IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 9, 2022, accepted May 17, 2022, date of publication May 23, 2022, date of current version May 26, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3176873

Multiview Gait Recognition on Unconstrained Path Using Graph Convolutional Neural Network

MD. SHOPON[1], (Member, IEEE), GEE-SERN JISON HSU[2], (Senior Member, IEEE), AND MARINA L. GAVRILOVA[1], (Senior Member, IEEE)

[1]Department of Computer Science, University of Calgary, Calgary, AB T2N 1N4, Canada
[2]Department of Mechanical Engineering, National Taiwan University of Science and Technology, Taipei 10607, Taiwan

Corresponding author: Md. Shopon (md.shopon@ucalgary.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada by the Discovery Grant (DG) on Machine Intelligence for Biometric Security under Grant 10007544, in part by the Strategic Partnership Grant on Biometric-Enabled Identity Management and Risk Assessment for Smart Cities under Grant 10022972, and in part by the Innovation for Defence Excellence and Security (IDEaS) Collaborative Project under Grant 10027075.

ABSTRACT Human gait recognition is a valuable biometric trait with vast applications in security domain. In most situations, the gait data is collected while the subject walks straight. Thus, the performance of the gait recognition system degrades when the subject changes walking direction. Previous gait recognition research was predominantly conducted for constrained paths, which limited the system's robustness and applicability. This paper introduces a novel approach for gait recognition which aims to recognize subjects walking along an unconstrained path. A graph neural network-based method is proposed for gait recognition along unconstrained path. The input of the architecture is the body joint coordinates and adjacency matrix representing the skeleton joints. Furthermore, a residual connection is incorporated to produce a smoothened output of the input feature. This graph neural network model utilizes the kinematic relationships of the body joints as well as spatial and temporal features. The findings demonstrate that the proposed method outperformed other state-of-the-art gait recognition methods on unconstrained paths. Multi-view Gait AVA and CASIA-B dataset are used to evaluate the efficacy of the proposed method.

INDEX TERMS Gait recognition, graph neural networks, residual connection, unconstrained gait recognition, biometrics.

I. INTRODUCTION

Biometric authentication is the method of identifying individuals based on their behavioral and physiological characteristics. Several biometric traits can be used for person authentication. Among them face, iris, voice, fingerprint, and gait hold most discriminative features [1]. The gait of a person denotes the walking pattern [2], [3]. It consists of information about the psychological and physical states, which is suitable for performing person identification. It has numerous advantages such as being unobtrusive, non-invasive, ubiquitous and acceptable. Due to its advantages, gait has been broadly used in different domains, including health [4], affective computing [5], assisted living [6], emotion recognition [7], and user identification [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir.

Research on gait recognition has gained a significant interest due to its unique characteristics and applications in different domains. A change in appearance due to change in viewing angle has a negative impact on performance in gait recognition methods. When a person walks along an unconstrained path, the perception angle between the subject's walking direction, and the camera optical axis varies in a single gait cycle.

Majority of security cameras in public spaces are used for remote observation or monitoring of human activities. The collected data would contain observations of persons walking at varied angles to the camera, wearing bulky clothing, or changing directions. However, these types of walking conditions are very rarely studied in literature. The two most common gait recognition approaches developed when the subject walks in a straight line are appearance-based methods [9]–[12] and the model-based methods [13]–[15].

The majority of appearance-based methods are not suitable for varied walking conditions. Model-based methods are slightly less sensitive to rotational effects and minor changes in viewpoint. Those are, nevertheless, distinguished by complicated mapping and searching operations, which increase the computation cost [16].

For gait recognition, there has been a noticeable trend toward deep learning-based systems due to the limitation of traditional machine learning algorithms [17]. Deep learning methods, such as Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and autoencoders, are distinguished by many layers of neurons that learn independent data representations [18]. Deep learning has the advantage of eliminating the need for independent feature development because the training method extracts the discriminative features from data. However, the disadvantage of using deep learning architectures is that they require high computation power. Moreover, the previous deep learning techniques did not fully utilize relative directions and motions of body joints in their architecture. Therefore, it is required to develop a more robust and lightweight method that utilizes the relative directions and motions of the skeleton body joints. The proposed GCNN-based model is perfectly suitable to remedy this situation. The feature propagation allows the GCNN to transmit the relative intra-join motions and the body joint motions to connected nodes, thus taking advantage of graph-based approach to extract highly discriminating features.

This study describes a novel method for classifying people who walk along unconstrained paths. This works primary research question is whether a residual connection-based graph convolutional neural network can extract more distinguishable and meaningful features for unconstrained path gait recognition. It is, to the best of our knowledge, a first deep-learning architecture proposed to correctly identify individuals from their walking patterns in the presence of unconstrained paths. Thus, the main contributions of this study are:

- A highly efficient deep-learning based architecture has been proposed for unconstrained gait recognition problem. The proposed method utilizes the kinematic dependency of the body joints, exploring both spatial and temporal features, which increases the accuracy of gait recognition.
- An adjacency matrix has been introduced to represent human joints as an input for the proposed graph convolution network.
- A Residual Connection has been introduced in the Graph Convolutional Neural Network allows to exploit the dynamic body joints relationship. This has resulted in very high recognition accuracy in the presence of varied walking directions.
- Global attention sum pooling layer is utilized to make the proposed model lightweight and reduce the training time.
- The proposed system was tested on two datasets, demonstrating its validity and outperforming all the previous works on unconstrained path gait recognition.

The proposed method is validated on the AVA Multi-View Gait Dataset [19] and CASIA-B dataset [20]. A comparison with recent state-of-the-art methods was also conducted, revealing that the proposed method outperformed all comparators. One potential application of this work is biometric authentication in special or restricted areas, including governmental facilities, military bases, and refugee camps [21]. This research can also be used for medical anomaly detection [22], [23], in smart homes for monitoring residents well-being [24], in robot navigation [25], and in virtual reality applications [26].

The rest of the paper is organized as follows: Section II will illustrate the previous research on gait recognition. Section III will discuss the data preparation and the proposed method. Experimental results of the proposed method will be presented in section IV. Finally, the findings of the work and future research direction will be discussed in section V.

II. RELATED WORK

Over the past decade, a growing body of literature has explored the gait recognition problem from various perspectives.

One of the first works on the appearance-based method was conducted in [9], where authors proposed a baseline algorithm that performs gait recognition by obtaining silhouettes and calculating the temporal correlation. Han *et al.* [10] proposed Gait Energy Image (GEI), which is another commonly used feature for gait recognition. GEI is obtained by averaging the silhouettes using statistical analysis. Hoffman and Rigoll [27] further proposed an improvement of gait energy images by integrating gradient histograms. GEI is a compact representation of all the silhouettes; it is computationally inexpensive and widely used. However, GEI is highly sensitive to different factors; for instance, if the subject carries accessories, the silhouette representation might not correctly represent the subject. Another drawback of GEI is that the system's performance reduces when the viewing angle changes. By stating this problem, the authors of [11] proposed a novel view transformation method that reconstructs one silhouette to another silhouette view, and all the silhouettes are transformed into one single view. However, when the viewing angle difference between the silhouettes is large, it causes degradation in the performance. Shakhnarovich *et al.* [12] proposed a view-normalization technique for multi-view gait recognition. They used a visual hull to synthesize images to obtain a 3D gait volume. Although this problem resolves the multi-view gait recognition problem, it requires placing multiple cameras at different viewing angles, which is inconvenient for frame synchronization. Various GEI modifications were proposed in [28]; however, these, like all other appearance-based approaches, suffered from view dependency.

Alternatively, model-based methods retrieved various features like motion and shape by fitting the videos to the model. These methods did not have a dependency from the viewing angle or scale. However, the quality of the video or image played a crucial role in extracting the features properly. BenAbdelkader *et al.* [13] were the first to propose a method for model-based gait recognition. They used two parameters in their work, namely cadence and stride. Cunado *et al.* [14] considered how human legs moved and linked the movement with pendulum. They used Hough transformation to obtain the feature representing harmonic motion pattern. Johnson and Bobick [15] first used different distance-based parameters for gait recognition. They calculated some static parameters of the subject, for instance, the height of the distance of head and feet from the pelvis and the maximum distance between pelvis and feet, and later used these features for performing gait recognition. Yoo *et al.* [16] transformed silhouettes obtained from the video frames into a two-dimensional stick-like figure. They obtained stick-like figures by using motion information from anatomical knowledge. Using 3D volumetric data, Seely *et al.* [29] created a proprietary dataset designed by mimicking the setting in airports and other high throughput environments. They produced silhouettes from a specific viewpoint and then silhouettes were then sent into a common 2D gait analysis method. The sequences were obtained from a multi-biometric tunnel in which subjects walked straight. To extract gait features, Ariyanto and Nixon [30] employed a model-fitting technique that utilized correlation filters and dynamic programming. To simulate the human lower legs, they utilized a structural model that included articulated cylinders with 3D degrees of freedom (DoF) fitted to a visible hull shape. Tian *et al.* [31] proposed a view adapting method for free view gait recognition. Their proposed method utilized a walking trajectory fitting method to compute viewing angles of a gait sequence, and a joint gait manifold was used to find the optimal manifold in the gait sequence.

Although the above-mentioned methods work well in cross-view settings, they are vulnerable to variations due to their reliance on human appearance and shape. Furthermore, it is challenging to acquire input silhouettes when the camera changes. Therefore, effective feature extraction and modeling strategies to address the highly nonlinear association between gait features in complex cross-view have been lacking. Moreover, these methods cannot handle the curved trajectory path of subjects.

In addition to the above methods, kinect based gait recognition gained attention over past couple of years. Ahmed *et al.* [32] used a Microsoft Kinect device to extract skeleton joints of the body. Afterward, distance feature vectors were obtained for each of the joints in relation to the other joints in the gait cycle. The computed distance-based features were later passed into the k-Nearest Neighbors (KNN) method for classification. Bari *et al.* [33] used a similar method to extract skeleton joint features; however, in addition to the distance-based features, the authors proposed two geometric features, which proved to be effective for classifying a person. For classification, the authors employed an artificial neural network. Although these methods make use of the human body's joint relationships, the architectures have a significant number of parameters, which makes the system very difficult to train.

