A Unified Convolutional Neural Network for Gait Recognition

Sonam Nahar(✉), Sagar Narsingani, and Yash Patel

Pandit Deendayal Energy University, Gandhinagar, Gujarat, India
{sonam.nahar,sagar.nce19,yash.pce19}@sot.pdpu.ac.in

Abstract. Gait recognition stands as a crucial method for distant person identification. Most state-of-the-art gait recognition frameworks consists of two modules: feature extraction and feature matching. The fixed nature of each part in these modules leads to suboptimal performance in challenging conditions as they are mutually independent. This paper presents the integration of those steps into a single framework. Specifically, we design a unified end-to-end convolutional neural network (CNN) for learning the efficient gait representation and gait recognition. Since dynamic areas contain the most informative part of the human gait and are insensitive to changes in various covariate conditions, we feed the gait entropy images as input to CNN model to capture mostly the motion information. The proposed method is evaluated through experiments conducted on the CASIA-B dataset, specifically for cross-view and cross-walking gait recognition. The experimental results strongly indicate the effectiveness of the approach.

Keywords: Gait Recognition · CNN · Cross Walking

1 Introduction

'Gait' is a measure of how a person walks and is used as a behavioral biometric [12]. A gait recognition system can uniquely identify individuals based on their way of walking. In general, gait is considered to be a poorer biometric than the other image-based biometrics such as fingerprint [3], face [15] and iris [13]. This is largely due to the fact that changes in covariate circumstances are more likely to have an impact on gait than other behavioural biometrics. Some of these factors, like the state of one's clothing, how they are carried objects, and the angle at which they are seen, primarily affects how a person appears, while others, like time, surface, and one's shoes, have an impact on gait. Despite this drawback, gait offers a notable advantage over other image-based biometrics in situations like visual surveillance because it doesn't require close proximity to the subject and can operate without interrupting or interfering with the subject's activity.

Today, gait recognition is an active area of research and development in computer vision and biometrics community, with applicationsranging from security

H. Lu et al. (Eds.): ACPR 2023, LNCS 14407, pp. 230–242, 2023.
https://doi.org/10.1007/978-3-031-47637-2_18

and surveillance to healthcare and sports performance analysis [16,18]. Existing vision-based gait recognition techniques mainly fall into two broad categories, namely model based and model free approaches. Model free approaches [1,6] use model information directly extracted from silhouettes, whilst model-based approaches [25,26] fit a model to body and represent gait using the parameters of the model. Model based approaches are more complex and computationally more expensive than model free approaches, and generally require good quality images to correctly extract the model parameters from a gait sequence, which may not be available in a real-world application scenario such as CCTV surveillance in public space. Most of the recent and state of the art research in gait recognition adopt model free approaches since they are computationally less insensitive, more robust to noise, insensitive to the quality of gait sequences, and have a comparable, or better performance when compared to model-based approaches on benchmark datasets [16,18,24]. The work in this paper fall into the model-free category.

Model free gait recognition approaches use silhouettes for human gait representation that can be computed by applying background subtraction on gait sequences. However, these silhouettes are susceptible to variations in the subject's appearance caused by factors such as different carrying styles, clothing and view conditions. To address this issue, several state-of-art gait recognition methods have been proposed, aiming to develop representations that remain unaffected by covariate conditions. Examples of such representations include gait energy image [6], gait entropy image [1], chrono gait image [21], gait flow image [9], frequency domain gait features [10], and more. Despite achieving reasonably satisfactory outcomes in recent years, these gait recognition strategies typically rely on manually designed features and have limited ability to learn intrinsic patterns within the data.

Deep learning has experienced a surge in popularity within various computer vision domains, including image recognition [7], face recognition [20], and human activity recognition [22], due to its remarkable performance. One key advantage of deep learning is its ability to automatically learn features from a large volume of training samples across different layers of a deep neural network. Additionally, it offers a unified framework for both feature learning and classification. Recently, deep learning-based gait recognition methods have emerged as the dominant approach in the field, enabling practical applications [16]. Among the deep architectures, convolutional neural networks (CNNs) have been extensively utilized for gait recognition. CNNs use convolutional layers to extract local features, pooling layers to reduce dimensionality, and fully connected layers to classify images based on learned features. Most existing CNN based gait recognition methods comprise of two steps: feature learning and feature matching [4,17,23]. CNN is used to extract the higher-level features from the gait silhouettes and then a traditional classifier is further used to compare the probe features with the gallery ones in order to identify most similar gait patterns and label them as being from the same subject. However, these frameworks demonstrate subpar performance when faced with demanding tasks such as cross-view

or cross-clothing gait recognition on extensive datasets. A potential explanation for this is the sequential approach employed, which leads to a lack of compatibility between the feature learning and classifier training processes.

