Gait Recognition Using Hybrid LSTM-CNN Deep Neural Networks

Entesar T. Burges[1,*], Zakariya A. Oraibi [2], and Ali Wali [3]

[1] National School of Electronics, University of Sfax, Sfax 3029, Tunisia
[2] Department of Computer Science, College of Education for Pure Sciences, University of Basrah, Basrah, Iraq
[3] Research Groups in Intelligent Machines, National Engineering School, University of Sfax, Sfax 3029, Tunisia
Email: entesar.barges@gmail.com (E.T.B.), zakaria_au@uobasrah.edu.iq (Z.A.O.), ali.wali@isims.usf.tn (A.W.)
*Corresponding author

***Abstract*—Identifying individuals based on their gait is a crucial aspect of biometric authentication. It is complicated by several factors, such as altering one's walking posture, donning a coat, and wearing high heels. With the advent of artificial intelligence, deep learning, in particular, has made significant strides in this area. The conditional Generative Adversarial Network (cGAN), together with hybrid Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs), are used in this research to create images using a novel technique. The framework comprises three parts. The first involves extracting silhouettes, necessitating computing the gait cycle and energy. The technique of creating images using discriminator models and cGANs is the second part. Image classification using hybrid LSTM and CNN networks is the third step. Experiments were conducted to assess our approach using the CASIA database, a publicly available gait recognition dataset. Our proposed approach achieved a high classification accuracy of 97.11%. Our results outperform state-of-the-art techniques, especially when it comes to carrying bags and donning coats.**

***Keywords*—gait recognition, Generative Adversarial Networks (GANs), Long Short-Term Memory (LSTM), deep learning**

I. Introduction

Biometrics is the science of using physical characteristics, such as fingerprints, faces, and gait, to identify people. It is a form of identification that uses unique biological traits to verify a person's identity. Biometric authentication is used in many applications, including access control, computer security, and banking. These systems are becoming increasingly popular as they provide a secure and convenient way to identify individuals. Gait is the pattern of movement of the limbs of a person or animal during locomotion. Specific limb movement sequences characterize it and describe how a person or animal generally walks, runs, or moves [1].

Gait recognition is a biometric technology that identifies individuals using their unique walking patterns. It is a non-invasive and cost-effective way to authenticate people. The current gait analysis techniques are divided into two parts. The first is known as model-based [2], and the second is known as appearance-based [3]. Model-based gait analysis techniques use mathematical models to analyze the motion of a person's body. These models measure a person's gait, speed, acceleration, and other parameters. This type of analysis is often used in clinical settings to diagnose and treat gait abnormalities. Appearance-based gait analysis techniques use computer vision algorithms to analyze the visual appearance of a person's gait. This type of analysis is used to identify unique characteristics in a person's walking pattern, such as stride length, step width, and foot placement. In addition, it can be used for applications such as biometric identification and motion capture for animation. Deep learning is a subset of machine learning that uses algorithms inspired by the structure and function of the brain's neural networks. It is a form of Artificial Intelligence (AI) that enables machines to learn from large amounts of data, identify patterns, and make decisions with minimal human intervention. Deep learning models are used in a variety of applications, such as computer vision, natural language processing, speech recognition, and robotics [4, 5].

In contrast to machine learning, which commonly utilizes shallow architectures, deep learning closely mirrors the organizational structure of the human brain by adopting a deep architecture [6]. The information goes through several modifications because of these deep structures before it is finally displayed. The input is routed through several simulated neural network layers to attain greater precision. In this paper, we proposed a method to use an innovative technique based on the use of conditional Generative Adversarial Networks (cGANs) and a deep learning network (hybrid Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs)) to produce a more accurate system in determining the identity of the individual by his gait.

The rest of the paper is arranged into the following sections. Section II discusses the related work. Section III introduces the gait recognition pipeline used in this paper and the dataset used in the experiments. Section IV shows the experiments, followed by Section V, which reveals the results and discussion of the experiments. Finally, Section VI concludes the paper.

Manuscript received November 6, 2023; revised January 15, 2024; accepted January 26, 2024; published June 6, 2024.