Deep learning-based approaches have proven to be beneficial in gait recognition research in the recent years. Siamese Neural Networks were utilized by Zhang *et al.* [34] to transform a sequence of frames into gait energy images. The authors used a contrastive loss function for the provided inputs, which allows the system to minimize loss for similar-looking inputs while increasing it for distinct ones. He *et al.* [35] utilized Multi-task Generative Adversarial Networks (MGAN) for generating view-specific feature encoding. Later these encodings were used for further classification. In [36], the authors employed two different methods for extracting different features. Long Short Term Memory (LSTM) based architecture was used to extract spatio-temporal features, and autoencoder was used to model discriminative features. Chao *et al.* [37] considered the gait recognition problem from a different point of view. Instead of considering gait sequence as a continuous feature, they considered set of independent silhouettes. Their approach could extract invariant properties from the set, such as speed and step distance. Batistone *et al.* [38] proposed an LSTM network based on time-graph. This method extracted key points from a person's skeleton representation and then trained joint characteristics using a fully connected neural network (FCNN) and LSTM. Recently, Lin *et al.* [39] applied a GCNN based method for reconstructing 3D human pose and its mesh from a single image.

Because of the specific structure of the network, deep learning-based methods have the benefit of being able to capture spatial and temporal information. However, these deep learning-based approaches typically have high computational cost. Furthermore, these approaches ignore the kinematic link between joints, an essential feature in gait recognition.

Some recent research addressed changing of view point during walking. Thus, authors of [31] proposed a walking trajectory fitting method to compute viewing angles of a gait sequence. In addition, authors of [40] extracted motion descriptors from densely sampled short-term trajectories and used them for gait recognition. The Fisher Vector encoding method was used to encode the feature descriptors for further classification. Their method was tested on CASIA [20] and TUM GAID [41] datasets. However, very few works touched on unconstrained gait recognition. D. López-Fernández *et al.* [42] proposed a gait descriptor method for unconstrained gait recognition named gait entropy volume. This method focuses on identifying 3D dynamical information of a subject using the concept of entropy. The authors validated their method on AVA dataset, described in [19]. Later, D. López-Fernández *et al.* [43] proposed another method for unconstrained gait recognition. This work used volumetric gait recognition to capture

3D morphological information. Three gait morphological descriptors were developed to extract information from 3D volumes and validated it on the same AVA dataset. Arshad *et al.* [44] developed a framework for unconstrained gait recognition using deep neural network and Fuzzy Entropy Controlled Skewness (FECS). They use a pretrained CNN architecture to extract features in the first step. The second step computed the entropy and skewness vectors from the extracted features. Later, the best subset of features was used for classification. They performed experiments on both constrained and unconstrained paths on CASIA [20] and AVA [19] datasets. Zhu *et al.* [45] recently developed a gait database that consists of subjects walking in an unconstrained environment. In their work, the authors used GaitSet method introduced by Chao *et al.* [37].

A recent paper by Lui *et al.* [46] presents an interesting idea on modelling of destinations for data-driven pedestrian trajectory in public buildings. The proposed method first identifies most probable destination of the pedestrian's by employing Destination Classifier (DC) and later predicts the future trajectories by utilizing destination-specific trajectory model (DTM). Another paper tackles robot navigation system based on human trajectory prediction in unconstrained environments [25]. The proposed architecture first predicts the trajectory of the human movement and later prepares a path for the robot. Li *et al.* [47] tackled an interesting problem of finding the effect of environment during the trajectory estimation. Authors proposed a counterfactual analysis for human trajectory estimation and investigation of the effect of environmental bias. A casual graph model was constructed for forecasting the current, past and future trajectories.

The above methods achieved good recognition accuracy on unconstrained paths. However, they did not utilize the kinematic relationships of the body joints, which can lead to an increased recognition performance. Moreover, they did not utilize graph neural network architecture to capture an intrinsic relationship among body joints.

The advantages that the proposed architecture has over the above-mentioned methods are:

- The proposed method has a highly resistant view and appearance variation.
- The proposed method achieved significant accuracy on unconstrained trajectories.
- The proposed Residual Connection-based Graph Convolutional Neural Network (RGCNN) architecture is lightweight, making the model deployable in practice.
- The proposed method utilizes the kinematic dependency of the body joints. Both spatial and temporal features are exploited during the training phase, which increases the accuracy of gait recognition.

All of the recent methods proposed for unconstrained gait recognition and discussed in this section were implemented in this paper. Their performance was compared to the proposed method on the benchmark AVA dataset. In addition, multi-view methods by Seely *et al.* [29] and Ariyanto and Nixon [30] were implemented for comparison.

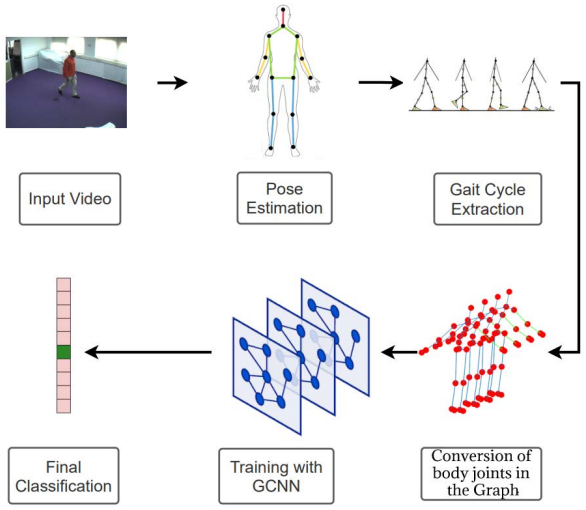


FIGURE 1. Overall architecture of the proposed method.

FIGURE 2. Pose estimation result of an input frame.

III. PROPOSED METHOD

A. OVERALL METHODOLOGY

This work proposes a residual connection-based graph convolutional neural network architecture, which utilizes the kinematic dependency of body joints to perform multi-view unconstrained path gait recognition. There are six steps in our proposed framework. Initially, we transform the image frames into a video. Later, the video is passed into a pose estimation architecture where we obtain the keypoints of body joints. In order to make the data sample uniform, we extracted gait cycles from each video. The gait cycles are later converted into the graph-like form to apply a GCNN. Our residual connection-based GCNN trained the transformed body joints. The proposed residual connection architecture is formed of residual blocks, batch normalization layer, global attention sum pooling layer, and dense layers. The residual blocks propagate spatio-temporal features to a deeper layer for amplifying the feature set. Furthermore, the global attention sum pooling layer reduces the number of trainable parameters by removing the unnecessary and noisy features. In addition, it reduces the model's total parameters count, making it a lightweight model. Finally, we apply the learned

model to the final classification. The overall methodology of this work is depicted in Fig. 1. Each of the steps will be briefly explained in the following subsections.

B. POSE ESTIMATION

Pose estimation is a method for tracking a person's or object's movements. This method is widely utilized in augmented reality, animation, gaming, and robotics [48]. OpenPose [49] is the first deep learning based pose estimation architecture that recognizes the key-points of different body joints on single pictures. To extract the feature maps from the input, the image is first processed through a baseline CNN network. The first ten layers of the VGG-19 network are employed. Following that, the feature map is processed in a multi-stage CNN pipeline to yield the Part Confidence Maps (PCM) and Part Affinity Field (PAF). At the last stage, the PCM and PAF's are computed using a bipartite matching algorithm to acquire the key points for each individual in the image. We have used the OpenPose network to obtain 25 body joints from the video. Fig. 2 shows the result of OpenPose estimating the keypoints from an input frame. To achieve a higher resistance to poor-quality gait sequences, the proposed approach utilizes OpenPose in conjunction with the dynamic modality of the skeletal sequences.

C. GAIT CYCLE EXTRACTION

The gait cycle is a walking pattern that begins with a single heel strike and progresses to the same heel strike. There are typically two approaches for estimating the gait cycle [50]. The first approach is the mid-stance or local minima, which involves keeping both feet close together. The second approach is the double support phase, in which the space between the two feet is at its greatest. The earlier research proved that the double support phase produced smoother gait cycles [51]. Therefore, we employed the double support phase method for better consistency and accuracy in this work. Fig. 3 exhibits a visual representation of one gait cycle computed from a walking sequence. A gait cycle is determined by calculating the Euclidean norm between the left and right ankles. As previously stated, a gait cycle consists of three maximum lengths between the feet. In Fig. 3, there is one gait cycle that is characterized by the red marked portion. The gait cycles of a subject always follow a specific pattern. To show that, we visualized the Euclidian distance of the left and right ankles joints for every frame in three video sequences of a subject. This can be observed in Fig. 4. We extracted multiple gait cycles from a single video. Several gait cycles are included as data augmentation for model training.

D. DATA REPRESENTATION

After extracting the gait cycles from videos, they are processed for Graph Neural Networks (GNN). Gait data is represented as a vector sequence in traditional methods. Therefore, such models ignore the kinematic relationship between joints

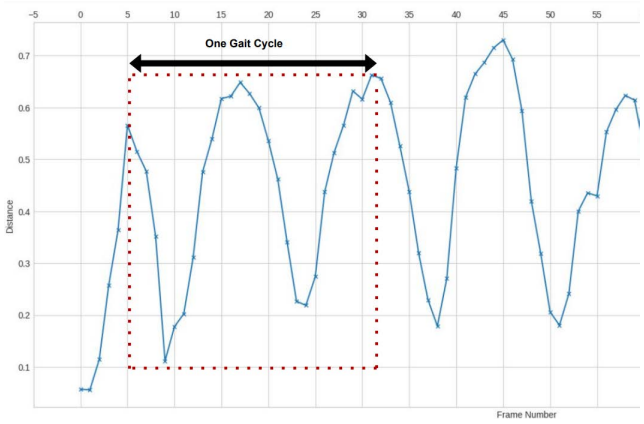


FIGURE 3. Euclidean distance between left and right ankles joint of a subject.

