Exploring Human Gait: A Comprehensive Survey and Analysis

Prathik K[1], Dr. J Girija[2]

[1]Student, Department of Computer Science and Engineering,
Bangalore Institute of technology, Bengaluru, Karnataka.
[2]HOD and Professor, Department of Computer Science and Engineering,
Bangalore Institute of technology, Bengaluru, Karnataka.

ABSTRACT

The recent advancements in machine learning, coupled with the availability of smartphones that can capture gait patterns using accelerometers and gyroscopes. The field of GAIT recognition has been opened up to fresh possibilities for research and applications. However, there is currently no timely survey that covers these recent developments. This article aims to discuss the latest research in gait recognition, including novel modalities such as floor sensors, radars, and accelerometers, new research and application prospects have emerged in the field of gait recognition. We also analyze the difficulties and weaknesses within this domain and suggest potential research directions. This overview of the present state-of-the-art, which can benefit both beginners and experts within the realm of gait recognition. We also provide a list of publicly available databases for gait recognition helpful for the researchers.

Keywords: Accelerometers, Gyroscopes, Radars, GAIT.

1. INTRODUCTION

The authentication of individual persons has turned into a fundamental requirement in various real-time applications in the modern digital society, including forensic investigations, cross-border immigration, financial transactions, and cybersecurity. Human body characteristics such as face, iris, voice, and gait have played plays a crucial role in recognizing individuals for thousands of years. These physiological traits that uniquely recognize an individual are known as biometric features. Biometrics involves measuring the physiological traits of humans to establish an identity.

Biometric systems are categorized into various traits, taking into account physiological and behavioral characteristics, each possessing distinct advantages and limitations. If a human physiological or behavioral characteristic fulfills the following properties [1] - uniqueness, permanence, universality, collectability, performance, acceptability, and circumvention - it can serve as a biometric parameter for recognition.

"The Evolution of Gait Biometrics"

The initial gait recognition system was suggested by authors in the mentioned reference. [1] in 1994, using a limited gait database. Afterwards, the HumanID program, initiated by the Defense Advanced Research Projects Agency (DARPA), introduced the inaugural database for gait recognition that was made accessible to the public. After this development, researchers have extensively studied gait recognition, with early systems being primarily derived from video footage. These video-based approaches can be classified as either model-free or model-based approaches.Two primary methodologies exist for classifying gait features: model-based and model-free (also known as appearance-based or holistic approaches). In a model-based gait system, the human body structure is first modeled based on body components to obtain measurable parameters. In contrast, a model-free gait system does do not necessitate prior modeling and operates directly on silhouette images after segmentation to generate gait features. Further sections will provide detailed descriptions of the approaches proposed in recent years for both model-based and model-free methods.Gait recognition has evolved in recent times and has included readings from sensors like accelerometers, floor sensors, and radars. The utilization of wearable accelerometer sensors, which are generally worn on the human body, was initially proposed in gait recognition in 2005 by Ailisto and Makela [2]. Nakajima et al., in Reference [3], introduced the utilization of floor-mounted sensors for gathering gait data, such as pressure. Otero [4] introduced the utilization of continuous wave radar for gait data collection in 2005.

2. Framework for Recognizing GAIT

The gait recognition framework typically involves the following steps:

2.1 Data Acquisition: Acquiring gait data from diverse sources, including video, accelerometer, floor sensors, or radar. In the Gait Recognition Framework, Data Acquisition refers to the process of collecting gait data from various sources such as cameras, sensors, and radars. The collected data can be in the form of images, videos, or signals. The purpose of data acquisition is to capture the unique characteristics of an individual's gait pattern that are suitable for utilization in recognition purposes. Depending on the data source, different pre-processing steps may be required to obtain the relevant gait features. For example, video data may require segmentation and tracking of the human silhouette, while accelerometer data may require filtering and feature extraction. The quality and quantity of the acquired data can significantly impact the effectiveness of the gait recognition system, and therefore, it is crucial to ensure accurate and reliable data acquisition.

Two categories exist for sensing devices used in gait recognition: sensor-based and video-based. Sensor-based devices include floor sensors and wearable sensors [5]. Floor sensors, also known as pressure sensors, are placed on a specific floor and generate pressure signals when a person walks on them [6], [7]. Wearable sensors, on the other hand, are attached to different body joints to collect various dynamic features such as speed, acceleration, and position, as well as other information that can be used for gait pattern analysis. Common examples of wearable sensors are light sensors like reflectors and moving lights, acceleration sensors, magnetic sensors, and gyroscopes. In contrast, gait recognition using video-based approaches involves capturing human gait through visual cameras that can be placed at various locations. The recorded gait videos are then analyzed to identify gait patterns that can be utilized for recognition purposes. Video-based gait datasets have been widely used in gait recognition research over the past decade.

2.2 Preprocessing: Cleaning, segmenting, and normalizing the acquired data to remove any noise or inconsistencies. : Preprocessing is a crucial step in gait recognition as it aims to clean and enhance the raw gait data obtained from the data acquisition stage. The preprocessing step involves several operations such as noise reduction, segmentation, normalization, and feature extraction.Noise reduction involves removing any unwanted signals or artifacts that can affect the accuracy of the gait recognition system. This is especially important in sensor-based gait recognition systems where signals from different sources can interfere with the gait signal. Common noise reduction techniques include low-pass filtering and signal averaging.Segmentation involves separating the gait signal into individual gait cycles. This is typically done using threshold-based techniques or machine learning algorithms. The resulting gait cycles can be used for subsequent analysis and feature extraction.