II. Literature Review

There are three fundamental steps in the gait recognition system. Image capture is the first phase, followed by initial processing, in which binary silhouettes are extracted from still or moving images; characterizing the silhouettes is the second step, and training or discrimination is the final step [7–10]. The earliest attempts at modeling the human body as a whole were made in the 1990s [11]. This technique is referred to as the "model-based method," which may be summed up as an effort to depict all the significant body points, such as the length and width of the body, by locating the pelvis and knees along with the different body joints from bilateral silhouettes. In Ref. [12], the researchers took 22 significant spots from the human body and portrayed them using a deformable layer. For binary silhouettes, these characteristics can be changed over time. Although this technique has produced useful results for identifying human gait, it still has drawbacks, including difficulties collecting silhouette images from distant regions and challenges with shadows and light. Researchers have tried several techniques to increase accuracy, including creating two or three-dimensional models [13]. The sharpness of the image is not a factor in the model-free approach, another method of differentiation, because the images taken are from distant security cameras. The model-free approach is further broken down into sequential motion-based and spatiotemporal motion-based approaches. While sequential motion approaches depict gait as a time sequence of human positions, the spatiotemporal approach portrays gait by mapping the distribution of motion through space and time [14]. The sequential motion-based method presented in [15] entails capturing the history of these motions as well as displaying the motion through temporal templates that show where the motion has happened. The pre-processing, feature extraction, and classification methods applied to silhouette-based gait sequences are principally responsible for the discrepancies between the spatiotemporal approaches. The Gait Energy Image (GEI), a feature selection strategy that captures a history of gait movements in a single 2D template rather than storing them as a collection of templates, was proposed by Hu *et al.* [16]. The spatiotemporal GEI characteristic is calculated by averaging the pixels of the silhouette over many frames during a gait cycle. The statistical integration of both natural and artificially (distorted) gait templates was necessary for the recognition process. This technique reports excellent recognition performance while also saving space. A gait entropy image that will be used as an automatic feature selection process is shown for the gallery (ground truth) and probing (testing) photos [17].

The adaptive component and discriminant analysis, a fast recognition method, has been shown to lessen the effects of covariate walking. Various academics have conducted research and proposed different tactics to cope with recognition from distinct points of view in an identical circumstance. In Ref. [18], the authors used the silhouette sequence to create a deep learning algorithm based on ResNet and LSTM. Recently, researchers have used Generative Adversarial Networks (GANs) extensively to recognize the gaits on very large scales of data [19] with a two-stream GAN model. In the same context, Dupuis *et al.* [20] developed an architecture based on LSTM and autoencoder networks that employed RGB image sequences as input. With silhouette-based data captured by certain cameras, Alvarez and Sahonero-Alvarez [21] proposed a Convolutional Neural Network (CNN) model for predicting the angle and also used it to detect the gait. The prediction results of several gait detection methods are convincing, yet there is still an opportunity to improve their effectiveness further.

III. Materials and Methods

In this study, our primary goal was to find a solution to the issue of vision variations brought on by altered walking angles, the wearing of a coat, wearing high heels, or carrying a bag. Numerous earlier studies [22–25] suggested that altering any of the previously listed parameters results in a change in human gait. To lessen the impact of the problem of various visions, an approach relying on the employment of the cGAN in addition to the hybrid LSTM and CNN networks

to create images was presented. The side view angle was used because it reveals a wide range of qualities. The framework can be divided into three parts. The first part is the process of extracting silhouettes, calculating the gait cycle, and then calculating gait energy images. The second part generates images through Generative Adversarial Networks (GANs) and discriminators' models, and the third part classifies images using two networks. The first network is Patch GAN, and the second is a hybrid CNN and LSTM network. Fig. 1 shows the proposed algorithm.

A. Preprocessing

Obtaining silhouettes during a single walking cycle is the first stage in this technique. The approach described in [25] generates human silhouettes from the provided gait sequence. Size normalization and horizontal alignment are applied to all images. Noise is removed from the images using dilation and erosion. After that, the gait cycle segmentation is estimated by measuring the silhouette's bounding box's length and width and then calculating the interval between the two highest lengths. After that, GEI is calculated by using the samples in Fig. 1 and Eq. (1):

$$G(x,y) = \sum_{t=1}^{N} I(x,y,t)/N \quad (1)$$

where X and Y are the image coordinates, N is the number of images in a whole gait cycle, I is the image, and t is the gait cycle frame number.