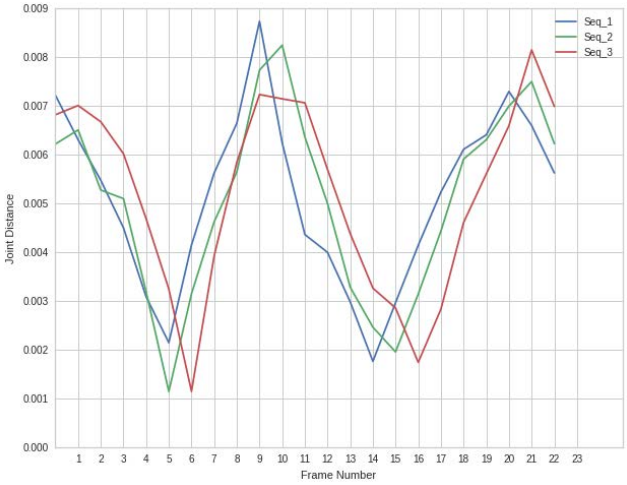


FIGURE 4. Smoothed Euclidean distance between the left and right ankles joint distance of subjects for three different video sequences.

that we employed in this study. The gait cycles must be transformed into a graph-like structure for the data to fit into Graph Convolutional Neural Network (GCNN). The gait cycles are represented as an undirected graph with vertices representing body joints and edges representing bones. There are two distinct subsets of the edges. The first subset denotes the relationship between intra joints in every frame, expressed as $E_S = \{(VT_{ti}, vT_{tj})|(i,j)\epsilon H\}$. Here H represents the body joints, and VT_{ti} and VT_{tj} are the vertices of the current frame and the following frame, respectively. The skeleton sequence's spatial information is stored in the first subset of edges. The second subset contains the intra-frame edges, denoted as $E_F = \{(VT_{ti}, VT_{(t+1)i})|i\epsilon H\}$. Here, all the edges in E_F denote its trajectory over the video frame sequence. The skeleton sequence's temporal information is stored in the second subset of edges. The temporal and spatial relationship in the body joints is depicted in Fig. 5. The proposed graph generation method has the advantage of preserving

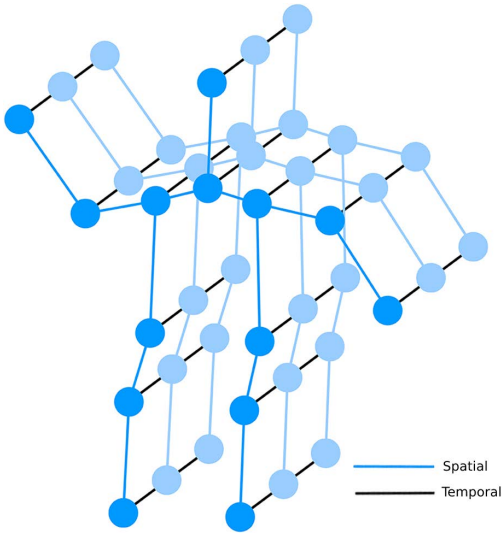


FIGURE 5. Spatial and temporal connection in the skeleton body joints.

the hierarchical structure of the skeleton sequences. Natural connectivity in human body structures and movements are represented as graph edges, and body joints will be represented as graph nodes. As a result, manual assignment of connections is not required.

E. RESIDUAL CONNECTION-BASED GRAPH CONVOLUTIONAL NEURAL NETWORK

1) GRAPH CONVOLUTIONAL NEURAL NETWORK

Graphs are some of the most versatile data structures due to their expressive power. A GCNN is a form of CNN that can exploit graph structure. GCNN can make more informed predictions about entities than traditional models that consider distinct entities in isolation. This is achieved by utilizing and exploiting underlying features in a graph.

To understand the mathematical foundation behind GCNN, let's consider the example of an undirected graph $G = \{E, V, A\}$, where E represents the set of edges, V represents the set of vertices, and A denotes the adjacency matrix, that establishes the connection between the edges and vertices. Later, the Fourier transform is applied to the graph to perform basic operations such as convolution.

The GCNN is a highly versatile architecture due to its feature aggregation property. GCNN accumulates information from previous layers, and generates effective feature representations of graph nodes. At the beginning, the feature vector x_i of node n_i is averaged with its neighborhood nodes feature vector. This process can be expressed as:

$$h^{(p+1)} = \sigma\left(h^{(p)} + \Theta^{(p)}\right) \tag{1}$$

Here, σ denote the sigmoid activation function, $h^{(p)}$ and $\Theta^{(p)}$ denote the activation matrix and the feature matrix of the p^{th} layer respectively. After this operation, node n_i will have all the features from its neighboring nodes. However, a node's aggregated representation does not include its own features. Therefore, a self-loop is added into the adjacency matrix to normalize the aggregated feature representation and include node n_i's feature into the feature aggregation process. This method can be denoted as:

$$A = \widetilde{d}^{-1/2}\widetilde{A}\widetilde{d}^{-1/2} \tag{2}$$

Here, $\widetilde{d}$ is the degree matrix of $\widetilde{A}$ and $\widetilde{A}$ is calculated as:

$$\widetilde{A} = A + I \tag{3}$$

where I is the identity matrix and A is the adjacency matrix without self loops.

2) GRAPH CONVOLUTIONAL NEURAL NETWORK WITH RESIDUAL CONNECTION

One of the difficulties in training neural networks is that deeper neural networks achieve higher accuracy and performance. However, the deeper the network, the harder it is for the training to converge. The idea of Residual connection (ResNet) was first designed by He *et al.* [52]. Earlier research demonstrated that residual connections are guaranteed to produce better results on an image or vision-based tasks [53]. In conventional feedforward neural networks, data flows sequentially through each layer. The output of one layer is the input for the following layer. Information is propagated in targeted levels via residual connections. As a result, the residual connection amplifies the discriminative information. Our architecture uses residual learning to extract spatio-temporal information from joints and propagate spatio-temporal features to a deeper layer. Identity mapping is the key feature for residual connection. The identity map employed in the GCNN layer differs from the originally proposed ResNet [52]. The Hadamard product was employed in the original study to concatenate the output of the activation layer with the output of a subsequent layer. However, in the proposed method, the dot product combines the feature maps. The model's convergence is amplified as a result of this additive property.

F. PROPOSED ARCHITECTURE

The proposed model uses two distinct inputs. The first one is the feature vector, which contains the location of each body joint, and the second is the adjacency matrix. The input is passed through an RGCNN layer, and the output is passed onto a batch normalization layer. The batch normalization layer normalizes each layer's input values so that the mean output and the standard deviation become zero and one, respectively. The momentum rate of the batch normalization layer is 0.99. Rectified Linear Unit (ReLU) is used as the activation function of the RGCNN layers. ReLU has the advantage of making the gradient less likely to vanish. ReLU's continuous gradient results in rapid learning. These RGCNN layers learn a hierarchical approximation of the gait pattern's spatio-temporal dynamics. The normalized features

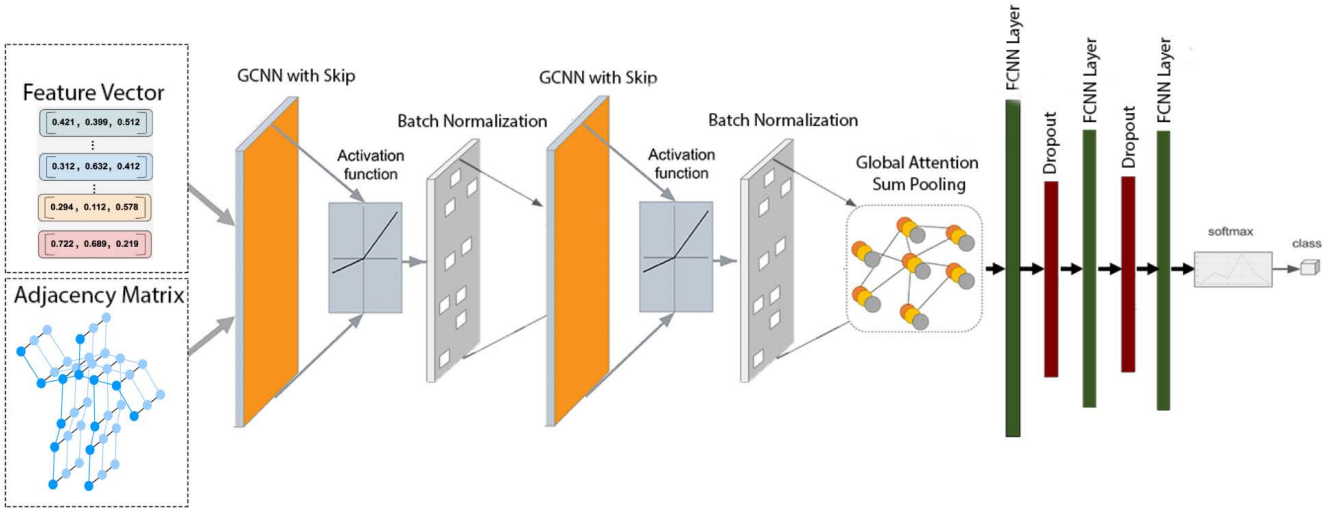


FIGURE 6. Proposed architecture of the RGCNN.

and adjacency matrix are passed to the next RGCNN layer followed by a batch normalization regularizer. This feature map becomes the pooling layer's input. The Global Attention Sum Pooling (GASP) is utilized to eradicate the impact of unnecessary nodes as well as minimize computation costs by pruning the graph. The GASP layer can be represented as follows:

$$x' = \sum_{i=1}^{N} \Big(\sigma\left(xw_1 + b_1\right) \bigodot \left(xw_2 + b_2\right) \Big)_i \tag{4}$$

The sigmoid activation function is used in the GASP layer, which is denoted as σ in Equation 4. N represents the data samples, x is the feature map, and w and b represent the weights and biases. The global attention sum pooling layer transforms the data into one dimension, which is later passed into the Multi-Layer Perceptron (MLP) network. The MLP sub-network is structured with three stacked MLP layers, and the number of fully connected nodes in there are 512, 256, and 128, respectively. In order to reduce the overfitting, a dropout regularizer is used between the MLP layers. In traditional methods, the output of a convolutional layer is converted into one large one-dimensional feature vector. Using this traditional method for converting the output of a CNN layer to a one-dimensional vector produces a large number of trainable parameters, which increases the training and inference time. Using the global attention sum pooling layer reduces the number of trainable parameters by removing the unnecessary and noisy features. The trainable and non-trainable parameters are presented in Table 1. The proposed architecture prevents overfitting by extracting generalized walking pattern characteristics for person identification because it contains a few model parameters. The overall architecture is depicted in Fig. 6.