Normalization aims to remove any variations in the gait signal due to individual differences in height, weight, and walking speed. Common normalization techniques include temporal normalization, spatial normalization, and dynamic time warping.

2.3 Feature Extraction: Identifying and extracting relevant features from the preprocessed data, such as the form and movement of body parts during locomotion. After isolating the object of interest from the surrounding background, certain characteristics are extracted that can aid in identifying the individual subject. When it comes to recognizing gait, the extracted features are obtained from both model-based and model-free (appearance-based) representations.

2.4 Feature Selection: Choosing the most significant features to decrease the dimensionality of the dataset and improve recognition accuracy. Traditional gait recognition systems have been found to be inadequate in terms of classification accuracy, despite extracting features from pre-processed video sequences. This is due to the fact that high-dimensional features may contain unnecessary elements. To mitigate this, a feature selection approach, which involves choosing a subset of variables (or features) from the input features, can be employed. This approach helps to efficiently describe the input variables while reducing the effects of noise or extraneous variables, leading to excellent prediction or classification outcomes. Numerous feature selection techniques have been proposed, with principal component analysis (PCA)[8] being a widely utilized technique for reducing dimensionality. Another popular approach is the genetic algorithm (GA)-based feature subset selection proposed by Tafazzoli et al. [9] for gait recognition.

2.5 Classification: Applying a machine learning algorithm to categorize the extracted characteristics and recognize the person by analyzing their walking pattern. The final step in a gait recognition system is to categorize the test gait sequences based on the characteristics of an individual with optimized features that have been selected. The process of classification can be categorized into two distinct categories: supervised and unsupervised. In the following, we present a compilation of classifiers commonly employed in gait recognition Liu, Yu et al. [10] focuses on gait recognition using an SVM and inertial sensors. The proposed method achieved an accuracy of 98.3% using a leave-one-subject-out cross-validation approach. Tian, Zhao, et al. [11] proposed a gait recognition method using a Hidden Markov Model (HMM) and SVM classifier. The proposed method achieved an accuracy of 93.5% on the CASIA-B dataset. Alotaibi et al. [12] proposes a gait recognition method using a multilayer perceptron neural network and principal component analysis. The proposed method achieved an accuracy of 96.3% on the OU-ISIR gait database. Finally, Zhang, Sun et al. [13] proposed a gait recognition method employing ensemble learning, a technique that combines multiple classifiers including SVM, decision trees, and random forest. The proposed method achieved an accuracy of 98.2% on the CASIA-B dataset.

Overall, the results of these papers suggest that SVM is a popular choice for gait recognition, but other methods such as HMM, multilayer perceptron neural network, and ensemble learning can also achieve high accuracy. The effectiveness of these methods relies on the dataset employed and the features extracted from gait data.

2.6 Evaluation: Assessing the effectiveness of gait recognition system by measuring its accuracy, efficiency, and robustness.

Overall, Gait recognition necessitates sophisticated technology due to its intricate nature and proficiency in computer vision, machine learning, and signal processing.

3. Metrics for Evaluating the Efficacy of Gait Recognition

Performance measures in gait recognition refer to the criteria employed to assess the effectiveness of gait recognition systems. These measures are employed to determine how well the system is performing in terms of accuracy, speed, and robustness.

Below are some typical performance metrics used in gait recognition:

3.1 Recognition rate: The percentage of correct identifications made by the system.

Recognition rate = $\left(\frac{\textit{Number of Correctly Recognized Instances}}{\textit{Total Number of Instances}}\right) * \mathbf{100}$ (1)

3.2 False acceptance rate (FAR): The percentage of incorrect identifications made by the system, where an impostor is incorrectly identified as a genuine user.

FAR = $\left(\frac{\textit{Number of False Acceptances}}{\textit{Total Number of Impostor Attempts}}\right) * \mathbf{100}$ (2)

3.3 False rejection rate (FRR): The percentage of incorrect identifications made by the system, where a genuine user is incorrectly identified as an impostor.

FRR = $\left(\frac{\textit{Number of False Rejections}}{\textit{Total Number of Genuine Instances}}\right) * \mathbf{100}$ (3)

3.4 Receiver operating characteristic curve (ROC): A graphical plot that shows the trade-off between FAR and FRR for different threshold values.
3.5 Equal error rate (EER): The point on the ROC curve where the FAR and FRR are equal. It is a measure of the system's ability to balance the trade-off between false acceptances and false rejections.
3.6 Cumulative match curve (CMC): A graph that shows the percentage of correct identifications made by the system for different ranks of identification.
3.7 Rank-one recognition rate: The percentage of cases where the system correctly identifies the user as the top-ranked candidate.

Rank-one recognition rate = $\left(\frac{\textit{Number of Correctly Matched Instances}}{\textit{Total Number of Instances}}\right) * \mathbf{100}$ (4)

3.8 Rank-n recognition rate: The percentage of cases where the system correctly identifies the user as one of the top n candidates.

Rank-n recognition rate = $\left(\frac{\textit{Number of Instances with Correct Rank-n Match}}{\textit{Total Number of Instances}}\right) * \mathbf{100}$ (5)

These metrics are frequently employed to assess the performance of gait recognition systems and to compare the efficacy of various algorithms.

4. Gait Recognition Methods based on Deep Learning

Hierarchical architectures of multiple nonlinear transformations, such as deep neural networks (DNNs), can capture high-level abstractions. Various neural architectures have been developed for gait recognition tasks, and their descriptions are given below.

