

[11] A. S. Alharthi, A. J. Casson, and K. B. Ozanyan, “Gait Spatiotemporal Signal Analysis for Parkinson’s Disease Detection and Severity Rating,” *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1838–1848, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9171856

[12] G. Cicirelli, D. Impedovo, V. Dentamaro, R. Marani, G. Pirlo, and T. D’Orazio, “Human Gait Analysis in Neurodegenerative Diseases: A Review,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 229–242, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9466394

[13] J. M. Echterhoff, J. Haladjian, and B. Brügge, “Gait and jump classification in modern equestrian sports,” in *Proceedings of the 2018 ACM International Symposium on Wearable Computers*. Acm, 2018, pp. 88–91. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3267242.3267267

[14] H. Zhang, Y. Guo, and D. Zanotto, “Accurate Ambulatory Gait Analysis in Walking and Running Using Machine Learning Models,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 191–202, 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8930581

[15] S. Ali Etemad and A. Arya, “Classification and translation of style and affect in human motion using RBF neural networks,” *Neurocomputing*, vol. 129, pp. 585–595, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231213008588

[16] S. A. Etemad and A. Arya, “Expert-Driven Perceptual Features for Modeling Style and Affect in Human Motion,” *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 4, pp. 534–545, 2016. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7446325

[17] A. Sepas-Moghaddam and A. Etemad, “Deep Gait Recognition: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9714177

[18] R. Liao, C. Cao, E. B. Garcia, S. Yu, and Y. Huang, “Pose-Based Temporal-Spatial Network (PTSN) for Gait Recognition with Carrying and Clothing Variations,” in *Biometric Recognition*, J. Zhou, Y. Wang, Z. Sun, Y. Xu, L. Shen, J. Feng, S. Shan, Y. Qiao, Z. Guo, and S. Yu, Eds. Cham: Springer International Publishing, 2017, pp. 474–483. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-69923-3_51

[19] R. Liao, S. Yu, W. An, and Y. Huang, “A model-based gait recognition method with body pose and human prior knowledge,” *Pattern Recognition*, vol. 98, p. 107069, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003132031930370X

[20] A. Sokolova and A. Konushin, “Pose-based deep gait recognition,” *IET Biometrics*, vol. 8, no. 2, pp. 134–143, 2018. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-bmt.2018.5046

[21] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein, “Geometric Deep Learning on Graphs and Manifolds Using Mixture Model CNNs,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5115–5124. [Online]. Available: https://ieeexplore.ieee.org/document/8100059

[22] N. Li, X. Zhao, and C. Ma, “A model-based Gait Recognition Method based on Gait Graph Convolutional Networks and Joints Relationship Pyramid Mapping,” *CoRR*, vol. abs/2005.08625, 2020. [Online]. Available: https://arxiv.org/abs/2005.08625

[23] T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, “Gaitgraph: Graph Convolutional Network for Skeleton-Based Gait Recognition,” in *2021 IEEE International Conference on Image Processing (ICIP)*. Anchorage, Alaska, USA: IEEE, 2021, pp. 2314–2318. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9506717

[24] T. N. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” in *5th International Conference on Learn-*

ing Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. OpenReview.net, 2017. [Online]. Available: https://openreview.net/forum?id=SJU4ayYgl

[25] L. Wang, J. Chen, Z. Chen, Y. Liu, and H. Yang, "Multi-stream part-fused graph convolutional networks for skeleton-based gait recognition," *Connection Science*, vol. 34, no. 1, pp. 652–669, 2022. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/09540091.2022.2026294

[26] Aristotle, *On the gait of animals*. Kessinger Publishing, LLC, 350 Bc. [Online]. Available: http://classics.mit.edu/Aristotle/gait_anim.htmll

[27] G. A. Borelli, *On the Movement of Animals*. Springer Berlin, Heidelberg, 1679.

[28] E.-J. Marey, "LOCOMOTION: Étienne-Jules MAREY via Benjamin GODARD (1880s)." [Online]. Available: https://www.youtube.com/watch?v=gsTabC5d7zU

[29] W. Weber, E. Weber, and C. Lawrence, "Mechanics of the Human Walking Apparatus," *Annals of Science*, vol. 51, no. 2, 1836.