To compute the loss between the original label and the predicted label, categorical crossentropy loss function is used.

TABLE 1. Number of parameters in the proposed RGCNN.

Parameter Type	Number of Parameters
Trainable	202,955
Non-trainable	1,536
Total	**204,491**

The categorical cross-entropy loss function is described in Equation 5.

$$CCELoss = \sum_{i=1}^{N} \sum_{c=1}^{C} \tau\left(y_i, c\right) \times logP(c|S_{i1}, \ldots, S_{iT}) \tag{5}$$

$$\tau(y_i, C) = \begin{cases} 1, & \text{if } y_i = c; \\ 0, & \text{if } otherwise \end{cases} \tag{6}$$

Here, N denotes the number of training samples and S_{i1}, S_{i2}, S_{i3},…,S_{iT} represent the factorized input graphs of i^{th} training sample and their respective label y_i, C denotes the number class and c denotes the original label of the samples. $logp(x)$ is a probability distribution of the factorized input graphs.

IV. EXPERIMENTAL RESULTS

A. DATASETS

The proposed method is validated on AVA Multi-View Gait Dataset [19] and CASIA-B dataset [20]. The AVA dataset comprises 20 subjects performing nine walking sequences in an indoor setting. Six color cameras (c_1, c_2, c_3, c_4, c_5, c_6) recorded each sequence inside the room at different viewing angles. Since the subjects enter the room from various points, the dataset is best adapted for testing view-independent gait recognition. The first three walking sequence (T_1, T_2, T_3) are straight and the rest (T_4,T_5, T_6, T_7, T_8, T_9) are curved trajectories. Authors of [19] successfully established that

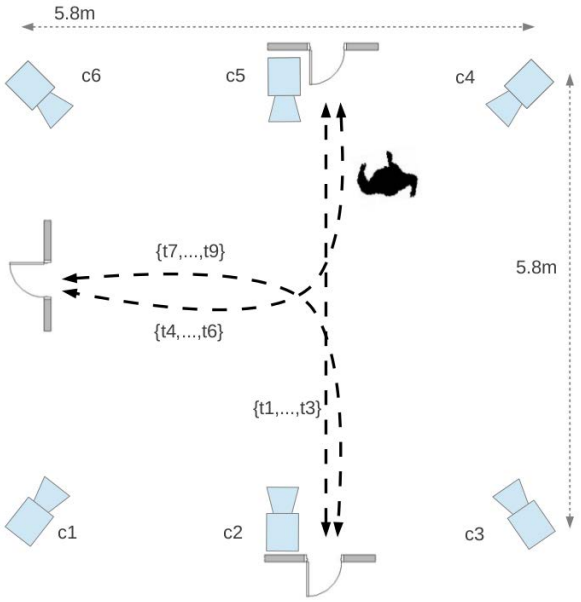


FIGURE 7. Samples of the walking trajectory, where $\{c1, \ldots, c6\}$ represents the set of cameras point of views and $\{t1, \ldots, t9\}$ represents the different trajectories followed by the subjects of the dataset [19].

curved trajectories closely approximate walking patterns along varied unconstrained paths. The map of the trajectories is shown in Fig. 7. The videos were sampled at a resolution of 640 × 480 pixels and the frame rate per second was 25. Some examples from the dataset are shown in Fig. 8. The CASIA-B gait dataset, a widely used standard dataset for gait recognition, is used for cross-validating the efficacy of the proposed method. This dataset comprises gait sequences from 124 participants under three distinct walking circumstances (normal, bag carrying, bulky cloth wearning), each captured from 11 different viewing angles (0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, 180°). Fig. 9 depicts some sample data from CASIA-B dataset.

B. EXPERIMENTS FOR SELECTING HYPERPARAMETERS

We performed several experiments to choose the key hyperparameters such as activation function, batch size, dropout rate, and optimizer. For selecting other hyperparameters, we adopted commonly used automated techniques. We employed an early stopping method [54] for the number of epochs, which resulted in the number of optimal iterations set to 150. For learning rate, we have used the adaptive learning rate method [55], For the network weight initialization, the Xavier initialization method was employed [56]. For choosing optimizer, batch size, activation function, and dropout rate we have performed several experiments. As there are nine trajectories in the dataset, we kept $T3$, $T6$, $T9$ in the testing set and the rest of the trajectories in the training set.

1) OPTIMIZER SELECTION EXPERIMENT

The node variables of neural networks are automatically adjusted throughout the training process to minimize the loss function. Depending on the optimizer, the direction and magnitude of the parameters are adjusted. Learning rate is the most crucial weight used to evaluate the optimizer's performance. A learning rate that is very high or very low results in non-convergence of the loss function or the range of the local minima, but not the absolute minima. As a result, the classifier's generalization capability improves when exposed to new data [57]. We experimented with four different optimization techniques to determine the best-performing optimizer in this work. The optimizers we experimented with are 1) Adam Optimizer, 2) Adadelta Optimizer, 3) Stochastic Gradient Descent (SGD) Optimizer, 4) Root Mean Squared Propagation (RMSProp) Optimizer. Table 2 demonstrates the performance of the optimizers. Furthermore, Fig. 10 shows the accuracy and loss curves for Adam, RMSProp, SGD, and Adadelta optimizers. Adam optimizer outperformed others. The accuracy and loss curves depict how the learning curve moved smoothly. Although RMSProp optimizer resulted in a smooth learning curve, it failed to produce optimal results.

Learning rate is another vital factor for the optimizer. We have used an adaptive learning rate in this work. Adaptive learning rate adapts the best learning rate with respect to loss during the training process. The magnitudes of Adam optimizer parameter updates are invariant to gradient rescaling, and their stepsizes are constrained by the stepsize hyperparameter, which explains why Adam optimizer performed well.

TABLE 2. Performance of the proposed RGCNN model with different optimization methods.

Optimizer	Training Accuracy	Testing Accuracy
Adam	**99.59%**	**99.14%**
Adadelta	97.78%	95.88%
RMSProp	95.89%	95.34%
SGD	98.18%	96.89%

TABLE 3. Performance of the proposed RGCNN model with different mini batch sizes.

Batch Size	Training Accuracy	Testing Accuracy
8	97.35%	96.53%
16	98.69%	98.04%
32	**99.48%**	**99.16%**
64	99.10%	98.89%

TABLE 4. Performance of the proposed RGCNN model with different dropout rates.

Dropout Rate	Training Accuracy	Testing Accuracy
No Dropout	98.46%	95.24%
0.2	98.21%	97.98%
0.3	**99.38%**	**99.12%**
0.4	98.07%	95.79%

2) BATCH SIZE SELECTION EXPERIMENT

One essential hyperparameter to tune in modern deep learning systems is batch size. Practitioners frequently prefer to

FIGURE 8. Samples of AVA Multi-view gait dataset [19].

FIGURE 9. Samples of Casia-B gait dataset [20].

TABLE 5. Performance of the proposed RGCNN model with different activation functions.

Activation Function	Training Accuracy	Testing Accuracy
Hyperbolic Tangent	98.25%	97.83%
Rectified Linear Unit	**99.48%**	**99.16%**

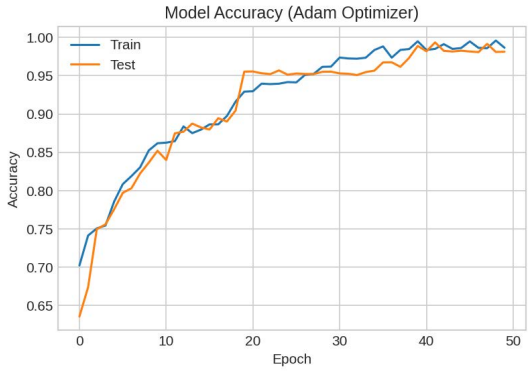
train their model with a larger batch size because it allows for computational speedups due to GPU parallelism [58]. However, it is well acknowledged that a large batch size leads to poor generalization [59]. Therefore, we experimented with different batch sizes to determine the optimal mini-batches. Table 3 shows that batch size 32 produced the highest accuracy among all batch sizes. Fig. 11 demonstrates the loss and accuracy graph for different batch size experiments.

3) ACTIVATION FUNCTION SELECTION EXPERIMENT

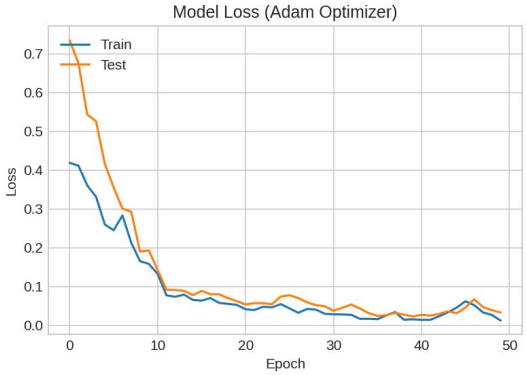
Activation functions are an essential component of neural network architecture [60]. The choice of activation function significantly influences the performance and effectiveness of neural networks. Different activation functions may be used in different parts of the model. Usually, either hyperbolic tangent or rectified linear unit activation function in the hidden layers provides better accuracy. In this work, we have experimented with two commonly used activation functions. The results are provided in Table 5. Rectified linear unit activation function attained 1.23% higher training and 2.31% higher testing accuracy over Hyperbolic Tangent activation function. The benefit of using ReLU is that the gradient is less likely to disappear. The constant gradient of ReLU leads to quick learning [61]. The accuracy and loss graph is depicted in Fig. 12.