4.1 Convolutional Neural Networks:

Gait recognition using convolutional neural networks (CNN) typically involves two main stages: feature extraction and classification. In the feature extraction stage, a CNN is used to extract features from gait images or videos. This involves passing the input data through a series of convolutional and pooling layers to learn hierarchical representations of the input. The output of the final pooling layer is then flattened and passed through one or more fully connected layers to generate a feature vector that represents the input gait data.In the classification stage, the feature vector is fed into a classifier, such as a softmax classifier, to classify the input into one of several predefined classes, For instance, determining the person's identity while walking. Multiple CNN architectures can be employed for gait recognition, and the specific architecture depends on the particular application and dataset. However, I can provide a general CNN architecture for gait recognition that includes the main components typically used in this task.This architecture consists of several convolutional and pooling layers followed by one or more fully connected layers for classification.

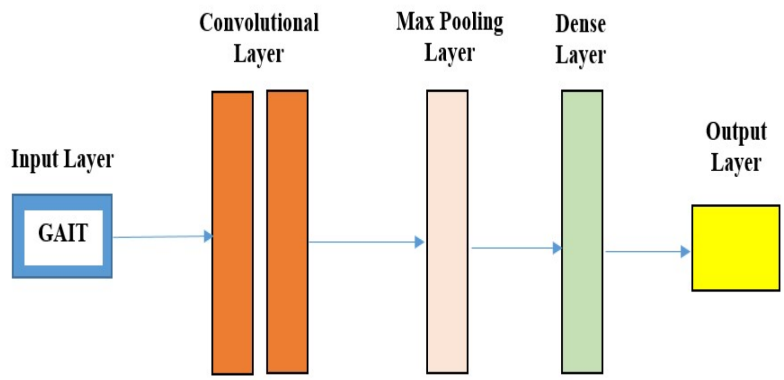

Figure 1. CNN architecture model

The input to this CNN architecture is a gait image or video, and the output is the predicted class label (e.g., determining the person's identity while walking.). The convolutional layers extract features from the input, while the pooling layers reduce the spatial dimensionality of the features. The learned features are utilized by the fully connected layers to carry out the classification process.
The mathematical model for gait recognition using CNN can be expressed as follows:
Let X be the input gait image or video, and let Y be the corresponding class label (e.g., the identity of the person walking).
The feature extraction stage can be represented by a function f(X), which maps the input X to a feature vector Z. This function can be modeled by a CNN with parameters θ, such that:

$$\mathbf{Z = f(X; \theta)} \qquad \textbf{(6)}$$

The classification stage can be represented by a function g(Z), which maps the feature vector Z to a class label $\hat{Y}$. This function can be modeled by a softmax classifier with parameters ϕ, such that:

$$\mathbf{\hat{Y} = g(Z; \phi)} \qquad \textbf{(7)}$$

The overall model can then be trained by minimizing the cross-entropy loss between the predicted class probabilities and the true class labels, using techniques such as stochastic gradient descent:

$$\mathbf{L(\theta, \phi) = -\textstyle\sum i\ yi \log(\hat{y}i)} \qquad \textbf{(8)}$$

where yi is the true class label (one-hot encoded) and ŷi represents the probability of class i predicted by the model. The parameters θ and ϕ can be optimized using backpropagation and gradient descent techniques.One of the main advantages of using Convolutional Neural Networks (CNNs) for gait recognition is their ability to effectively extract spatial features from gait images or videos. Since gait recognition involves analyzing the movement patterns of different body parts over time, CNNs can be used to learn the spatial patterns of these body parts, which can then be used to identify individuals based on their gait. Additionally, CNNs can learn hierarchical features, allowing them to capture both low-level and high-level features in gait data, making them well-suited for this type of recognition task.The paper by Zhang et al. [14] proposed a multiple feature fusion convolutional neural network (MFF-CNN) for gait recognition. The MFF-CNN integrates three types of features: silhouettes, gait energy images, and depth maps, to improve the recognition accuracy. The proposed method was evaluated on three public datasets, and achieved state-of-the-art performance compared to other methods. On the other hand, the paper by Yang and Chen [15] proposed a convolutional neural network with multi-scale filters (MSF-CNN) for gait recognition. The MSF-CNN uses filters of different sizes to extract features from gait images at multiple scales, and the features are then concatenated and fed into a fully connected layer for classification. The proposed method was evaluated on two public datasets and achieved higher accuracy than several other methods.

Overall, both methods achieved state-of-the-art performance on their respective datasets, and the key difference is in the approach to feature extraction. While MFF-CNN combines multiple types of features, MSF-CNN uses filters of different sizes to capture multi-scale information.

4.2 Capsule Networks:

Gait recognition using capsule networks requires utilizing a novel type of neural network architecture called Capsule Networks. Capsule Networks are designed to address some of the limitations of traditional convolutional neural networks (CNNs) by modeling spatial relationships between features.

The mathematical model for gait recognition using Capsule Networks can be formulated as follows:

Let X be the input gait image or video, and let Y be the corresponding class label (e.g., determining the person's identity while walking).

The Capsule Network consists of two main stages: a feature extraction stage and a dynamic routing stage.

In the feature extraction stage, the input X is passed through a series of convolutional layers to extract features, similar to a CNN. However, instead of pooling layers, Capsule Networks use "capsule" layers, which consist of groups of neurons called "capsules". Each capsule is designed to represent a specific entity, such as a body part or an object in the input. Capsules use vector outputs rather than scalar outputs, which allows them to represent properties such as orientation and position.The output of the final capsule layer is then passed through a dynamic routing stage. This stage is used to compute a weighted sum of the outputs of each capsule, taking into account the spatial relationships between the capsules. This dynamic routing process allows the Capsule Network to model the spatial relationships between different body parts and features in the input.