[30] E. Muybridge, "Race Horse First Film Ever 1878 Eadweard Muybridge." [Online]. Available: https://www.youtube.com/watch?v=IEqccPhsqgA

[31] J. B. Saunders, V. T. Inman, and H. D. Ebergart, "The major determinants in normal and pathological gait," *The Journal of Bone & Joint Surgery*, vol. 35, no. 3, pp. 543–558, 1953. [Online]. Available: https://journals.lww.com/jbjsjournal/Abstract/1953/35030/THE_MAJOR_DETERMINANTS_IN_NORMAL_AND_PATHOLOGICAL.3.aspx

[32] D. H. Sutherland and J. L. Hagy, "Measurement of gait movements from motion picture film," *The Journal of Bone & Joint Surgery*, vol. 54, no. 4, pp. 787–797, 1972. [Online]. Available: https://journals.lww.com/jbjsjournal/Abstract/1972/54040/Measurement_of_Gait_Movements_from_Motion_Picture.9.aspx

[33] J. Perry, "Clinical Gait Analyzer," *Bulletins of Prosthetic Research*, pp. 188–192, 1974. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/4462898/

[34] M. P. Murray, A. B. Drought, and R. C. Kory, "Walking patterns of normal men," *The Journal of Bone & Joint Surgery*, vol. 46, no. 2, pp. 335–360, 1964. [Online]. Available: https://journals.lww.com/jbjsjournal/Abstract/1964/46020/Walking_Patterns_of_Normal_Men.9.aspx

[35] M. P. Murray, "Gait as a Total Pattern of Movement: Including a Bibliography on Gait," *American Journal of Physical Medicine &*

Rehabilitation, vol. 46, no. 1, pp. 290–333, 1967. [Online]. Available: https://journals.lww.com/ajpmr/Citation/1967/02000/Gait_As_A_Total_Pattern_of_Movement__Including_A.26.aspx

[36] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception & Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973. [Online]. Available: https://link.springer.com/article/10.3758/BF03212378

[37] J. E. Cutting and L. T. Kozlowski, "Recognizing friends by their walk: Gait perception without familiarity cues," *Bulletin of the Psychonomic Society*, vol. 9, no. 5, pp. 353–356, 1977. [Online]. Available: https://link.springer.com/article/10.3758/BF03337021

[38] J. E. Cutting, D. R. Proffitt, and L. T. Kozlowski, "A biomechanical invariant for gait perception." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4, no. 3, p. 357, 1978. [Online]. Available: https://psycnet.apa.org/record/1980-00173-001

[39] S. A. Niyogi and E. H. Adelson, "Analyzing and recognizing walking figures in XYT," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 469–474. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/323868

[40] M. Addlesee, A. Jones, F. Livesey, and F. Samaria, "The ORL active floor [sensor system]," *IEEE Personal Communications*, vol. 4, no. 5, pp. 35–41, 1997. [Online]. Available: https://ieeexplore.ieee.org/document/626980

[41] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S.-M. Makela, and H. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 2, 2005, pp. ii/973–ii/976 Vol. 2. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1415569

[42] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316–322, 2005. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1561189

[43] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A Full-Body Layered Deformable Model for Automatic Model-Based Gait Recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1–13, 2007. [Online]. Available: https://link.springer.com/article/10.1155/2008/261317

[44] H. Murase and R. Sakai, "Moving object recognition in eigenspace representation: gait analysis and lip reading," *Pattern Recognition Letters*, vol. 17, no. 2, pp. 155–162, 1996. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0167865595001093

[45] P. Huang, C. Harris, and M. Nixon, "Recognising humans by gait via parametric canonical space," *Artificial Intelligence in Engineering*, vol. 13, no. 4, pp. 359–366, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0954181099000084

[46] C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis, "EigenGait: Motion-Based Recognition of People Using Image Self-Similarity," in *Audio- and Video-Based Biometric Person Authentication*, J. Bigun and F. Smeraldi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 284–294. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-45344-X_42