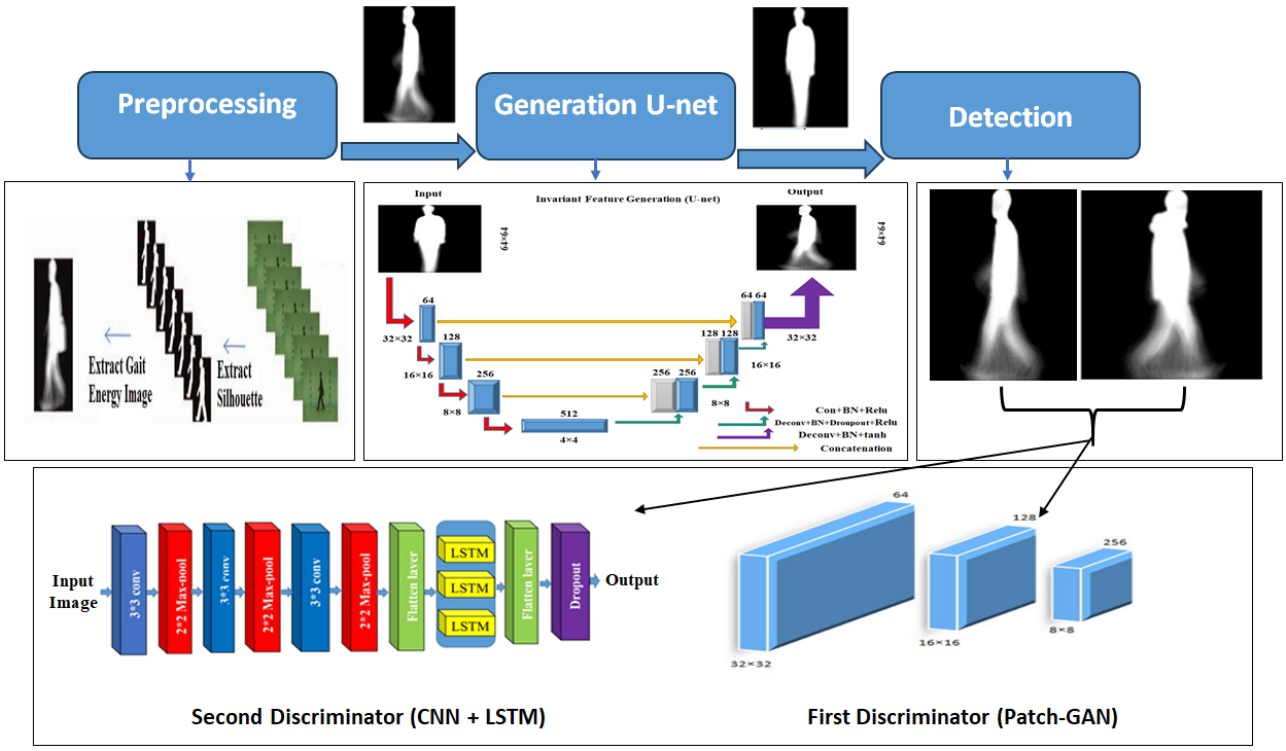


Fig. 1. The proposed pipeline for gait recognition.

B. *Invariant Feature Generation*

We propose a conditional generative model to convert gait representations from any perspective and appearance condition to representations at side view under typical conditions using an architecture based on a U-Net.

1) *Input data*

Before the GAN can be applied, data must first be organized. As a result, the GEIs from all views in the regular walking, carrying a bag, and wearing a coat sequences are set as the source information. On the other hand, the GEIs from normal walking at 90° (side view) are set as the goal data. Then, 40M source-target representation pairings were gathered to train the GAN.

2) *Conditional generative adversarial nets*

Mirza and Osindero proposed the cGANs in [26]. Because attributes are not explicitly provided, simple GAN cannot be controlled. Thus, they introduced the conditional GAN version. When using external data, cGAN includes a condition y that decides whether an image is being produced or a distinction is being made. The class label or properties from several distinct modalities may be associated with these conditions. Conditioning is performed by adding extra data y as an additional input layer for the generator and discriminator.

The cost function for cGAN is presented in [26] and is nearly identical to that for GAN. The only distinction between discriminator and generator networks is the addition of a condition y. Our conditional GAN's objective can be summed up as follows:

$$(G,D) = E_{x,y}[log\ D(x,y)] + E_{x,y}[[log\ (1 - D(x,G(x,z)))]] \quad (2)$$

In which the generator G tries to minimize this function. In contrast, the discriminator D aims to maximize it. Past studies have also demonstrated that combining the past loss with more traditional loss functions aids in obtaining results that are quite close to the truth.

$$L_{L1}(G) = E_{x,z}[\| y - G(x,z) \|_1] \quad (3)$$

The final objective can be defined as:

$$G* = arg\ \min_G \max_D L_{CGAN}(G,D) + \lambda L_{L1}(G) \quad (4)$$

where λ is the regularizing hyperparameter. For example, when using λ, the cGAN generates high defined outputs, but the classification accuracy decreases.

3) *Classification*

A subject is classified to determine whether or not it belongs to a class in the database. A hybrid LSTM-CNN deep neural network was adopted to eliminate the weaknesses of traditional classification methods, as the discriminative information is not within the means of classes and a small sample size problem. The proposed image classification model is a layered deep neural network consisting of a CNN and an LSTM. The LSTM is a type of Recurrent Neural Network (RNN) that, in contrast to conventional feed-forward neural networks, has feedback connections. The ability to have feedback connections makes LSTM a type of "general purpose computer", allowing it to perform all computations that a Turing machine can.

a) *Long Short-Term Memory (LSTM)*

According to Fig. 2, a unit of an LSTM is defined as a group of vectors consisting of a forget gate f_t, an input

gate i_t, a memory cell c_t, an output gate o_t, and a hidden state h_t at each time step, where d is the magnitude of the memory dimension [24, 25]. The LSTM equations have numbers ranging from 5–10. The terms *b* and *W* in the equations represent the input gate, output gate, forget gate, memory cell, tanh layer, a hidden layer bias vector, and weight matrices, respectively. σ Indicates the logistic sigmoid function.