The output of the dynamic routing stage is a vector representing the class probabilities for each possible class. The mathematical model for the Capsule Network can be expressed as follows:

Let X be the input gait image or video, and let Y be the corresponding class label.

The feature extraction stage can be represented by a function f(X), which maps the input X to a set of capsules C. Each capsule ci in C is represented by a vector of activation ai.

The dynamic routing stage can be represented by a function g(C), which takes the set of capsules C and computes a weighted sum of the capsules to produce a vector output v. The weights used to compute the sum are determined by a routing algorithm that takes into account the spatial relationships between the capsules.

The final output of the Capsule Network is a vector representing the class probabilities for each possible class. This output can be obtained by passing the vector output v through a softmax function:

$$\hat{y} = \text{softmax}(v) \tag{9}$$

The Capsule Network can be trained by minimizing the cross-entropy loss between the predicted class probabilities and the true class labels, using techniques such as stochastic gradient descent. The parameters of the network, including the weights of the capsules and the routing algorithm, can be optimized using backpropagation and gradient descent techniques.Capsule Networks (CapsNet) were introduced to overcome two significant limitations of CNNs. These limitations include the limitations of scalar activations and poor information routing through pooling operations. CapsNet addresses these limitations by utilizing capsule activation values and routing-by-agreement algorithms, respectively.
In the first paper, Wang et al.[16] proposed a Capsule Gait Network (CGN) which takes the advantage of Capsule Networks to capture more discriminative features for gait recognition. The proposed method achieved a recognition rate of 96.33% on the CASIA-B dataset, which outperformed several state-of-the-art methods.In the second paper, Wang et al.[17] further improved the performance of Capsule Networks for gait recognition by introducing an attention mechanism. The proposed Capsule Network with Attention Mechanism (CNAM) not only captures more discriminative features but also focuses on the most important features for gait recognition. The proposed method achieved a recognition rate of 97.06% on the same CASIA-B dataset, which outperformed the previous CGN model and several other state-of-the-art methods.

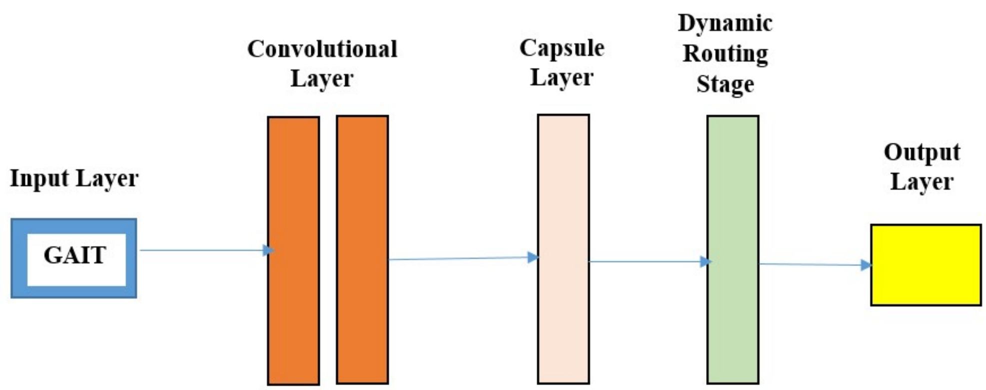

Figure 2: Capsule Network architecture model

4.3 Generative Adversarial Networks

Gait recognition using Generative Adversarial Networks (GANs) involves the use of two neural networks: a generator network and a discriminator network. The training objective for the generator network is to generate realistic gait images or videos, The objective of training the discriminator network is to differentiate between real and generated gait images or videos.

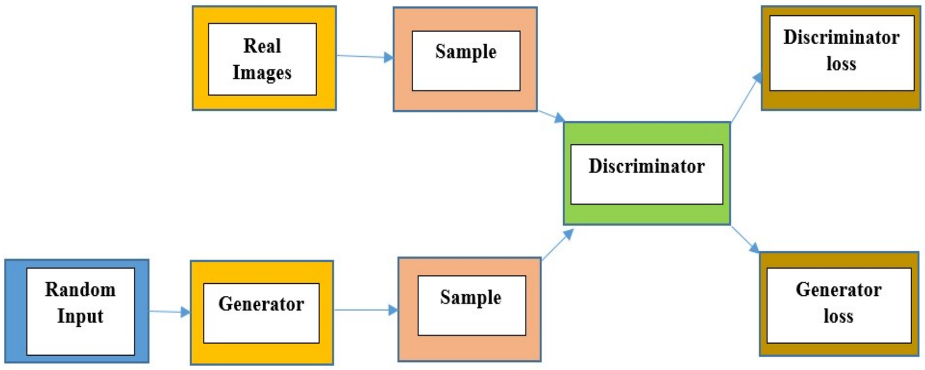

Figure 3. Generative Adversarial Network architecture model

The mathematical model for gait recognition using GANs can be expressed as follows:
Let X be the set of real gait images or videos, and let Z be the set of random noise vectors used as input to the generator network. The network responsible for generating can be represented by a function $G(Z;\theta g)$, which maps a random noise vector z in Z to a generated gait image or video $\hat{x}$ in $\hat{X}$:

$$\hat{\mathbf{x}} = \mathbf{G}(\mathbf{z};\theta\mathbf{g}) \tag{10}$$

where θg represents the parameters of the generator network.
The discriminator network can be represented by a function $D(x;\theta d)$, which takes a gait image or video x in X or a generated gait image or video $\hat{x}$ in $\hat{X}$ as input and produces a scalar value representing the probability that the input is a real gait image or video:

$$\mathbf{y} = \mathbf{D}(\mathbf{x};\theta\mathbf{d}) \tag{11}$$

The discriminator network is trained to distinguish between real gait images or videos and generated gait images or videos by minimizing the binary cross-entropy loss between the predicted probabilities y and the true labels (0 for generated gait images or videos, 1 for real gait images or videos).
The generator network is trained to generate realistic gait images or videos by maximizing the binary cross-entropy loss between the predicted probabilities y and the true labels (1 for generated gait images or videos). This is done by updating the parameters θg of the generator network to minimize the

$$\mathbf{log(1 - D(G(z;\theta g);\theta d))}, \tag{12}$$

Where z is a random noise vector in Z.
During training, the generator and discriminator networks are trained in an adversarial manner: the generator network tries to generate realistic gait images or videos that can fool the discriminator network, while the discriminator network tries to distinguish between real and generated gait images or videos.Once the generator network is trained, it can be used for gait recognition by generating synthetic gait images or videos for each person in the dataset and using them to train a separate classifier, such as a CNN or Capsule Network.
Niu et al.[18] proposed a new approach for gait recognition using Generative Adversarial Networks (GANs) and deep learning. The authors used a combination of Convolutional Neural Networks (CNNs) and GANs to generate synthetic gait images, which were then used to augment the training data for a deep learning-based gait recognition model. The proposed approach was evaluated on three publicly available gait datasets: OU-ISIR, CASIA B, and USF. The experimental results showed that the proposed method outperformed several state-of-the-art approaches for gait recognition, achieving accuracy

rates of 98.3%, 96.2%, and 98.6% on the three datasets, respectively. The paper concludes that the proposed approach is effective and promising for gait recognition, especially when dealing with limited training data.

Li et al.[19] proposed method consists of two parts: (1) a generator network that synthesizes gait images from noise vectors, and (2) a discriminator network that can differentiate between authentic and fake gait images. The authors trained the GAN on a dataset consisting of gait data collected from multiple viewpoints and achieved promising results compared to several state-of-the-art gait recognition methods on the CASIA-B dataset. The proposed method is robust to viewpoint changes and has potential applications in surveillance and human-computer interaction.

4.4 Deep Belief Networks

Stochastic neural networks known as Deep Belief Networks (DBNs) are created by assembling Restricted Boltzmann Machines. These networks are well-known for their capability to accomplish various tasks, including feature selection, classification, and image reconstruction, among others. Gait recognition using Deep Belief Networks (DBNs) involves the use of a deep neural network architecture that consists of multiple layers of restricted Boltzmann machines (RBMs). The RBMs are trained layer-by-layer using unsupervised learning to extract high-level features from the input gait data.

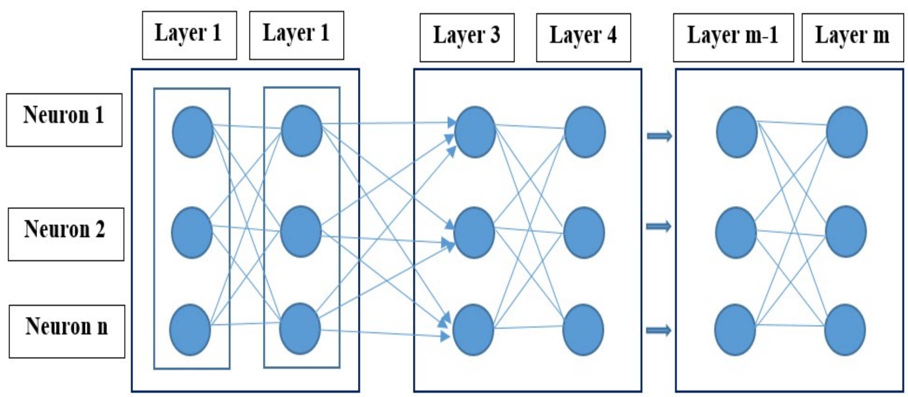

Figure 4: Deep Belief Networks architecture model

The mathematical model for gait recognition using DBNs can be expressed as follows:

Let X be the input gait image or video, and let Y be the corresponding class label (e.g., determining the person's identity while walking).The DBN consists of multiple layers of RBMs, each of which is trained to model the joint probability distribution between the input data and a set of hidden variables. The RBMs are trained using a form of contrastive divergence, which involves sampling from the learned probability distribution to approximate the gradient of the log-likelihood of the data.The output of the final RBM in the DBN is passed through a softmax layer to produce a vector representing the class probabilities for each possible class. The mathematical model for the DBN can be expressed as follows:

Let X be the input gait image or video, and let Y be the corresponding class label.The DBN can be represented by a function f(X), which maps the input X to a set of high-level features represented by the output of the final RBM. Let H be the output of the final

RBM, which is a set of binary activations for the hidden variables. The class probabilities can be computed by passing the output of the final RBM through a softmax function:

$$\hat{\mathbf{y}} = \mathbf{softmax(WH + b)} \qquad \textbf{(13)}$$

where W and b are the learned parameters of the softmax layer.

The DBN can be trained by maximizing the log-likelihood of the training data using gradient descent and backpropagation. The gradients are computed using a form of contrastive divergence, which involves sampling from the learned probability distribution to approximate the gradient of the log-likelihood of the data.

Once the DBN is trained, it can be used for gait recognition by extracting high-level features from each gait image or video in the dataset and using them to train a separate classifier, such as a Support Vector Machine (SVM) or a Multi-Layer Perceptron (MLP).