[47] R. T. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*. IEEE, 2002, pp. 366–371. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1004181

[48] Z. Liu and S. Sarkar, "Improved gait recognition by gait dynamics normalization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 863–876, 2006. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1624352

[49] 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. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1374864

[50] A. Y. Johnson and A. F. Bobick, "A Multi-view Method for Gait Recognition Using Static Body Parameters," in *International Conference on Audio-and Video-Based Biometric Person Authentication*. Springer, 2001, pp. 301–311. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-45344-x_44

[51] C. BenAbdelkader, R. Cutler, and L. Davis, "View-invariant Estimation of Height and Stride for Gait Recognition," in *International Workshop on Biometric Authentication*. Springer, 2002, pp. 155–167. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-47917-1_16

[52] A. Kale, A. Rajagopalan, N. Cuntoor, and V. Kruger, “Gait-based recognition of humans using continuous HMMs,” in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*. IEEE, 2002, pp. 336–341. [Online]. Available: https://ieeexplore.ieee.org/document/1004176

[53] P. C. Cattin, “Biometric authentication system using human gait,” Doctoral Thesis, ETH Zürich, Rämistrasse, Zürich, Switzerland, 2002. [Online]. Available: https://www.research-collection.ethz.ch/handle/20.500.11850/146786

[54] N. Cuntoor, A. Kale, and R. Chellappa, “Combining multiple evidences for gait recognition,” in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP ’03).*, vol. 3. IEEE, 2003, pp. Iii–33. [Online]. Available: https://ieeexplore.ieee.org/document/1199100

[55] J. P. Foster, M. S. Nixon, and A. Prügel-Bennett, “Automatic gait recognition using area-based metrics,” *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2489–2497, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0167865503000941

[56] S. D. Mowbray and M. S. Nixon, “Automatic Gait Recognition via Fourier Descriptors of Deformable Objects,” in *International Conference on Audio-and Video-Based Biometric Person Authentication*. Springer, 2003, pp. 566–573. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-44887-x_67

[57] L. Wang, T. Tan, H. Ning, and W. Hu, “Silhouette analysis-based gait recognition for human identification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1505–1518, 2003. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1251144

[58] L. Wang, T. Tan, W. Hu, and H. Ning, “Automatic gait recognition based on statistical shape analysis,” *IEEE transactions on image processing*, vol. 12, no. 9, pp. 1120–1131, 2003. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1221765

[59] S. Yu, L. Wang, W. Hu, and T. Tan, “Gait analysis for human identification in frequency domain,” in *Third International Conference on Image and Graphics (ICIG’04)*. IEEE, 2004, pp. 282–285. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1410440

[60] S. Hong, H. Lee, I. F. Nizami, and E. Kim, “A New Gait Representation for Human Identification: Mass Vector,” in *2007 2nd IEEE Conference on Industrial Electronics and Applications*. IEEE, 2007, pp. 669–673. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4318491

[61] S. Lee, Y. Liu, and R. Collins, "Shape Variation-Based Frieze Pattern for Robust Gait Recognition," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4270163

[62] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Speed-invariant gait recognition based on Procrustes Shape Analysis using higher-order shape configuration," in *2011 18th IEEE International Conference on Image Processing*. IEEE, 2011, pp. 545–548. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6116403

[63] H. El-Alfy, I. Mitsugami, and Y. Yagi, "A New Gait-Based Identification Method Using Local Gauss Maps," in *Asian Conference on Computer Vision (ACCV)*. Springer, 2014, pp. 3–18. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-16628-5_1

[64] P. J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "The gait identification challenge problem: Data sets and baseline algorithm," in *Object recognition supported by user interaction for service robots*, vol. 1. IEEE, 2002, pp. 385–388. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1044731

[65] J. Han and B. Bhanu, "Statistical feature fusion for gait-based human recognition," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2. IEEE, 2004, pp. II–II. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1315252

[66] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General Tensor Discriminant Analysis and Gabor Features for Gait Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700–1715, 2007. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4293202