4) Dropout Rate Selection Experiment

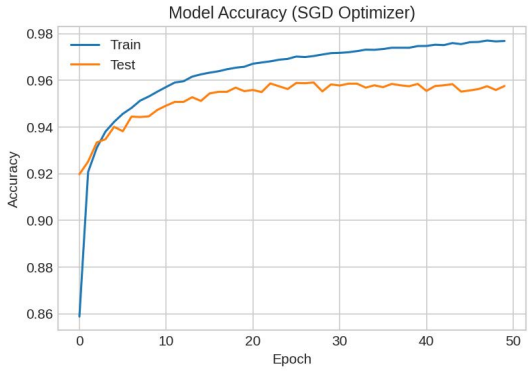
Regularization is a technique used in machine learning to prevent over-fitting [62]. The model is trained to avoid learning an interdependent set of feature weights by including this penalty. Dropout is one of the most effective regularization methods for deep learning. Some layer outputs are disregarded during training or "dropped out" at random. This causes the layer to appear as if it had a different



(a) Training vs validation accuracy graph for adam optimizer

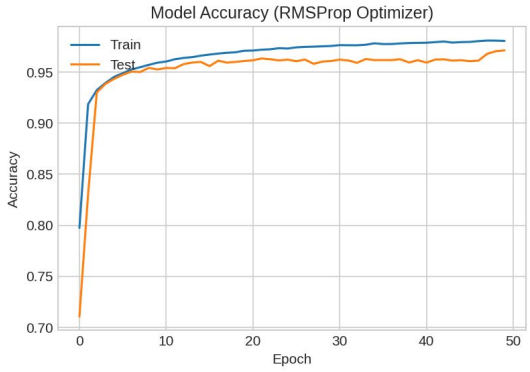


(b) Training vs validation loss graph for adam optimizer

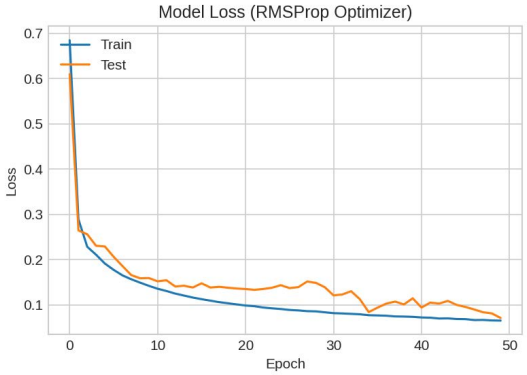


(c) Training vs validation accuracy graph for SGD optimizer

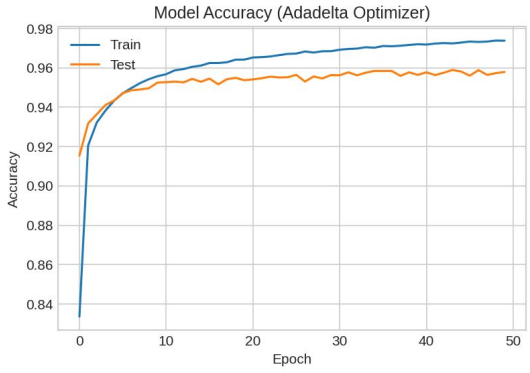
(d) Training vs validation loss graph for SGD optimizer



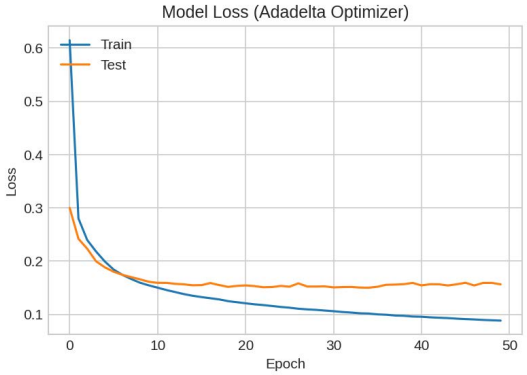
(e) Training vs validation accuracy graph for RMSProp optimizer



(f) Training vs validation loss graph for RMSProp optimizer

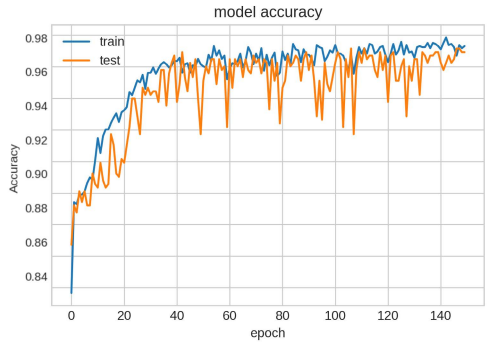


(g) Training vs validation accuracy graph for adadelta optimizer

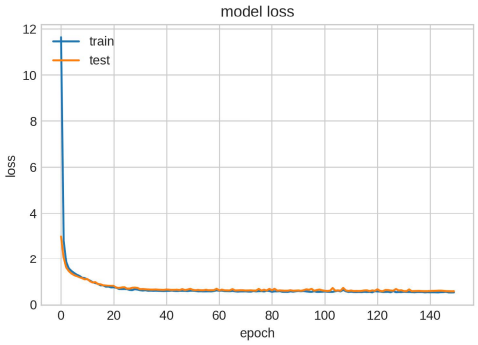


(h) Training vs validation loss graph for adadelta optimizer

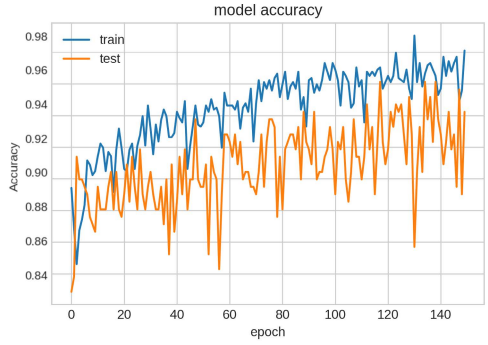
FIGURE 10. **Accuracy and loss curve for Adam, RMSProp, Adadelta, and SGD optimizer.**



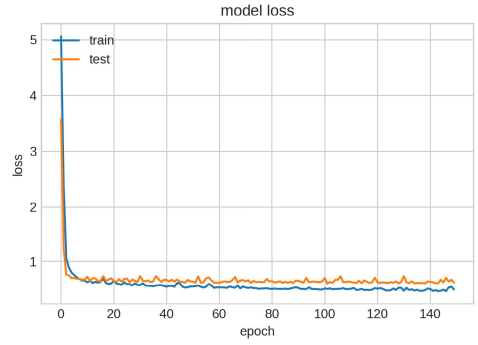
(a) Training vs validation accuracy graph for batch size 8.



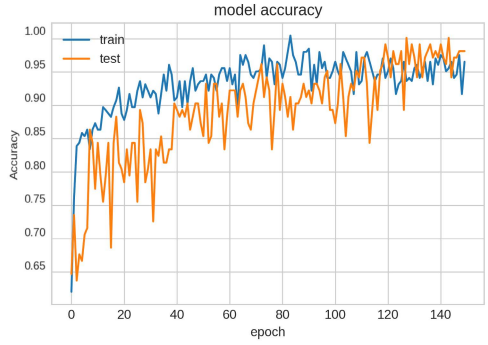
(b) Training vs validation loss graph for batch size 8.



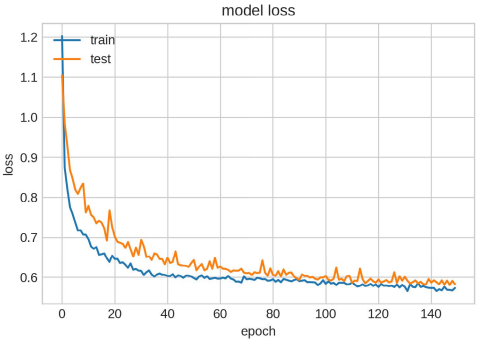
(c) Training vs validation accuracy graph for batch size 16.



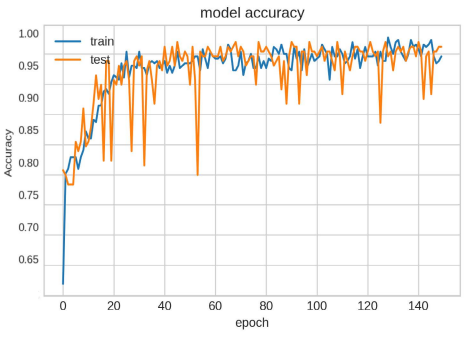
(d) Training vs validation loss graph for batch size 16.



(e) Training vs validation accuracy graph for batch size 32.



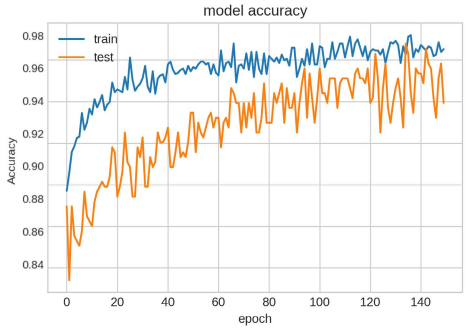
(f) Training vs validation loss graph for batch size 32.



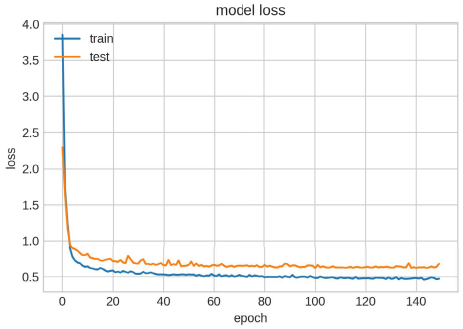
(g) Training vs validation accuracy graph for batch size 64.

(h) Training vs validation loss graph for batch size 64.