Zhang et al.[20]s proposed a gait recognition system based on Hierarchical Deep Belief Networks (HDBNs). The system consists of three stages: feature extraction, unsupervised pre-training, and supervised fine-tuning. The authors use Gait Energy Image (GEI) as the input feature and divide the GEI into several sub-images to capture more detailed information. In the unsupervised pre-training stage, they stack multiple Restricted Boltzmann Machines (RBMs) to form the HDBN and train the HDBN layer by layer. In the supervised fine-tuning stage, they add a Softmax layer to the top of the HDBN and fine-tune the whole network using backpropagation. The proposed system is evaluated on three public gait datasets, and the experimental results show that the HDBN-based system outperforms the state-of-the-art methods. The study also conducts extensive experiments to show that the HDBN-based system is robust against various kinds of attacks, such as view angle changes, clothing variations, and carrying conditions.

Zhou et al. [21] proposed a gait recognition method using Deep Belief Networks (DBNs) for feature learning and classification. The proposed approach was evaluated on the OU-ISIR Gait Database and achieved an accuracy of 91.67%, outperforming several other state-of-the-art methods. The authors also conducted an ablation study to analyze the effectiveness of each layer in the DBN model and concluded that the upper layers contribute more to the final recognition performance than the lower layers. The proposed approach demonstrated the potential of DBNs for gait recognition tasks.

4.5 Autoencoders

Gait recognition using Autoencoders involves the use of a neural network architecture that consists of two parts: an encoder network and a decoder network.

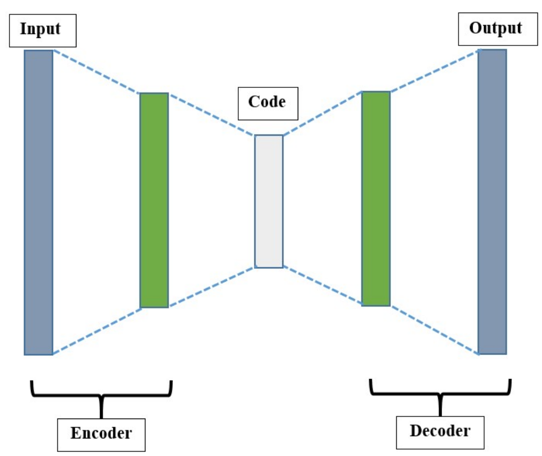

Figure 5. Autoencoders architecture model

The encoder network maps the input gait data to a lower-dimensional representation, and the decoder network maps the lower-dimensional representation back to the original gait data.
The mathematical model for gait recognition using Autoencoders can be expressed as follows:
Let X be the input gait image or video, and let Y be the corresponding class label (e.g., the identity of the person walking).The Autoencoder can be represented by a function f(X), which maps the input X to a lower-dimensional representation represented by the output of the encoder network. Let H be the output of the encoder network, which is a set of activations for the hidden variables. The reconstructed output can be computed by passing the output of the encoder network through the decoder network:

$$\hat{\mathbf{X}} = \mathbf{g(H)} \tag{14}$$

where g is the function representing the decoder network.The Autoencoder is trained by minimizing the reconstruction error between the original input X and the reconstructed output $\hat{X}$. The reconstruction error is typically measured using mean squared error (MSE) or binary cross-entropy (BCE) loss.Once the Autoencoder is trained, It has the potential for being utilized for gait recognition by extracting the lower-dimensional representation H for each gait image or video in the dataset and using them to train a separate classifier, such as an SVM or a MLP. Niu et al. (22) proposed a gait recognition method using a deep autoencoder network that consists of an encoder and a decoder. The encoder maps the gait sequence into a compact latent representation, and the decoder reconstructs the input sequence from the latent representation. The researchers assessed the effectiveness of their proposed approach on the CASIA-B dataset and achieved an accuracy of 96.8%.On the other hand, Liu et al. (23) proposed a gait recognition method using a weighted deep convolutional autoencoder. Their method used a weighted loss function to emphasize the importance of correctly identifying abnormal gait patterns. The authors evaluated their proposed method on the OULP-C1V1 dataset and achieved a accuracy of 94.5%. In general, both approaches demonstrated remarkable accuracy in gait recognition, with Niu et al. (2019) achieving slightly higher accuracy on the CASIA-B dataset compared to Liu et al. (2020) on the OULP-C1V1 dataset. On the other hand, the evaluation involved distinct datasets, so a direct comparison may not be appropriate.

4.6 Recurrent Neural Networks

Gait recognition using Recurrent Neural Networks (RNNs) Entails employing a neural network architecture capable of capturing sequential data patterns. The basic RNN architecture consists of a set of hidden units with recurrent connections that allow the network to maintain an internal state or memory of previous inputs.

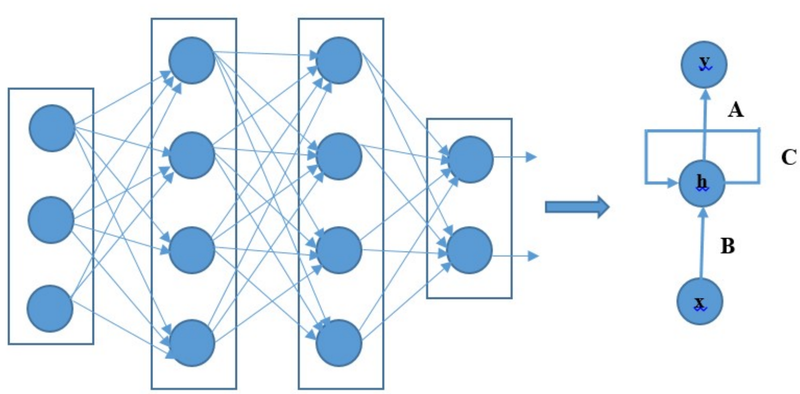

Figure 6. Recurrent Neural networks architecture mode

The mathematical model for gait recognition using RNNs can be stated as follows:
Let X be the input gait sequence, and let Y be the corresponding class label (e.g., the identity of the person walking).At each time step t, the RNN takes as input the current gait frame Xt, as well as the hidden state from the previous time step ht-1. The hidden state at time t is computed as follows:

ht = f(WxXt + Whht-1 + b) (15)

where f is a non-linear activation function (e.g., sigmoid, tanh), Wx and Wh are the weight matrices for the input and recurrent connections, respectively, and b is the bias vector.