[67] R. D. Seely, S. Samangooei, M. Lee, J. N. Carter, and M. S. Nixon, "The University of Southampton Multi-Biometric Tunnel and introducing a novel 3D gait dataset," in *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*. IEEE, 2008, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4699353

[68] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition," *IEEE Transactions on Information Forensics*

and Security, vol. 7, no. 5, pp. 1511–1521, 2012. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6215042

[69] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, and K. Moustakas, "Gait Recognition Using Compact Feature Extraction Transforms and Depth Information," *IEEE Transactions on Information Forensics and security*, vol. 2, no. 3, pp. 623–630, 2007. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4291549

[70] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, "Gait energy volumes and frontal gait recognition using depth images," in *2011 International Joint Conference on Biometrics (IJCB)*. IEEE, 2011, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6117504

[71] M. Hofmann, S. Bachmann, and G. Rigoll, "2.5D gait biometrics using the Depth Gradient Histogram Energy Image," in *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 2012, pp. 399–403. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6374606

[72] K. Bashir, T. Xiang, and S. Gong, "Feature selection on Gait Energy Image for human identification," in *2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2008, pp. 985–988. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4517777

[73] C. Wang, J. Zhang, J. Pu, X. Yuan, and L. Wang, "Chrono-Gait Image: A Novel Temporal Template for Gait Recognition," in *European Conference on Computer Vision*. Springer, 2010, pp. 257–270. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-15549-9_19

[74] Y. Liu, J. Zhang, C. Wang, and L. Wang, "Multiple HOG templates for gait recognition," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 2012, pp. 2930–2933. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6460779

[75] M. Hofmann and G. Rigoll, "Improved Gait Recognition using Gradient Histogram Energy Image," in *2012 19th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2012, pp. 1389–1392. [Online]. Available: https://ieeexplore.ieee.org/document/6467128

[76] J. Little and J. Boyd, "Describing motion for recognition," in *Proceedings of International Symposium on Computer Vision-ISCV*. IEEE, 1995, pp. 235–240. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/477007

[77] ——, "Recognizing people by their gait: the shape of motion," *Videre: A Journal Of Computer Vision Research*, vol. 1, no. 2, pp. 1–32, 1998. [Online]. Available: https://www.cs.rochester.edu/~brown/Videre/001/abstracts/v1n2001.html

[78] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition.* IEEE, 2002, pp. 155–162. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1004148

[79] B. Bhanu and J. Han, "Human Recognition on Combining Kinematic and Stationary Features," in *International Conference on Audio-and Video-Based Biometric Person Authentication.* Springer, 2003, pp. 600–608. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-44887-x_71

[80] D. K. Wagg and M. S. Nixon, "On automated model-based extraction and analysis of gait," in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.* IEEE, 2004, pp. 11–16. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1301502

[81] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, "Self-Calibrating View-Invariant Gait Biometrics," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 4, pp. 997–1008, 2009. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/5299188

[82] J.-H. Yoo and M. S. Nixon, "Automated Markerless Analysis of Human Gait Motion for Recognition and Classification," *Etri Journal*, vol. 33, no. 2, pp. 259–266, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.4218/etrij.11.1510.0068

[83] R. Tanawongsuwan and A. Bobick, "Performance Analysis of Time-Distance Gait Parameters under Different Speeds," in *International Conference on Audio-and Video-Based Biometric Person Authentication.* Springer, 2003, pp. 715–724. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-44887-x_83

[84] I. Bouchrika and M. S. Nixon, "Gait recognition by dynamic cues," in *2008 19th International Conference on Pattern Recognition.* IEEE, 2008, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4760994

[85] G. Ariyanto and M. S. Nixon, "Model-based 3D gait biometrics," in *2011 IEEE International Joint Conference on Biometrics (IJCB).* IEEE, 2011, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6117582

[86] J. Preis, M. Kessel, M. Werner, and C. Linnhoff-Popien, “Gait recognition with kinect,” in *Proceedings of the First Workshop on Kinect in Pervasive Computing*. New Castle, UK, 2012, pp. 1–4.