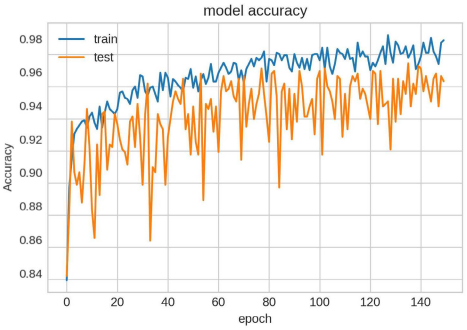
FIGURE 11. Accuracy and loss curve for batch size 8, 16, 32 and 64.



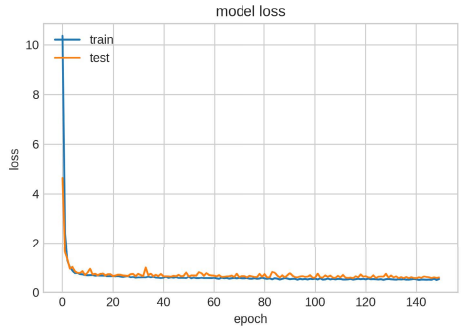
(a) Training vs validation accuracy graph for hyperbolic tangent activation function.



(b) Training vs validation loss graph for hyperbolic tangent activation function.



(c) Training vs validation accuracy graph for rectified linear unit activation function.



(d) Training vs validation loss graph for rectified linear unit activation function.

FIGURE 12. **Accuracy and loss curve for hyperbolic tangent and rectified linear unit activation function.**

number of nodes and connectedness to the preceding layer. We experimented with different dropout rates in this work. The experiments showed that the dropout rate of 0.3 produced the smooth loss and accuracy graph. The results attained from the experiment are demonstrated in Table 4. The accuracy and loss graph are shown in Fig. 13.

C. ABLATION STUDY

To verify that a residual connection in the GCNN and global attention sum pooling layer play a major role in the efficacy, we have performed an ablation study. AVA Multi-view Gait dataset was utilized to perform the ablation study.

1) EFFECTIVENESS OF THE RESIDUAL CONNECTION

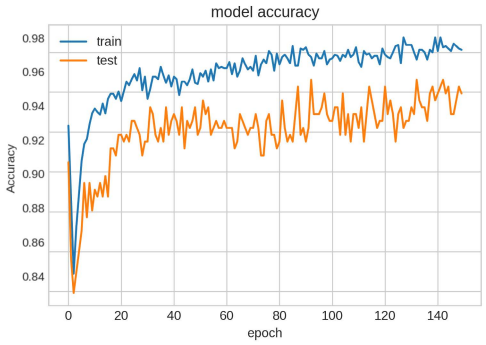
The first experiment we performed was removing the residual connection from the proposed architecture, while keeping the global attention sum pooling layer. Table 6, Row 2 demonstrates the result of the proposed model with this configuration. This ablation study showed that the performance of this architecture reduced drastically as a result of removing the residual connection. This architecture resulted in drop of 11.7% accuracy for trajectory 1, 8.24% accuracy for trajectory 2, 11.84% accuracy for trajectory 4, 10.09% accuracy for trajectory 5, 11.86% accuracy for trajectory 7, and in drop of 10.46% accuracy for trajectory 8.

2) EFFECTIVENESS OF THE GLOBAL ATTENTION SUM POOLING LAYER

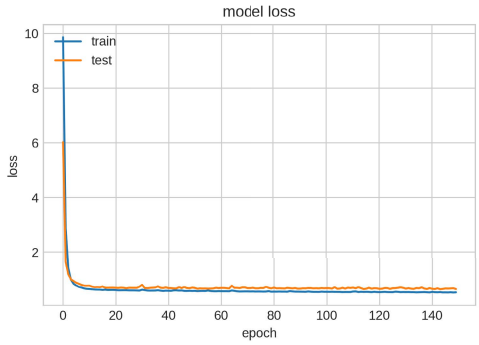
The second experiment we performed was removing the global attention sum pooling layer from the proposed architecture, with residual connection being retained. This ablation study showed that the performance of the revised architecture was reduced. As seen in Table 6, Row 3,this architecture attained 2.32% less accuracy for trajectory 1, 0.97% less accuracy for trajectory 2, 1.42% less accuracy for trajectory 4, 1.91% lesser accuracy for trajectory 5, 2.76% less accuracy for trajectory 7, and 2.01% less accuracy for trajectory 8.

3) EFFECTIVENESS OF THE RESIDUAL CONNECTION AND GLOBAL ATTENTION SUM POOLING LAYER

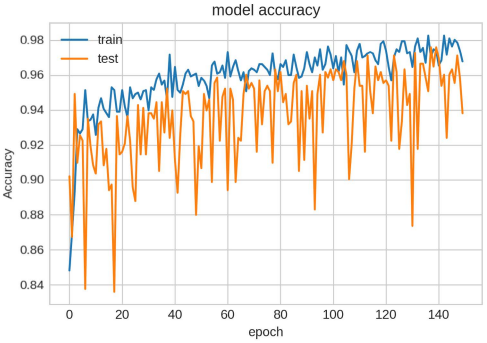
The third experiment we conducted for the ablation study was removing both the residual connection and global attention sum pooling layer. The results demonstrated that the performance of the revised architecture reduced drastically. Table 6, Row 1 demonstrates the results. The architecture resulted in loss of accuracy of 12.53% trajectory 1, 13.24% for trajectory 2, 12.71% for trajectory 4, 13.77% for trajectory 5, 15.45% for trajectory 7, and 14.82% for trajectory 8.



(a) Training vs validation accuracy graph for no dropout.

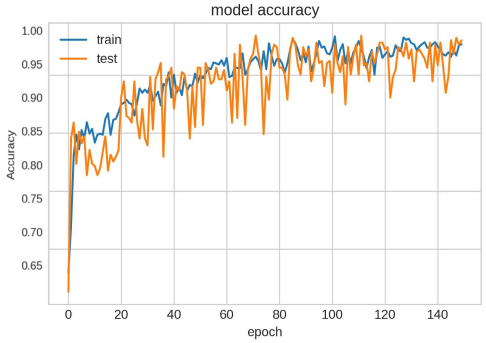


(b) Training vs validation loss graph for no dropout.

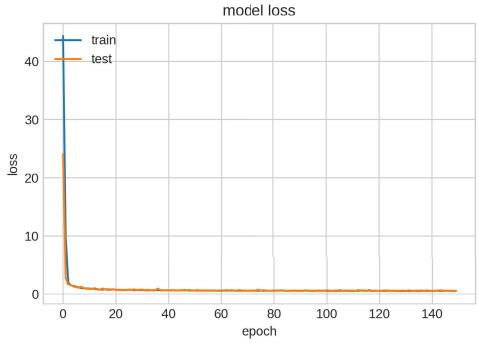


(c) Training vs validation accuracy graph for dropout rate 0.2

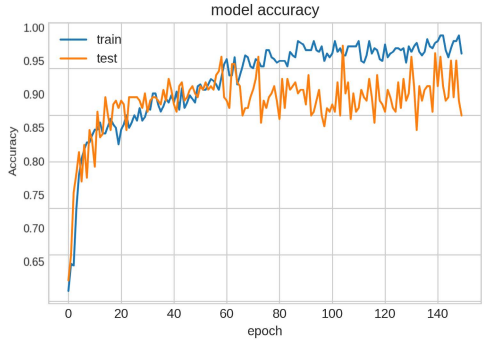
(d) Training vs validation loss graph for dropout rate 0.2.



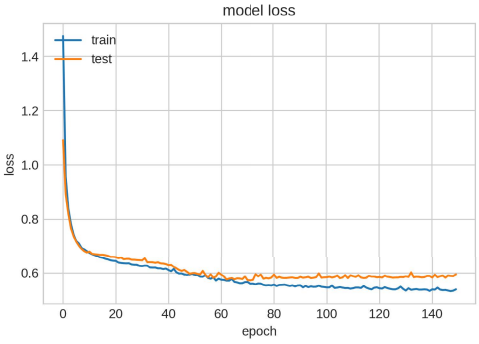
(e) Training vs validation accuracy graph for dropout rate 0.3.



(f) Training vs validation loss graph for dropout rate 0.3.



(g) Training vs validation accuracy graph for dropout rate 0.4.



(h) Training vs validation loss graph for dropout rate 0.4.

FIGURE 13. **Accuracy and loss curve for no dropout, 0.2, 0.3, and 0.4 dropout.**

TABLE 6. Ablation study for the different components of the proposed RGCNN architecture.

Experiment No.	Architecture		T-1	T-2	T-4	T-5	T-7	T-8
1	GCNN	✓	87.37%	82.76%	86.89%	86.23%	83.55%	81.98%
	Residual connection	×						
	Global attention sum pooling	×						
2	GCNN	✓	88.20%	89.02%	87.76%	89.91%	87.14%	86.34%
	Residual connection	×						
	Global attention sum pooling	✓						
3	GCNN	✓	97.58%	95.03%	98.18%	98.09%	96.24%	94.79%
	Residual connection	✓						
	Global attention sum pooling	×						
4	GCNN	✓	99.90%	96.00%	99.60%	100%	99.00%	96.8%
	Residual connection	✓						
	Global attention sum pooling	✓						

TABLE 7. Comparison of rank-1 accuracies with the proposed method on AVA Multi-view gait dataset for different unconstrained trajectories.

Method	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9	Mean
Ariyanto and Nixon [30]	55.00%	45.00%	52.60%	45.00%	26.30%	35.00%	35.00%	31.50%	40.00%	40.60%
Seely et al. [29]	90.00%	80.00%	94.70%	90.00%	60.00%	100%	80.00%	84.20%	90.00%	85.40 %
D.López-Fernández et al. [43]	100%	88.00%	100%	99.30%	99.20%	97.70%	96.20%	84.80%	100%	96.10%
Chao et al. [37]	100%	97%	97.26%	99.2%	98.5%	96.2%	98.0%	93.6%	99.4%	96.69%
Zhang et al. [36]	98.66%	91.30%	97.86%	100%	98.58%	98.20%	95.24%	94.2%	98.36%	96.93%
Arshad et al. [44]	98.36%	96.58%	98.12%	98.90%	97.80%	97.56%	98.16%	94.54%	98.88%	97.65%
Basic GCNN	94.30%	94.16%	97.54%	95.42%	97.28%	97.46%	93.70%	88.38%	96.40%	94.96%
RGCNN+ SGD + ReLU + 32 Batch	98.70%	95.15%	97.26%	97.22%	97.56%	98.46%	97.58%	95.23%	97.70%	97.42%
RGCNN+ Adam + ReLU + 64 Batch	99.16%	95.50%	98.80%	98.10%	98.85%	99.10%	98.16%	96.7%	97.95%	98.03%
RGCNN+ Adam + ReLU + 32 Batch	**99.9%**	**96%**	**99.9%**	**99.60%**	**100%**	**99.9%**	**99.0%**	**96.8%**	**98.86%**	**98.85%**

This ablation study clearly demonstrates that the residual connection plays an important role in the performance of the proposed method.