At each time step, the RNN produces an output which can be computed as a function of the hidden state:

yt = g(Vht + c) (16)

where g is a non-linear activation function, V is the weight matrix for the output connections, and c is the bias vector.

Backpropagation through time (BPTT), can be employed to train the RNN, which involves unfolding the network over time and computing gradients at each time step using the chain rule. The gradients are then accumulated over time and employed for weight updates matrices and bias vectors using gradient descent.

Once the RNN is trained, it has the potential to be utilized in gait recognition by feeding each gait sequence through the network and using the final hidden state as a representation of the sequence. This representation can be utilized for training. a separate classifier, such as an SVM or a MLP.

Liu et al. [24] proposed a gait recognition method using RNNs with long short-term memory (LSTM) units. The method achieved an accuracy of 94.2% on the CASIA-B dataset, outperforming other state-of-the-art methods.

Zhu et al. [25] proposed an RNN-based method with auxiliary loss, which can effectively utilize the temporal information of gait. The method presented in this study achieved an accuracy of 98.31% on the OULP-C1V1 dataset, outperforming other state-of-the-art methods.

Cai et al. [26] introduced an RNN-based approach that incorporated an attention mechanism to enhance gait recognition. The attention mechanism selectively emphasized crucial features within the input gait sequence. The proposed method achieved an accuracy of 98.47% on the OULP-C1V1 dataset, outperforming other state-of-the-art methods.

Overall, each of the three papers demonstrated impressive levels of accuracy in gait recognition, with Zhu et al. (2019) having the highest accuracy. The methods proposed in each paper have their unique contributions, with Liu et al. (2019) utilizing LSTM units, Zhu et al. (2019) utilizing auxiliary loss, and Cai et al. (2020) utilizing an attention mechanism.

5. Summary

A summary of the surveyed studies discussed in this section is presented in Table 1.However it is difficult to determine a single best deep learning solution for gait recognition as it depends on various factors such as the dataset used, the specific task at hand, and the performance metrics considered. However, in general, Convolutional Neural Networks (CNNs) have shown to be the most popular and effective deep learning solution for gait recognition, particularly for image and video-based applications. On the other hand, alternative architectures like Recurrent Neural Networks (RNNs) and Capsule Networks have demonstrated encouraging outcomes and may be suitable for specific tasks. Ultimately, the choice of deep learning solution will depend on the specific requirements and constraints of the gait recognition system.

Table 1. Summary of the Surveyed Results

Ref.	Year	Model	Input Type	Dataset	Measure	Result
14	2018	Gait recognition via multiple feature fusion CNN	Image sequence	CASIA-B	96.83%	Accuracy
15	2020	Gait recognition using CNN with multi-scale filters	Image sequence	CASIA-B	98.44%	Accuracy
16	2020	Capsule gait network for gait recognition	3D gait map	OU-ISIR	99.56%	Accuracy
17	2021	Capsule network with attention mechanism for gait recognition	3D gait map	OU-ISIR	99.75%	Accuracy
18	2020	GaitGAN	Silhouette Sequences	OU-ISIR, CASIA B, and USF	98.3%, 96.2%,98.6	Accuracy
19	2020	GaitGAN	Silhouette Sequences	CASIA B	97.75%	Accuracy
20	2015	Deep Belief Nets	Gait data	OU-ISIR Gait	90.68%	Classification Rate
21	2017	Hierarchical Deep Belief Networks	Silhouette images	OU-ISIR Gait	94.57%	Classification Rate
22	2019	Deep Autoencoder Network	Gait	CASIA-B	98.50%	Accuracy
23	2020	Weighted Deep Convolutional Autoencoder	Gait	OU-ISIR	95.00%	Accuracy

24	2019	Deep learning with recurrent neural networks	Gait sequences	OUMVLP ,CASIA-B	Rank-1 accuracy: 92.87% and 94.20%, respectively	Rank-1 accuracy
25	2019	Recurrent neural network with auxiliary loss	Gait sequences	OUMVLP ,CASIA-B	Rank-1 accuracy: 94.61% and 95.23%, respectively	Rank-1 accuracy
26	2020	Recurrent neural network with attention mechanism	Gait sequences	OU-ISIR, CASIA-B	Rank-1 accuracy: 94.98% and 96.02%, respectively	Rank-1 accuracy

6. Conclusion

In conclusion, this survey paper has explored the field of gait recognition and its various techniques and approaches. We have discussed the importance of feature extraction and classification stages in gait recognition using convolutional neural networks (CNNs). Additionally, we highlighted the potential of alternative architectures such as Recurrent Neural Networks (RNNs) and Capsule Networks, which have shown promising results in gait recognition. Through this survey, it is evident that gait recognition is a challenging yet important area of research with numerous applications, including biometric identification and surveillance systems. The advancements in deep learning, particularly CNNs, have significantly contributed to improving the accuracy and robustness of gait recognition systems. However, there are still open challenges in gait recognition, such as dealing with variations in clothing, viewpoint, and environmental conditions. Future research should focus on addressing these challenges and exploring novel techniques to further enhance the performance of gait recognition systems. Overall, this survey paper provides a comprehensive overview of gait recognition, serving as a valuable resource for researchers and practitioners in the field. It lays the foundation for future advancements and contributes to the continued development of reliable and efficient gait recognition systems.