[87] A. Świtoński, A. Polański, and K. Wojciechowski, “Human Identification Based on Gait Paths,” in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2011, pp. 531–542. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-23687-7_48

[88] ——, “Human identification based on the reduced kinematic data of the gait,” in *2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, 2011, pp. 650–655. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6046684

[89] T. Yu and J.-H. Zou, “Automatic Human Gait Imitation and Recognition in 3D from Monocular Video with an Uncalibrated Camera,” *Mathematical Problems in Engineering*, vol. 2012, 2012. [Online]. Available: https://www.hindawi.com/journals/mpe/2012/563864/

[90] J. Kovač and P. Peer, “Human Skeleton Model Based Dynamic Features for Walking Speed Invariant Gait Recognition,” *Mathematical Problems in Engineering*, vol. 2014, 2014. [Online]. Available: https://www.hindawi.com/journals/mpe/2014/484320/

[91] B. Dikovski, G. Madjarov, and D. Gjorgjevikj, “Evaluation of different feature sets for gait recognition using skeletal data from Kinect,” in *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2014, pp. 1304–1308. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6859769

[92] D. Cunado, M. S. Nixon, and J. N. Carter, “Using gait as a biometric, via phase-weighted magnitude spectra,” in *International Conference on Audio- and Video-Based Biometric Person Authentication*. Springer, 1997, pp. 93–102. [Online]. Available: https://link.springer.com/chapter/10.1007/BFb0015984

[93] ——, “Automatic extraction and description of human gait models for recognition purposes,” *Computer Vision and Image Understanding*, vol. 90, no. 1, pp. 1–41, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1077314203000080

[94] C. Yam, M. S. Nixon, and J. N. Carter, “Automated person recognition by walking and running via model-based approaches,” *Pattern recognition*, vol. 37, no. 5, pp. 1057–1072, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0031320303003996

[95] I. Bouchrika and M. S. Nixon, “Model-Based Feature Extraction for Gait Analysis and Recognition,” in *International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications.* Springer, 2007, pp. 150–160. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-540-71457-6_14

[96] E. Fix and J. L. Hodges, “Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties,” *International Statistical Review/Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238–247, 1989. [Online]. Available: https://www.jstor.org/stable/1403797?seq=1

[97] Wikipedia contributors, “K-nearest neighbors algorithm — Wikipedia, The Free Encyclopedia,” https://en.wikipedia.org/w/index.php?title=K-nearest_neighbors_algorithm&oldid=1077256005, 2022, [Online; accessed 29-April-2022].

[98] R. Tanawongsuwan and A. Bobick, “Gait recognition from time-normalized joint-angle trajectories in the walking plane,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 2. IEEE, 2001, pp. Ii–ii. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/991036

[99] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, “GEINet: View-invariant gait recognition using a convolutional neural network,” in *2016 international conference on biometrics (ICB).* IEEE, 2016, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7550060

[100] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, “A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 2, pp. 209–226, 2016. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7439821

[101] Y. Wang, B. Du, Y. Shen, K. Wu, G. Zhao, J. Sun, and H. Wen, “EV-Gait: Event-Based Robust Gait Recognition Using Dynamic Vision Sensors,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE, 2019, pp. 6351–6360. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8953966

[102] C. Song, Y. Huang, Y. Huang, N. Jia, and L. Wang, “GaitNet: An end-to-end network for gait based human identification,” *Pattern Recognition*, vol. 96, p. 106988, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0031320319302912

[103] H. Chao, Y. He, J. Zhang, and J. Feng, "GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01. PKP Publishing Services Network, 2019, pp. 8126–8133. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/4821

[104] Y. Zhang, Y. Huang, L. Wang, and S. Yu, "A comprehensive study on gait biometrics using a joint CNN-based method," *Pattern Recognition*, vol. 93, pp. 228–236, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320319301694

[105] A. Sepas-Moghaddam and A. Etemad, "View-Invariant Gait Recognition With Attentive Recurrent Learning of Partial Representations," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 1, pp. 124–137, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9229117