D. PERFORMANCE COMPARISON WITH STATE-OF-THE ART RESULTS

The majority of existing unconstrained path gait recognition methods use silhouettes gait energy images as features. Although these characteristics restrain necessary details for identifying individuals, they do not consider the kinematics dependency between various body parts. We employed a leave-one-out cross-validation to compare our method with existing state-of-the-art methods. In the experiments, each fold consists of a tuple formed from 20 sequences for testing and the eight sequences of each subject for training. The proposed method was compared with D.López-Fernández *et al.* [43], Seely *et al.* [29], Ariyanto and Nixon [30], and Arshad *et al.* [44]. In addition, two recent deep learning-based gait recognition methods, Chao *et al.* [37] and Zhang *et al.* [36] were chosen as comparators. Both of these methods are deep learning-based and showed excellent performance on gait recognition. Furthermore, to show the superiority of the proposed architecture, we trained a basic GCNN network without the residual connection. From the results reported in Table 7 it is observed that the proposed method outperformed all of the existing state-of-the-art methods. The proposed method outperformed all the previous works for all the trajectories except for trajectory 9. The proposed method outperformed Ariyanto and Nixon [30] by 58.25%, Seely *et al.* [29] by 13.45%, D. López-Fernández *et al.* [43] by 2.75%, Chao *et al.* [37] by 2.16%, and Zhang *et al.* [36] by 1.92% in terms of mean accuracy.

From the results presented in Table 7 it is evident that the proposed RGCNN architecture achieves higher accuracy than the existing methods. One of the reasons for such performance is the proposed temporal features. Integrating GCNN and residual learning results in a reduced number of model parameters and facilitates resolving the vanishing gradient problem. Overfitting is more likely to occur in networks with less number of parameters. It is avoided in this network using dropouts and Batch Norm Layers. The running time is decreased significantly as well. The most contributing feature map is passed through the successive layers using identity mapping during optimization. It enhances the efficacy on curved trajectories and in difficult walking conditions. Furthermore, the residual connection in the GCNN architecture helps in overcoming the vanishing gradient problem.

E. EXPERIMENTS ON CASIA-B GAIT DATASE

To show the robustness of the proposed method, we further validated the performance of CASIA-B multi-view gait dataset. This dataset comprises of video gait data collected from 124 participants under three distinct walking circumstances (normal, bag carrying, bulky cloth wearing), each captured from 11 different viewing angles (0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, 180°) [63], [64]. Training dataset was divided into three groups, Normal, Bag carrying and Bulky cloth wearing. The goal is to validate the hypothesis that training the developed architecture on normal walking conditions will yield a high accuracy of recognition even in the presence of bag carrying or bulky cloth wearing conditions. Moreover, the training set was formed from specific

TABLE 8. Different groups formed from CASIA-B Dataset.

Group name	Data
Group A	Normal: 0°, 18, 36°, 54°, 72°, 90°, 108°
Group B	Normal, Bag carrying, Bulky cloth wearing: 0°, 18°, 36°, 54°, 72°, 90°, 108°
Group X	Normal: 126°, 144°, 162°, 180°
Group Y	Bag carrying: 126°, 144°, 162°, 180°
Group Z	Bulky Cloth wearing: 126°, 144°, 162°, 180°

TABLE 9. Experimental results on CASIA-B gait dataset.

Ex. No.	Training Set	Testing Set	Accuracy
1	Group A	Group X	**98.86%**
2	Group A	Group Y	**92.13%**
3	Group A	Group Z	**90.79%**
4	Group B	Group X	**99.19%**
5	Group B	Group Y	**97.25%**
6	Group B	Group Z	**96.63%**

angles, whereas the test set comprised different from the training set viewing angles. The dataset was divided into five different groups for training and testing. The groups formed from CASIA-B dataset are shown in Table 8.

Table 9 shows the experimental results on CASIA-B gait dataset. First, the proposed RGCNN model was trained only on normal walking condition (Group A) and tested on bag carrying and bulky cloth wearing conditions. The proposed method attained 98.86% testing accuracy on the normal walking testing set (Group X), 92.13% on the bag carrying testing set (Group Y), 90.79% on bulky cloth wearing testing set (Group Z). These are very high results considering that the architecture has been trained both on different conditions than it was tested on and on different viewing angels. Notably, when the training set included challenging conditions, the accuracies increased to 99.19% for normal walking (Group X), 97.25% for bag carrying (Group Y), and to 96.63% for bulky cloth wearing (Group Z). This study convincingly established that the proposed method is not dataset biased and performs very well on a different dataset with challenging walking conditions.

V. CONCLUSION

This paper proposed a residual connection-based graph convolutional neural network for unconstrained path gait recognition. The proposed method successfully identifies subjects regardless of viewpoint or direction change. Unlike existing view-independent systems, which limit view change to a few degrees and cannot handle curved trajectories, the proposed method allows users to walk freely in the scene without compromising recognition. The proposed method outperformed the existing unconstrained path gait recognition works and was validated on the AVA Multi-View and CASIA-B gait recognition dataset. The number of model parameters is small, making it suitable for deploying in real-life scenarios. In the future, different architectures can be investigated to extract more discriminating feature maps. In addition, the integration of recurrent units into GCNN architecture can be evaluated. Combining video sequences along with skeletal sequences can be another interesting directions for future research.

REFERENCES

[1] Q. Xiao, "Technology review-biometrics-technology, application, challenge, and computational intelligence solutions," *IEEE Comput. Intell. Mag.*, vol. 2, no. 2, pp. 5–25, May 2007.
[2] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, "Gait recognition: A challenging signal processing technology for biometric identification," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 78–90, Nov. 2005.
[3] F. Ahmed, P. P. Paul, and M. L. Gavrilova, "DTW-based kernel and rank-level fusion for 3D gait recognition using Kinect," *Vis. Comput.*, vol. 31, nos. 6–8, pp. 915–924, Jun. 2015.
[4] J. Juen, Q. Cheng, V. Prieto-Centurion, J. A. Krishnan, and B. Schatz, "Health monitors for chronic disease by gait analysis with mobile phones," *Telemed. e-Health*, vol. 20, no. 11, pp. 1035–1041, Nov. 2014.
[5] T. A. L. Wren, G. E. Gorton, III, S. Õunpuu, and C. A. Tucker, "Efficacy of clinical gait analysis: A systematic review," *Gait Posture*, vol. 34, no. 2, pp. 149–153, Jun. 2011.
[6] A.-K. Seifert, A. M. Zoubir, and M. G. Amin, "Radar-based human gait recognition in cane-assisted walks," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2017, pp. 1428–1433.
[7] F. Ahmed, A. S. M. H. Bari, and M. L. Gavrilova, "Emotion recognition from body movement," *IEEE Access*, vol. 8, pp. 11761–11781, 2020.
[8] C. Luo, J. Wu, J. Li, J. Wang, W. Xu, Z. Ming, B. Wei, W. Li, and A. Y. Zomaya, "Gait recognition as a service for unobtrusive user identification in smart spaces," *ACM Trans. Internet Things*, vol. 1, no. 1, pp. 1–21, Mar. 2020.
[9] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, Feb. 2005.
[10] J. Man and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.
[11] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, "Gait recognition using a view transformation model in the frequency domain," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2006, pp. 151–163.
[12] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait recognition from multiple views," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2001, p. 439.
[13] C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2002, pp. 372–377.
[14] D. Cunado, M. S. Nixon, and J. N. Carter, "Using gait as a biometric, via phase-weighted magnitude spectra," in *Proc. Int. Conf. Audio Video-Based Biometric Person Authentication*. Cham, Switzerland: Springer, 1997, pp. 93–102.
[15] A. Y. Johnson and A. F. Bobick, "A multi-view method for gait recognition using static body parameters," in *Proc. Int. Conf. Audio Video-Based Biometric Person Authentication*. Cham, Switzerland: Springer, 2001, pp. 301–311.
[16] J.-H. Yoo, D. Hwang, K.-Y. Moon, and M. S. Nixon, "Automated human recognition by gait using neural network," in *Proc. 1st Workshops Image Process. Theory, Tools Appl.*, Nov. 2008, pp. 1–6.

[17] A. Sepas-Moghaddam and A. Etemad, "Deep gait recognition: A survey," 2021, *arXiv:2102.09546*.

[18] K. R. Shetty, V. S. Soorinje, and P. Dsouza, "Deep learning for computer vision: A brief review," *Int. J. Adv. Res. Sci., Commun. Technol.*, vol. 2018, pp. 450–463, Mar. 2022.

[19] D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, M. J. Marín-Jiménez, and R. Muñoz-Salinas, "The AVA multi-view dataset for gait recognition," in *Proc. Int. Workshop Activity Monitor. Multiple Distrib. Sens.*, Cham, Switzerland: Springer, 2014, pp. 26–39.

[20] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, Dec. 2003.

[21] P. Connor and A. Ross, "Biometric recognition by gait: A survey of modalities and features," *Comput. Vis. Image Understand.*, vol. 167, pp. 1–27, Feb. 2018.

[22] A. Zhao, L. Qi, J. Dong, and H. Yu, "Dual channel LSTM based multi-feature extraction in gait for diagnosis of Neurodegenerative diseases," *Knowl. Based Syst.*, vol. 145, pp. 91–97, Apr. 2018.