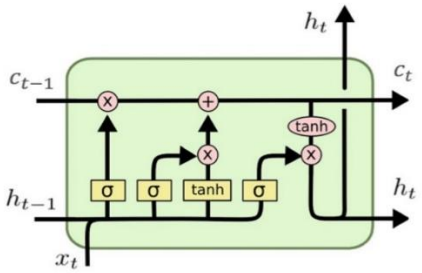


Fig. 2. LSTM cell.

In an LSTM unit, the input gate is in charge of deciding whether to remember the data it processes and is fed with a fresh stream of information at every time step t. On the other hand, the amount of information that should be erased from the memory cell is controlled by the forget gate.

b) The CNN-LSTM neural network

Fig. 1 represents the architecture of our proposed CNN-LSTM model. Our system uses grayscale images as input images of 100×100 size. The network consists of three convolutional layers. A max pooling layer follows each layer, which is composed of three LSTM layers, followed by a fully connected layer and a dropout layer. The details of the network structures are shown in Table I.

TABLE I. DETAILS OF THE CNN AND LSTM NETWORKS

Layers	Number of Filters	Filter Size	Windows Size	Stride	Padding	Activation Function
Conv.1	32	3×3	N	N	Same	Relu
MaxPool	N	N	2×2	2	N	N
Conv.2	16	3×3	N	N	Same	Relu
MaxPooll	N	N	2×2	2	N	N
Conv.3	32	3×3	N	N	Same	Relu
MaxPool	N	N	2×2	2	N	N
flatten	N	N	N	N	N	N
LSTM	32	N	N	N	N	N
LSTM	64	N	N	N	N	N
LSTM	128	N	N	N	N	N
Flatten	N	N	N	N	N	Softmax

IV. EXPERIMENTS AND ANALYSIS

This CASIA-B database [2] was captured indoors when the subject was walking, with 11 cameras positioned around the person's left side. Eighteen degrees separated the two closest view directions. Fig. 3(a) shows the images captured by the cameras. The viewing angles are named 0°, 18°, 36°, 54°, 72°, 90°, 108°, 126°, 144°, 162°, and 180° from left to right. Gait data from 124 subjects were captured, among whom 93 were men, and 31 were women. Every subject was asked to walk 10 times in the scene (6 normal + 2 with a bag + 2 with a coat). Example frames are shown in Fig. 3(b).

Thus, there were a total of 10×11×124 = 13,640 video sequences in the database [31]. Fig. 4 shows the GEI from one subject in all the conditions at 11 views.

We applied the experimental approach recommended in [18, 19] in order to accurately compare the proposed strategy with cutting-edge methods. As a result, we divided the dataset in half.

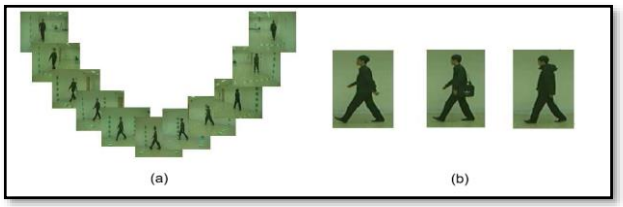


Fig. 3. Sample frames of CASIA database. (a) 11 different capturing views. (b) 3 different walking conditions.

The first 62 participants, composed of six regular, two carrying-bag, and two wearing-coat sequences, made up the training set. The remaining 62 people were employed in the test phase. In order to evaluate the variations in view, carrying, and clothing circumstances, the first four normal sequences, denoted as "*nm1*", were used as the gallery set, but the two left sequences, along with the "*bg*" and "*cl*" sequences, were used as the proving set. This is displayed in Table II. The generator and two discriminators make up the cGAN's two components, as depicted in Fig. 1. The GEI generation process makes use of the generator.

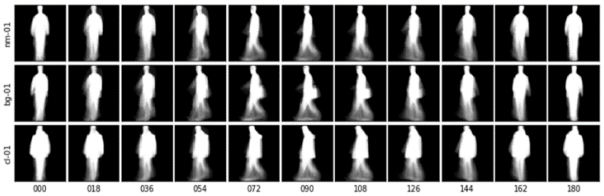


Fig. 4. Walking sequences at all the conditions at the 11 views of the CASIA-B dataset.