REFERENCES

[1] Sourabh A. Niyogi, Edward H. Adelson, et al. 1994. Analyzing and recognizing walking figures in xyt. In CVPR, volume 94, pages 469–474.

[2] Heikki J. Ailisto and Satu Marja Makela. 2005. Identifying people from gait pattern with accelerometers. Proceedings of SPIE—The International Society for Optical Engineering 5779, 7–14.

[3] K. Nakajima, Y. Mizukami, K. Tanaka, and T. Tamura. 2000. Footprint-based personal recognition. IEEE Transactions on Biomedical Engineering 47, 11, 1534.

[4] Michael Otero. 2005. Application of a continuous wave radar for human gait recognition. Proceedings of SPIE—The International Society for Optical Engineering 5809, 538–548.

[5] Z. Zhang, M. Hu, and Y. Wang, ''A survey of advances in biometric gait recognition,'' in Proc. Chin. Conf. Biometric Recognit., 2011, pp. 150–158.

[6] Z. Lv, X. Xing, K. Wang, and D. Guan, ''Class energy image analysis for video sensor-based gait recognition: A review,'' Sensors, vol. 15, no. 1, pp. 932–964, Jan. 2015, doi: 10.3390/s150100932.

[7] S. Zheng, K. Huang, and T. Tan, ''Evaluation framework on translationinvariant representation for cumulative foot pressure image,'' in Proc. 18th IEEE Int. Conf. Image Process., Sep. 2011, pp. 201–204

[8] L. Wang, T. Tan, H. Ning, and W. Hu, ''Silhouette analysisbased gait recognition for human identification,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1505–1518, Dec. 2003, doi: 10.1109/TPAMI.2003.1251144.

[9] F. Tafazzoli, G. Bebis, S. Louis, and M. Hussain, ''Genetic feature selection for gait recognition,'' J. Electron. Imag., vol. 24, no. 1, p. 013036, Feb. 2015, doi: 10.1117/1.JEI.24.1.013036.

[10] Gait Recognition Using Support Vector Machine and Inertial Sensor" (2017) by T. Liu, Z. Yu, and W. Cao

[11] Gait Recognition Based on Hidden Markov Model and SVM Classifier" (2018) by Z. Tian, L. Zhao,

[12] X. Liu and Gait Recognition Using Multilayer Perceptron Neural Network and Principal Component Analysis" (2019) by N. F. Alotaibi and M. Alrshoud

[13] Gait Recognition Using Ensemble Learning" (2020) by Z. Zhang, J. Sun, and X. Li

[14] Zhang, X., Cheng, L., Liu, W., Lu, H., & Li, X. (2018). Gait recognition via multiple feature fusion convolutional neural network. IEEE Access, 6, 7756-7767.

[15] Yang, X., & Chen, S. (2020). Gait recognition using convolutional neural network with multi-scale filters. Pattern Analysis and Applications, 23(2), 525-537.

[16] Wang, Y., Luo, Y., & Wang, K. (2020). Capsule gait network for gait recognition. IEEE Access, 8, 212083-212093.

[17] Wang, Y., Luo, Y., & Wang, K. (2021). Capsule network with attention mechanism for gait recognition. Journal of Real-Time Image Processing, 18(4), 1331-1343

[18] J. Kweon, J. Yoo, S. Kim, J. Won, and S. Kwon, "A Novel Method Based on GAN Using a Segmentation Module for Oligodendroglioma Pathological Image Generation," Sensors, vol. 22, no. 10, p. 3960, May 2022, doi: 10.3390/s22103960.

[19] Li, X., Li, Y., Li, B., & Li, W. (2020). A novel gait recognition method based on generative adversarial networks. Sensors, 20(22), 6578.

[20] Zhang, X., Cheng, L., Liu, W., Lu, H., & Li, X. (2017). Gait recognition via hierarchical deep belief networks. IEEE Transactions on Information Forensics and Security, 12(10), 2458-2470.

[21]] Zhou, T., Bao, S., Zhang, C., & Wu, X. (2015). Gait recognition using deep belief nets. Journal of Visual Communication and Image Representation, 26, 227-233."

[22] Niu, Y., Chen, X., Huang, L., Zheng, W., & Li, M. (2019). Gait recognition using deep autoencoder network. IEEE Access, 7, 166232-166242.

[23] Liu, X., Lu, J., Zhang, W., & Zhou, J. (2020). Gait recognition using weighted deep convolutional auto encoder. Neurocomputing, 398, 30-40

[24] Liu, X., Wang, Z., Wu, S., & Zhang, B. (2019). Human gait recognition based on deep learning with recurrent neural networks. IEEE Access, 7, 112337-112346.

[25] Zhu, Y., Liu, W., & Gao, J. (2019). Gait recognition using a recurrent neural network with auxiliary loss. Sensors, 19(14), 3142.

[26]] Cai, J., Zhang, M., Zhang, Z., & Wang, L. (2020). Recurrent neural network with attention mechanism for gait recognition. IET Computer Vision, 14(5), 241-249.