[23] Q. Li, Y. Wang, A. Sharf, Y. Cao, C. Tu, B. Chen, and S. Yu, "Classification of gait anomalies from Kinect," *Vis. Comput.*, vol. 34, no. 2, pp. 229–241, Feb. 2018.

[24] I. Bouchrika, "A survey of using biometrics for smart visual surveillance: Gait recognition," in *Surveillance in Action*. Cham, Switzerland: Springer, 2018, pp. 3–23.

[25] Z. Chen, C. Song, Y. Yang, B. Zhao, Y. Hu, S. Liu, and J. Zhang, "Robot navigation based on human trajectory prediction and multiple travel modes," *Appl. Sci.*, vol. 8, no. 11, p. 2205, Nov. 2018.

[26] K. Pfeuffer, M. J. Geiger, S. Prange, L. Mecke, D. Buschek, and F. Alt, "Behavioural biometrics in vr: Identifying people from body motion and relations in virtual reality," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–12.

[27] M. Hofmann and G. Rigoll, "Exploiting gradient histograms for gait-based person identification," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 4171–4175.

[28] X.-T. Chen, Z.-H. Fan, H. Wang, and Z.-Q. Li, "Automatic gait recognition using kernel principal component analysis," in *Proc. Int. Conf. Biomed. Eng. Comput. Sci.*, Apr. 2010, pp. 1–4.

[29] 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 *Proc. IEEE 2nd Int. Conf. Biometrics, Theory, Appl. Syst.*, Sep. 2008, pp. 1–6.

[30] G. Ariyanto and M. S. Nixon, "Model-based 3D gait biometrics," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–7.

[31] Y. Tian, L. Wei, S. Lu, and T. Huang, "Free-view gait recognition," *PLoS ONE*, vol. 14, no. 4, Apr. 2019, Art. no. e0214389.

[32] M. W. Rahman and M. L. Gavrilova, "Kinect gait skeletal joint feature-based person identification," in *Proc. IEEE 16th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI*CC)*, Jul. 2017, pp. 423–430.

[33] A. S. M. H. Bari and M. L. Gavrilova, "Artificial neural network based gait recognition using Kinect sensor," *IEEE Access*, vol. 7, pp. 162708–162722, 2019.

[34] C. Zhang, W. Liu, H. Ma, and H. Fu, "Siamese neural network based gait recognition for human identification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 2832–2836.

[35] Y. He, J. Zhang, H. Shan, and L. Wang, "Multi-task GANs for view-specific feature learning in gait recognition," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 1, pp. 102–113, Jan. 2019.

[36] Z. Zhang, L. Tran, X. Yin, Y. Atoum, X. Liu, J. Wan, and N. Wang, "Gait recognition via disentangled representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4710–4719.

[37] H. Chao, Y. He, J. Zhang, and J. Feng, "Gaitset: Regarding gait as a set for cross-view gait recognition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 8126–8133.

[38] F. Battistone and A. Petrosino, "TGLSTM: A time based graph deep learning approach to gait recognition," *Pattern Recognit. Lett.*, vol. 126, pp. 132–138, Sep. 2019.

[39] K. Lin, L. Wang, and Z. Liu, "Mesh graphormer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 12939–12948.

[40] F. M. Castro, M. J. Marín-Jiménez, N. G. Mata, and R. Muñoz-Salinas, "Fisher motion descriptor for multiview gait recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 1, Jan. 2017, Art. no. 1756002.

[41] M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll, "The TUM gait from audio, image and depth (GAID) database: Multimodal recognition of subjects and traits," *J. Vis. Commun. Image Represent.*, vol. 25, no. 1, pp. 195–206, Jan. 2014.

[42] D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer, "Entropy volumes for viewpoint-independent gait recognition," *Mach. Vis. Appl.*, vol. 26, nos. 7–8, pp. 1079–1094, Nov. 2015.

[43] D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, M. J. Marín-Jiménez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Viewpoint-independent gait recognition through morphological descriptions of 3D human reconstructions," *Image Vis. Comput.*, vols. 48–49, pp. 1–13, Apr. 2016.

[44] H. Arshad, M. A. Khan, M. I. Sharif, M. Yasmin, J. M. R. S. Tavares, Y. Zhang, and S. C. Satapathy, "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition," *Expert Syst.*, Mar. 2020, Art. no. e12541.

[45] Z. Zhu, X. Guo, T. Yang, J. Huang, J. Deng, G. Huang, D. Du, J. Lu, and J. Zhou, "Gait recognition in the wild: A benchmark," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 14789–14799.

[46] A. K.-F. Lui, Y.-H. Chan, and M.-F. Leung, "Modelling of destinations for data-driven pedestrian trajectory prediction in public buildings," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2021, pp. 1709–1717.

[47] G. Chen, J. Li, J. Lu, and J. Zhou, "Human trajectory prediction via counterfactual analysis," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9824–9833.

[48] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, "Deep learning-based human pose estimation: A survey," 2020, *arXiv:2012.13392*.

[49] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," 2018, *arXiv:1812.08008*.

[50] E. Auvinet, F. Multon, C.-E. Aubin, J. Meunier, and M. Raison, "Detection of gait cycles in treadmill walking using a Kinect," *Gait Posture*, vol. 41, no. 2, pp. 722–725, Feb. 2015.

[51] C. BenAbdelkader, R. G. Cutler, and L. S. Davis, "Gait recognition using image self-similarity," *EURASIP J. Adv. Signal Process.*, vol. 2004, no. 4, Dec. 2004, Art. no. 721765.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[53] S. Pouyanfar, T. Wang, and S.-C. Chen, "Residual attention-based fusion for video classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 478–480.

[54] L. Rice, E. Wong, and Z. Kolter, "Overfitting in adversarially robust deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 8093–8104.

[55] J. Patterson and A. Gibson, *Deep Learning: A practitioner's Approach*. Sebastopol, CA, USA: O'Reilly Media, 2017.

[56] L. Datta, "A survey on activation functions and their relation with xavier and He normal initialization," 2020, *arXiv:2004.06632*.

[57] D. Choi, C. J. Shallue, Z. Nado, J. Lee, C. J. Maddison, and G. E. Dahl, "On empirical comparisons of optimizers for deep learning," 2019, *arXiv:1910.05446*.

[58] B. Schmeiser, "Batch size effects in the analysis of simulation output," *Oper. Res.*, vol. 30, no. 3, pp. 556–568, Jun. 1982.

[59] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," 2016, *arXiv:1609.04836*.

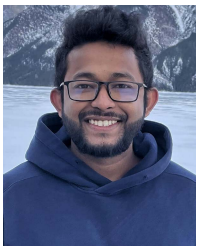
[60] S. Hayou, A. Doucet, and J. Rousseau, "On the impact of the activation function on deep neural networks training," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2672–2680.

[61] A. Fred Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.

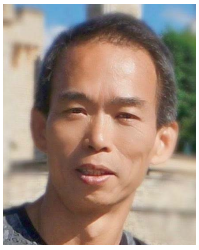
[62] X. Ying, "An overview of overfitting and its solutions," *J. Phys., Conf. Ser.*, vol. 1168, no. 2, 2019, Art. no. 022022.

[63] M. Bukhari, K. B. Bajwa, S. Gillani, M. Maqsood, M. Y. Durrani, I. Mehmood, H. Ugail, and S. Rho, "An efficient gait recognition method for known and unknown covariate conditions," *IEEE Access*, vol. 9, pp. 6465–6477, 2021.

[64] R. Delgado-Escano, F. M. Castro, N. Guil, and M. J. Marin-Jimenez, "GaitCopy: Disentangling appearance for gait recognition by signature copy," *IEEE Access*, vol. 9, pp. 164339–164347, 2021.



MD. SHOPON (Member, IEEE) received the B.Sc. degree in computer science and engineering from the University of Asia Pacific, in 2018. He is currently pursuing the M.Sc. degree in computer science with the University of Calgary, Canada, under the supervision of Prof. Marina L. Gavrilova. From September 2018 to September 2020, he worked as a Lecturer with the University of Asia Pacific. He has authored over 15 international conference papers and journals. His research interests include computer vision, deep learning, and behavioral biometrics.



GEE-SERN JISON HSU (Senior Member, IEEE) received the dual M.S. degree in electrical and mechanical engineering and the Ph.D. degree in mechanical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1993 and 1995, respectively. From 1995 to 1996, he was a Postdoctoral Fellow with the University of Michigan. From 1997 to 2000, he was a Senior Research Staff with the National University of Singapore, Singapore. In 2001, he joined PenPower Technology, where he led research on face recognition and intelligent video surveillance. In 2007, he joined the Department of Mechanical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan. His research interests include computer vision, deep learning, and particularly face recognition and license plate recognition. He is a member of ACM, IAPR, and IEICE. He received the Best/Outstanding Paper Award from ICMT 2011, CVGIP 2013, and CVPRW 2014. He and his team at PenPower Technology were recipients of the Best Innovation Award at the SecuTech Expo for three consecutive years from 2005 to 2007.

MARINA L. GAVRILOVA (Senior Member, IEEE) is currently a Full Professor at the University of Calgary and an international expert in the area of biometric security, machine learning, pattern recognition, and information fusion. She directs the Biometric Technologies Laboratory, and published over 300 books, conference proceedings, and peer-reviewed articles. She delivered over 50 invited keynote speeches and panel presentations at premium international conferences and research centers worldwide. Her professional excellence was recognized by the Canada Foundation for Innovation, the Killam Foundation, U Make a Difference Award, and the prestigious Order of the University of Calgary. Her research was recognized by the Best Paper Awards at the IEEE SMC International Conference on Systems, Men and Cybernetic, the IEEE International Conference on Cognitive Informatics Cognitive Computing, the IEEE Symposium on 3-D User Interfaces, and the IEEE/ACM International Conference on CyberWorlds. She is a Founding Editor-in-Chief of *Transactions on Computational Sciences* (Springer) and serves on the editorial boards for the IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SCIENCES, IEEE ACCESS, *The Visual Computer*, *Sensors*, and seven other journals.