TABLE II. THE EXPERIMENTAL DESIGN

Training set	Gallery set	Probe set			
		Probe NM	Probe BG	Probe CL	Probe ALL
ID: 001–062 nm01–nm06, bg01, bg02, cl01, cl02	ID: 063–124 nm01–nm04	ID: 063–124 nm05, nm04	ID: 063–124 bg01, bg02	ID: 063–124 cl01, cl02	ID: 063–124 nm05, nm04 bg01, bg02 cl01, cl02

It utilizes the U-Net architecture. We employed a similar setup to that of Isola [19]. It is composed of two elements, the first of which is the encoder and the second is the decoder. There are four convolutional layers in the encoder. Because we use a U-Net design, which concatenates activations from layer *i* to layer *n-i*, the

number of channels in the decoder doubled. The encoder's first layer differs from the others because it does not use batch normalization. After the final decoder layer, a convolution and *tanh* function are added to match the channel number of the output. Table III presents specifics regarding the encoder and decoder structures. To distinguish whether created photos are fake or real, the first discriminator shown is utilized as in Fig. 1, which comprises three convolutional layers. In addition, we decided to use a marginally unique design known as Patch GAN [20]. The benefit of using this method is that it only penalizes specific areas of the images, i.e., it identifies whether a specific area of an image is real or fake. As a result, we can better focus on particular GEI regions that correlate to the parts more resistant to changes in appearance, such as the head and feet, which contain the most important features [21, 22]. The identification discriminator, which uses a hybrid LSTM-CNN architecture, is the second discriminator. The identification discriminator uses the initial gait image sequence and the generated output gait image sequence as one training data pair and then computes the likelihood that the data pair belongs to the same person. When the training data pair represents a single subject, the identification discriminator should output 1; otherwise, it should output 0.

TABLE III. DETAILS OF THE ENCODER AND THE DECODER

phases	Layers	Number of Filters	Filter Size	Stride	Batch Norm	Dropout	Concatenation	Activation Function
Encoder	Conv.1	64	32×32	2	N	N	Y	L-Relu
	Conv.2	128	16×16	2	Y	N	Y	L-Relu
	Conv.3	256	8×8	2	Y	N	Y	L-Relu
	Conv.4	512	4×4	2	Y	N	N	L-Relu
Decoder	Conv.1	256	8×8	2	Y	Y	N	Relu
	Conv.2	128	16×16	2	Y	N	N	Relu
	Conv.3	64	32×32	2	Y	N	N	Relu

The generator and the first discriminator were trained using the Adam optimizer, with a learning rate of 0.0002 and momentum parameters of $\beta_1 = 0.5$ and $\beta_2 = 0.999$. We discovered that after 20 training epochs, satisfactory performance was reached when using $\lambda = 100$. The second discriminator was trained using root mean square prop optimization with a learning rate of 0.001 and a binary cross entropy function. The weights were initially obtained from a Gaussian distribution with a mean of 0 and a standard deviation of 0.02 because U-Net and Path GAN were trained from scratch (the CASIA database only has a limited number of sequences for each subject). In U-Net and Patch GAN, the image size is 64×64, whereas in LSTM, it is 150×150×3 reshaped to 100×100×3 to fit the LSTM input shape layer. We utilized Python programming language in the Collaborator Pro Environment and the Windows 11 operating system.

V. RESULTS AND DISCUSSION

In this paper, a method of hybrid CNN and LSTM classifiers on the spatiotemporal feature GEI generated using CGAN has been proposed for gait recognition systems. The CASIA-B [2] dataset has been used to assess the proposed method statistically. There are 124 people in the CASIA-B [2] dataset. The 124 individuals in the dataset were split into 62 individuals for the training set and 62 individuals for the testing set. The performance measure to evaluate the proposed model includes accuracy, score, precision, and recall [26]. Please see Table IV.

TABLE IV. ACCURACY, PRECISION, RECALL, AND F1-SCORE RESULTS

Class	Accuracy	Precision	Recall	F1 Score
Training Data	0.9700	0.9700	0.9700	0.9700
Val. Data	0.9700	0.9700	0.9700	0.9700
Testing Data	0.9700	0.9700	0.9700	0.9700

First, an image segmentation process was conducted, employing the Gaussian mixture technique to separate the silhouettes from RGB gait images due to this method's efficiency for fundus gait image segmentation, as shown in prior studies [25]. Some of the segmentation results are shown in Fig. 1. After this processing, size normalization and horizontal alignment are applied to each image, followed by removing all noise from the images using dilation and erosion. The estimation of the gait cycle segmentation that follows was done by first determining the length and width of the bounding box that is drawn around the silhouette and then by determining the interval between the two largest lengths. The samples in Fig. 1 are then used to calculate the GEI. To generate invariant features, GEI were created using cGAN for different conditions. This produced typical side view images for multiple views, wearing a coat and carrying a bag GEI, as shown in Fig. 5. After that, these views are discriminated using the first discriminator for the first 62 people. The test set was differentiated using a second discriminator, indicating whether the generated GEI belonged to the same person. Tables V–VII show the CCR that our model was able to attain. In the aforementioned tables, each column denotes a view angle from the probe set, whereas each row denotes a view angle from the gallery set. Because there are 11 views in the database, there are 121 possible combinations.