[106] C. Fan, Y. Peng, C. Cao, X. Liu, S. Hou, J. Chi, Y. Huang, Q. Li, and Z. He, "GaitPart: Temporal Part-Based Model for Gait Recognition," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 14 225–14 233. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9156784

[107] X. Li, Y. Makihara, C. Xu, Y. Yagi, S. Yu, and M. Ren, "End-to-End Model-Based Gait Recognition," in *Proceedings of the Asian conference on computer vision*, 2020. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-69535-4_1

[108] A. Sepas-Moghaddam, S. Ghorbani, N. F. Troje, and A. Etemad, "Gait Recognition using Multi-Scale Partial Representation Transformation with Capsules," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 8045–8052. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9412517

[109] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3422622

[110] Y. He, J. Zhang, H. Shan, and L. Wang, "Multi-Task GANs for View-Specific Feature Learning in Gait Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 1, pp. 102–113, 2018. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8374898

[111] B. Hu, Y. Gao, Y. Guan, Y. Long, N. D. Lane, and T. Ploetz, "Robust Cross-View Gait Identification with Evidence: A Discriminant Gait GAN (DiGGAN) Approach on 10000 People," *CoRR*, vol. abs/1811.10493, 2018. [Online]. Available: http://arxiv.org/abs/1811.10493

[112] Y. Wang, C. Song, Y. Huang, Z. Wang, and L. Wang, "Learning view invariant gait features with Two-Stream GAN," *Neurocomputing*, vol. 339, pp. 245–254, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0925231219302395

[113] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/2cad8fa47bbef282badbb8de5374b894-Abstract.html

[114] Z. Xu, W. Lu, Q. Zhang, Y. Yeung, and X. Chen, "Gait recognition based on capsule network," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 159–167, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1047320319300318

[115] A. Zhao, J. Li, and M. Ahmed, "SpiderNet: A spiderweb graph neural network for multi-view gait recognition," *Knowledge-Based Systems*, vol. 206, p. 106273, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0950705120304597

[116] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6795963

[117] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches," in *Proceedings of SSST@EMNLP 2014, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, 25 October 2014*, D. Wu, M. Carpuat, X. Carreras, and E. M. Vecchi, Eds. Association for Computational Linguistics, 2014, pp. 103–111. [Online]. Available: https://aclanthology.org/W14-4012/

[118] D. Liu, M. Ye, X. Li, F. Zhang, and L. Lin, "Memory-based Gait Recognition," in *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016*, R. C. Wilson, E. R. Hancock, and W. A. P. Smith, Eds. BMVA Press, 2016. [Online]. Available: http://www.bmva.org/bmvc/2016/papers/paper082/index.html

[119] Y. Feng, Y. Li, and J. Luo, "Learning effective Gait features using LSTM," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 325–330. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7899654

[120] F. Battistone and A. Petrosino, "TGLSTM: A time based graph deep learning approach to gait recognition," *Pattern Recognition Letters*, vol. 126, pp. 132–138, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0167865518301703

[121] T. Wolf, M. Babaee, and G. Rigoll, "Multi-view gait recognition using 3D convolutional neural networks," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 4165–4169. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7533144

[122] B. Lin, S. Zhang, and F. Bao, "Gait Recognition with Multiple-Temporal-Scale 3D Convolutional Neural Network," in *Proceedings of the 28th ACM International conference on Multimedia*, 2020, pp. 3054–3062. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3394171.3413861

[123] B. Lin, S. Zhang, and X. Yu, "Gait Recognition via Effective Global-Local Feature Representation and Local Temporal Aggregation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 14 648–14 656. [Online]. Available: https://ieeexplore.ieee.org/document/9710710

[124] B. Lin, S. Zhang, Y. Liu, and S. Qin, "Multi-Scale Temporal Information Extractor For Gait Recognition," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 2998–3002. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9506488

[125] S. Gao, J. Yun, Y. Zhao, and L. Liu, "Gait-D: Skeleton-based gait feature decomposition for gait recognition," *IET Computer Vision*, vol. 16, no. 2, pp. 111–125, 2022. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/cvi2.12070