The accuracy of the hybrid CNN and LSTM classifier has been assessed based on previously discussed changing neural network parameters. The resulting accuracy, as determined by the testing dataset, is 97%. Fig. 6 shows the training details after 25 epochs. Based on the overall accuracy of the data obtained, as demonstrated in Table VIII, the performance of our suggested technique has been superior in comparison with existing methods. As can be observed from the comparative analysis in the table, our study is performing well when compared to the other studies.

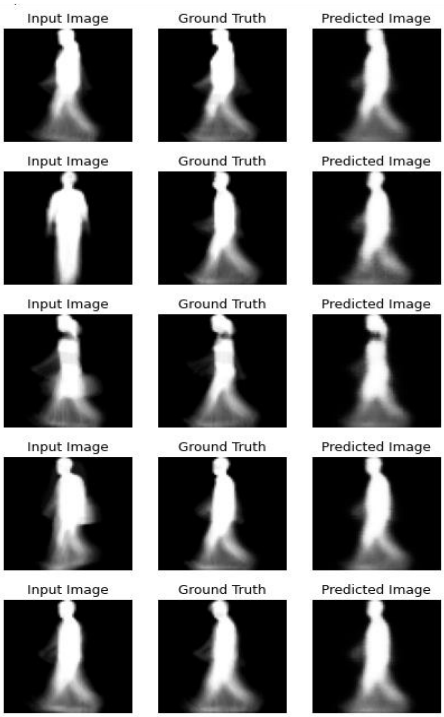


Fig. 5. Side views images for multiple views, wearing a coat and carrying a bag GEI.

TABLE V. CORRECT CLASSIFICATION RATE FOR NORMAL WALKING

		Probe Set View (bg01, bg02)										
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
Gallery Set View	0°	45.16	27.42	21.77	14.52	6.45	9.68	6.45	10.48	20.97	25.00	30.65
	18°	28.23	49.19	45.16	28.23	20.97	19.35	17.74	24.19	29.84	30.65	21.77
	36°	24.19	41.13	58.06	49.19	31.45	23.39	21.77	27.42	35.48	22.58	13.71
	54°	12.10	20.97	45.16	58.87	52.42	41.13	33.06	28.23	25.81	16.13	12.10
	72°	20.97	22.58	35.48	45.97	61.29	53.23	43.55	31.45	25.81	16.94	8.87
	90°	14.52	19.35	31.45	41.13	58.06	50.00	47.58	44.35	25.81	21.77	9.68
	108°	15.32	18.55	37.10	41.94	58.06	51.61	59.68	53.23	43.55	28.23	15.32
	126°	19.35	21.77	32.26	40.32	38.71	40.32	42.74	54.84	45.97	24.19	12.90
	144°	23.39	22.58	32.26	32.26	31.45	29.84	39.52	50.81	58.87	40.32	19.35
	162°	18.55	16.13	19.35	22.58	19.35	12.10	16.94	23.39	29.84	41.94	19.35
	180°	29.03	16.13	11.29	8.87	8.06	7.26	4.84	8.06	12.10	22.58	38.71

TABLE VI. CORRECT CLASSIFICATION RATE FOR CARRYING WALKING

		Probe Set View (cl01, cl02)										
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
Gallery Set View	0°	99.19	67.74	46.77	29.03	23.39	18.55	18.55	28.23	30.65	47.58	68.55
	18°	80.65	99.19	93.55	62.90	48.39	35.48	33.06	43.55	45.16	66.13	50.81
	36°	48.39	91.94	96.77	86.29	69.35	54.84	58.06	62.10	66.94	57.26	28.23
	54°	34.68	59.68	91.13	97.58	92.74	89.52	83.06	82.26	66.13	39.52	24.19
	72°	16.94	35.48	66.13	93.55	99.19	97.58	93.55	77.42	55.65	30.65	12.90
	90°	20.16	37.90	54.03	77.42	98.39	99.19	97.58	83.06	58.06	32.26	20.16
	108°	24.19	39.52	55.65	81.45	93.55	95.97	99.19	92.74	82.26	39.52	26.61
	126°	22.58	45.97	59.68	75.00	81.45	83.87	93.55	97.58	95.16	56.45	25.00
	144°	32.26	45.97	60.48	61.29	62.90	59.68	82.26	94.35	98.39	77.42	41.13
	162°	63.71	61.29	62.10	50.81	34.68	35.48	48.39	66.13	77.42	99.19	71.77
	180°	99.19	67.74	46.77	29.03	23.39	18.55	18.55	28.23	30.65	47.58	68.55