[126] Z. Chen, S. Li, B. Yang, Q. Li, and H. Liu, "Multi-Scale Spatial Temporal Graph Convolutional Network for Skeleton-Based Action Recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 2, pp. 1113–1122, May 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/16197

[127] L. Sorber, M. Van Barel, and L. De Lathauwer, "Optimization-Based Algorithms for Tensor Decompositions: Canonical Polyadic Decomposition,

Decomposition in Rank-\$(L_r,L_r,1)\$ Terms, and a New Generalization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 695–720, 2013. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/120868323

[128] M. Shopon, A. Bari, and M. L. Gavrilova, "Residual connection-based graph convolutional neural networks for gait recognition," *The Visual Computer*, vol. 37, no. 9, pp. 2713–2724, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s00371-021-02245-9

[129] Y.-F. Song, Z. Zhang, C. Shan, and L. Wang, *Stronger, Faster and More Explainable: A Graph Convolutional Baseline for Skeleton-Based Action Recognition*. New York, NY, USA: Association for Computing Machinery, 2020, p. 1625–1633. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3394171.3413802

[130] T. Teepe, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, "Towards a Deeper Understanding of Skeleton-based Gait Recognition," in *17th IEEE Computer Society Workshop on Biometrics 2022*. IEEE/CVF, 2022. [Online]. Available: https://arxiv.org/abs/2204.07855

[131] S. Yu, H. Chen, Q. Wang, L. Shen, and Y. Huang, "Invariant feature extraction for gait recognition using only one uniform model," *Neurocomputing*, vol. 239, pp. 81–93, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S092523121730276X

[132] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, "Joint Intensity Transformer Network for Gait Recognition Robust Against Clothing and Carrying Status," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 12, pp. 3102–3115, 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8695052

[133] ——, "Gait Recognition via Semi-supervised Disentangled Representation Learning to Identity and Covariate Features," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 13 309–13 319. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9156701

[134] M. Benouis, M. Senouci, R. Tlemsani, and L. Mostefai, "Gait recognition based on model-based methods and deep belief networks," *International Journal of Biometrics*, vol. 8, no. 3-4, pp. 237–253, 2016. [Online]. Available: https://www.inderscienceonline.com/doi/abs/10.1504/IJBM.2016.082598

[135] B. M. Nair and K. D. Kendricks, "Deep network for analyzing gait patterns in low resolution video towards threat identification." *Electronic*

Imaging, vol. 2016, no. 11, pp. 1–8, 2016. [Online]. Available: https://library.imaging.org/ei/articles/28/11/art00015

[136] S. Hou, C. Cao, X. Liu, and Y. Huang, "Gait Lateral Network: Learning Discriminative and Compact Representations for Gait Recognition," in *European Conference on Computer Vision*. Springer, 2020, pp. 382–398. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-58545-7_22

[137] X. Li, Y. Makihara, C. Xu, Y. Yagi, S. Yu, and M. Ren, "End-to-End Model-Based Gait Recognition," in *Computer Vision – ACCV 2020*, H. Ishikawa, C.-L. Liu, T. Pajdla, and J. Shi, Eds. Cham: Springer International Publishing, 2021, pp. 3–20. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-69535-4_1

[138] S. Hou, X. Liu, C. Cao, and Y. Huang, "Set Residual Network for Silhouette-Based Gait Recognition," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 3, pp. 384–393, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9410587

[139] T. Chai, X. Mei, A. Li, and Y. Wang, "Silhouette-Based View-Embeddings for Gait Recognition Under Multiple Views," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 2319–2323. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9506238

[140] G. Batchuluun, H. S. Yoon, J. K. Kang, and K. R. Park, "Gait-Based Human Identification by Combining Shallow Convolutional Neural Network-Stacked Long Short-Term Memory and Deep Convolutional Neural Network," *IEEE Access*, vol. 6, pp. 63 164–63 186, 2018. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8502031