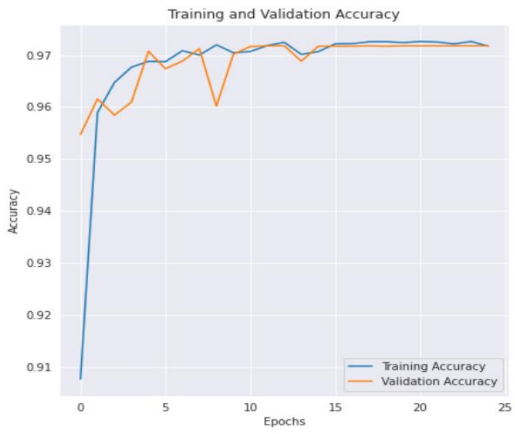
TABLE VII. CORRECT CLASSIFICATION RATE FOR CLOTHING WALKING

		Probe Set View (nm05, nm06)									
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
Gallery Set View	0°	99.19	67.74	46.77	29.03	23.39	18.55	18.55	28.23	30.65	47.58	68.55
	18°	80.65	99.19	93.55	62.90	48.39	35.48	33.06	43.55	45.16	66.13	50.81
	36°	48.39	91.94	96.77	86.29	69.35	54.84	58.06	62.10	66.94	57.26	28.23
	54°	34.68	59.68	91.13	97.58	92.74	89.52	83.06	82.26	66.13	39.52	24.19
	72°	16.94	35.48	66.13	93.55	99.19	97.58	93.55	77.42	55.65	30.65	12.90
	90°	20.16	37.90	54.03	77.42	98.39	99.19	97.58	83.06	58.06	32.26	20.16
	108°	24.19	39.52	55.65	81.45	93.55	95.97	99.19	92.74	82.26	39.52	26.61
	126°	22.58	45.97	59.68	75.00	81.45	83.87	93.55	97.58	95.16	56.45	25.00
	144°	32.26	45.97	60.48	61.29	62.90	59.68	82.26	94.35	98.39	77.42	41.13
	162°	63.71	61.29	62.10	50.81	34.68	35.48	48.39	66.13	77.42	99.19	71.77
	180°	79.03	47.58	28.23	21.77	16.94	15.32	20.97	22.58	41.13	68.55	99.19

TABLE VIII. COMPARISON WITH STATE-OF-THE-ART METHODS

Authors	Proposed Method	Accuracy	Database
Wang *et al.* [27]	Ensemble Learning	0.92%	CASIA-B
Wang *et al.* [27]	LSTM	0.95%	CASIA-B
Amin *et al.* [28]	Conv-BiLSTM	0.96%	CASIA-B
Proposed	cGAN + CNN + LSTM	**0.9711%**	CASIA-B

(a)



(b)

Fig. 6. The accuracy on CASIA-B experiments, (a) Legends are training loss vs validation loss; (b) Legends represent training accuracy vs validation accuracy.

On the CASIA-A and CASIA-B datasets, Wang *et al.*'s [27] ensemble learning approach for classifying human gait produced values of 0.95 and 0.92 CPR, respectively. On the CASIA-B dataset, the same technique achieved 0.95 CPR by using the LSTM model on the CASIA-B dataset to learn the sequential patterns of the input images. Amin *et al.* [28] achieved 0.96 CPR (0.88, human with bag and 0.92, normal) using the CNN-BiLSTM model for the classification of different types of humans.

Due to the adoption of the identical databases and data division, a comparison was made with the earlier works listed in Table VIII. It was demonstrated that creating images using CGAN was highly successful and effective, making it a superb method for differentiating the human gait.

VI. CONCLUSION

This work developed a technique for a gait identification system based on cGAN using both U-Net architecture and hybrid CNN and LSTM classifiers to overcome appearance fluctuations owing to changes in clothes, carrying conditions, and view angle. Due to the specificity of the data, it is difficult to distinguish the human gait, so we proposed an algorithm consisting of a generator that is used to generate standard images at a 90° angle. The first discriminators were proposed to distinguish between real and fake images produced by the generator. The second generator determines whether these images are for the same person.

The accuracy of gait recognition was improved by this design, as we have shown through the results. Performance is better using the suggested framework than in earlier proposals. In addition, our solution overcame the issues of previous studies that suffered from variations in carry bags and clothing outerwear. Because of this quality, our technology is suitable for advanced surveillance systems, among other practical applications. This model will be updated to handle more difficult circumstances, including temporal fluctuations, and show how benchmarking has improved when massive databases are used.

Additional subjects are necessary to produce more accurate results, and these additional subjects improve accuracy when there are major view changes between the prove set and the gallery. In addition, we will need to use more advanced and powerful models to handle cross-view identification. It is also noteworthy that the limitations of

current gait analysis techniques include inhibited environmental conditions and long processing duration. Future work will consider these limitations to improve the performance of the gait recognition system further.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Entesar T. Burges proposed the suggested pipeline for gait recognition. Entesar T. Burges and Zakariya A. Oraibi produced the written manuscript. Ali Wali and Entesar T. Burges approved the analytical methods. Entesar T. Burges and Zakariya A. Oraibi discussed the final results thoroughly. All authors had approved the final version.

REFERENCES

[1] L. Yao, W. Kusakunniran, Q. Wu, J. Zhang, and Z. Tang, "Robust cnn based gait verification and identification using skeleton gait energy image," in *Proc. Digital Image Computing: Techniques and Applications, Canberra*, Australia, December 2018.

[2] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle clothing and carrying condition on gait recognition," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 4, pp. 441–444, 2006.

[3] M. Balážia, "Human gait recognition based on body component trajectories," M.Sc. thesis, Faculty of Informatics, University of Masaryk, 2013.

[4] G. Ariyanto, "Model-based 3D gait biometrics," PhD thesis, Department of Electronics and Computer Science, Faculty of Physical and Applied Sciences, University of Southampton, March, 2013.

[5] K. Bashir, X. Tao, and G. Shaogang, "Feature selection on gait energy image for human identification," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 985–988.

[6] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.

[7] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *IEEE Trans. Syst., Man, Cybern.*, vol. 38, no. 2, pp. 342–352, Apr. 2008.

[8] J. Yoo, D. Hwang, K. Moon, and M. S. Nixon, "Automated human recognition by gait using neural network," in *Proc. Workshops on Image Processing Theory, Tools and Applications*, Sousse, Tunisia, November 2008.

[9] C. Yan, B. Zhang, and F. Coenen, "Multi-attributes gait identification by convolutional neural networks," in *Proc. International Congress on Image and Signal Processing*, Shenyang, China, October 2015.

[10] W. Kusakunniran, "Human gait recognition under changes of walking conditions," PhD thesis, University of New South Wales, June 2013.

[11] I. Rida, A. Somaya, and B. Ahmed, "Gait recognition based on modified phase-only correlation," *Signal Image Video Processing*, vol. 10, no. 3, pp. 463–470, 2016.

[12] T. Connie, M. K. O. Goh, and A. B. J. Teoh, "Human gait recognition using localized Grassmann mean representatives with partial least squares regression," *Multimedia Tools Applications*, vol 77, no. 21, pp. 28457–28482, 2018.

[13] H. Chao, Y. He, J. Zhang, J. Feng, "Gait set: Regarding gait as a set for cross-view gait recognition," in *Proc. the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, no. 1, pp. 8126–8133.

[14] C. Carley, E. Ristani, and C. Tomasi, "Person re-identification from gait using an autocorrelation network," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[15] Y. He, J. Zhang, H. Shan, and L. Wang, "Multi-task GANs for view-specific feature learning in gait recognition," in *Proc. IEEE Trans Information Forensics Security*, vol. 14, no. 1, pp. 102–113, 2018.

[16] B. Hu, Y. Gao, Y. Guan, Y. Long, N. Lane, and T. Ploetz, "Robust cross-view gait identification with evidence: A discriminant gait GAN (DiGGAN) approach on 10000 people," arXiv preprint, arXiv: 1811.10493, 2018.

[17] X. Wang, S. Feng, and W. Q. Yan, "Human gait recognition based on self-adaptive hidden Markov model," in *Proc. IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2019.

[18] I. R. T. Alvarez and G. Sahonero-Alvarez, "Cross-view gait recognition based on U-Net," in *Proc. International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2020, pp. 1–7.

[19] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with Conditional adversarial networks," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, 2017, pp. 5967–5976.

[20] Y. Dupuis, X. Savatier, and P. Vasseur, "Feature subset selection applied to model-free gait recognition," *Image Vis. Comput.*, vol. 31, no. 8, pp. 580–591, 2013.

[21] I. R. T. Alvarez and G. Sahonero-Alvarez, "Gait recognition based on modified gait energy image," in *Proc. 2018 IEEE Sciences and Humanities International Research Conference (SHIRCON)*, IEEE, 2018, pp. 1–4.

[22] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," arXiv preprint, arXiv: 1503.00075, 2015.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp.1735–1780, 1997.

[24] A. Kalsoom, M. Maqsood, M. A. Ghazanfar, F. Aadil, and S. Rho, "A dimensionality reduction-based efficient software fault prediction using Fisher Linear Discriminant Analysis (FLDA)," *Journal of Supercomputing*, vol. 74, no. 9, pp. 4568–4602, 2018.

[25] X. Wang and W. Q. Yan, "Cross-view gait recognition through ensemble learning," *Neural Computing and Applications*, vol. 32, no. 11, pp. 7275–7287, 2020.

[26] M. Mirza and O. Simon, "Conditional generative adversarial nets," arXiv preprint, arXiv: 1411.1784, 2014.

[27] X. Wang and W. Q. Yan, "Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory," *International Journal of Neural Systems*, vol. 30, no. 1, pp. 1950027–1950028, 2020.

[28] J. Amin, M. A. Sharif, K. Seifedine, N. Yunyoung, and W. ShuiHua, "Convolutional Bi-LSTM based human gait recognition using video sequences," *Computers, Materials & Continua*, 2021.