[141] W. An, R. Liao, S. Yu, Y. Huang, and P. C. Yuen, "Improving Gait Recognition with 3D Pose Estimation," in *Biometric Recognition*, J. Zhou, Y. Wang, Z. Sun, Z. Jia, J. Feng, S. Shan, K. Ubul, and Z. Guo, Eds. Cham: Springer International Publishing, 2018, pp. 137–147. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-97909-0_15

[142] Y. Zhang, Y. Huang, S. Yu, and L. Wang, "Cross-View Gait Recognition by Discriminative Feature Learning," *IEEE Transactions on Image Processing*, vol. 29, pp. 1001–1015, 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8759096

[143] S. Yu, H. Chen, E. B. Garcia Reyes, and N. Poh, "GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Networks,"

in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 30–37. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8014814

[144] S. Yu, R. Liao, W. An, H. Chen, E. B. García, Y. Huang, and N. Poh, "GaitGANv2: Invariant gait feature extraction using generative adversarial networks," *Pattern Recognition*, vol. 87, pp. 179–189, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320318303649

[145] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, "Gait recognition invariant to carried objects using alpha blending generative adversarial networks," *Pattern Recognition*, vol. 105, p. 107376, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320320301795

[146] S. Li, W. Liu, H. Ma, and S. Zhu, "Beyond View Transformation: Cycle-Consistent Global and Partial Perception Gan for View-Invariant Gait Recognition," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2018, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8486484

[147] Z. Zhang, L. Tran, X. Yin, Y. Atoum, X. Liu, J. Wan, and N. Wang, "Gait Recognition via Disentangled Representation Learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4710–4719. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8953845/

[148] Z. Zhang, L. Tran, F. Liu, and X. Liu, "On Learning Disentangled Representations for Gait Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9154576

[149] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2019, pp. 5693–5703. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8953615

[150] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *European Conference on Computer Vision*. Springer, 2014, pp. 740–755. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-10602-1_48

[151] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis," in *Proceedings of the*

IEEE conference on computer vision and pattern recognition. IEEE, 2016, pp. 1010–1019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7780484

[152] J. Liu, A. Shahroudy, M. Perez, G. Wang, L.-Y. Duan, and A. C. Kot, "NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2684–2701, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8713892

[153] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7780459

[154] Z. Liu, H. Zhang, Z. Chen, Z. Wang, and W. Ouyang, "Disentangling and Unifying Graph Convolutions for Skeleton-Based Action Recognition," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 143–152. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9156556

[155] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, p. 1929–1958, jan 2014. [Online]. Available: https://dl.acm.org/doi/10.5555/2627435.2670313

[156] K. Cheng, Y. Zhang, C. Cao, L. Shi, J. Cheng, and H. Lu, "Decoupling GCN with DropGraph Module for Skeleton-Based Action Recognition," in *European Conference on Computer Vision (ECCV)*. Springer, 2020, pp. 536–553. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-58586-0_32

[157] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[158] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, Z. Zhang, H. Lin, Y. Sun, T. He, J. Mueller, R. Manmatha, M. Li, and A. J. Smola, "ResNeSt: Split-Attention Networks," *CoRR*, vol. abs/2004.08955, 2020. [Online]. Available: https://arxiv.org/abs/2004.08955

[159] S. Yu, D. Tan, and T. Tan, "A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 4. IEEE, 2006, pp. 441–444. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1699873

[160] S. Khirirat, H. R. Feyzmahdavian, and M. Johansson, "Mini-batch gradient descent: Faster convergence under data sparsity," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 2880–2887. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8264077

[161] I. J. Goodfellow, Y. Bengio, and A. C. Courville, *Deep Learning*, ser. Adaptive computation and machine learning. MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org/

[162] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations (ICLR)*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[163] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised Contrastive Learning," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 18 661–18 673. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/d89a66c7c80a29b1bdbab0f2a1a94af8-Abstract.html

List of Publications

Conference

1. **M. B. Hasan**, T. Ahmed, and M. H. Kabir, “HEATGait: Hop-Extracted Adjacency Technique in Graph Convolution based Gait Recognition”. To appear in *Proceedings of the 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC 2022)*. 2022. [Online]. Available: https://arxiv.org/abs/2204.10238