To better address above problems, it is probably a good choice to adopt an end-to-end framework. A recent study introduced a comprehensive CNN architecture that simultaneously learns gait segmentation and recognition in an end-to-end manner [19]. The segmentation model extracts gait silhouettes, which are subsequently fused using a temporal fusion unit to improve the recognition model's performance. The study also proposes a unified model that trains both segmentation and recognition jointly. While this approach achieved superior results on benchmark datasets, it employs two separate CNN models, which are intricate in terms of architecture. Furthermore, training the CNNs individually and subsequently fine-tuning them in a unified framework is computationally demanding. In this paper, we present the integration of the steps feature learning and feature matching into a single framework. Specifically, we design a unified end-to-end convolutional neural network (CNN) that learns the gait features and perform recognition jointly. View angles, walking with carrying a bag and walking with wearing a coat are used as different covariate conditions for the experiments. We demonstrate the effectiveness of our method in the settings of cross view and cross walking using the large benchmark dataset. Here, cross-view means the test gait sequences are with different view angles from the view point in training sequences. In cross-walking setting, the subjects in test set have walking sequences either with a coat or with a bag, while subjects in the training set are under the normal walking condition. The key contributions of the paper are summarized as follows:

1. We introduce a straightforward CNN model that achieves efficient gait representation and performs gait recognition within a unified framework. The unified model offers two notable advantages: firstly, it significantly simplifies the conventional step-by-step procedures, and secondly, the joint learning of each component yields noticeable performance improvements compared to separate learning approaches.
2. Most of the existing CNN based approaches use gait energy images (GEI) as an input because GEI can efficiently capture both the static (e.g., head, torso) and dynamic parts (e.g., lower parts of legs and arms) of the human silhouette [16]. However, since GEI mainly contain body shape information, they are sensitive to changes in various covariate conditions, and hence is not an appropriate representation to feed in a CNN for learning robust gait features. In our work, we propose to use gait entropy images to be fed as input to our CNN model. Gait entropy image captures the dynamic areas of the human body by measuring the Shannon's entropy [1]. Dynamic areas contain the most informative part of the human gait and are insensitive to changes in various covariate conditions. With gait entropy images as input, our CNN model learns higher level gait features which are invariant to different camera viewpoints, clothing and carrying conditions.

3. We present extensive experimental results for cross-view and cross-walking gait recognition using the CASIA-B benchmark dataset [24].

The rest of the paper is organized as follows: in Sect. 2, related work is reviewed. The proposed gait recognition method is detailed in Sect. 3. Experimental results and conclusion are presented in Sect. 4 and Sect. 5, respectively.

2 Related Work

Typically, model-free gait recognition approaches comprise of two steps: feature extraction and feature matching. The gait features are mainly represented by the silhouettes that are extracted from human walking sequences. The gait energy image (GEI) has been widely used as an effective representation by averaging the pixel values of the silhouettes over the gait period [6]. However, GEI suffers from information loss in gait sequences, which hampers performance when dealing with changes caused by covariate conditions like clothing, carrying variations, and view differences. To address this issue, an entropy-based gait representation is proposed [1,2]. The gait entropy image focuses on dynamic regions and is computed by calculating the pixel-wise entropy of the GEI. Another variant, called Chrono-Gait Image (CGI) is introduced to preserve temporal information by utilizing a multi-channel temporal encoding scheme [21]. Additionally, a gait flow image (GFI) directly emphasizes the dynamic components by averaging the lengths of optical flow observed on the silhouette contour over the gait period [9]. Frequency domain gait features are also proposed [8,10], taking into account the periodic nature of gait. These frequency-based methods learn cross-view projections to normalize gait features, enabling comparison of normalized features from different views and computation of their similarity when comparing two videos.

The latest advancements in gait recognition methods utilizing such gait representations have demonstrated promising outcomes even in the presence of challenging covariate conditions [16,18]. However, these methods rely on manually designed gait features, which possess limited capability to capture the underlying patterns inherent in the data. Moreover, these image-based gait features are transformed into a feature vector, and techniques like linear discriminant analysis (LDA) [6], primal rank support vector machines [5], and multi-view discriminant analysis (MvDA) [11] are employed to extract relevant gait features that remain unaffected by various covariate conditions. Nevertheless, treating each dimension in the feature vector as a separate pixel for subsequent classification/recognition fails to capture the spatial proximity within the gait image, resulting in overfitting issues.

Deep learning-based gait recognition techniques have gained significant prominence in recent times by leveraging their ability to automatically acquire discriminative gait representations. Convolutional neural networks (CNNs) have been predominantly employed due to their capability to capture spatial proximity within images through convolutional operations, resulting in substantial enhancements in recognition accuracy. A noteworthy example is GEI-Net, which

directly learns gait representations from gait energy images (GEIs) using CNNs [17]. The authors in [23] proposed a deep CNN-based framework for cross-view and cross-walk gait recognition, where similarities between pairs of GEIs are learned, leading to state-of-the-art performance. Another recent approach called GaitSet treats gait as a set and assumes that the silhouette's appearance contains positional information [4]. It utilizes CNNs to extract temporal information from the gait set. All these methods utilize CNNs for learning gait features, while a separate feature matching module is employed for gait recognition. Typically, direct template matching between gallery and probe features or a K-NN classifier is used for this purpose. However, the fixed nature of the feature learning and feature matching modules results in suboptimal performance under challenging conditions, as they operate independently of each other. To address this limitation, the authors in [19] recently introduced a comprehensive CNN architecture named as GaitNet that simultaneously learns gait segmentation and recognition in an end-to-end manner.

Apart from CNN, various other deep architectures have emerged to address gait recognition challenges [16]. These include deep belief networks (DBN), long short-term memory (LSTM) networks (a type of recurrent neural network), deep autoencoders (DAE), generative adversarial networks (GAN), capsule networks, and hybrid networks that combine multiple architectures. While these methods have demonstrated remarkable performance in demanding scenarios, they often involve complex network structures and necessitate large amounts of labeled data for effective training.

3 Proposed Method

3.1 Generation of Gait Entropy Images

In the given human walking sequence, a silhouette is extracted from each frame utilizing background subtraction [14]. Subsequently, the height of the silhouettes is normalized and aligned at the centre. Gait cycles are then estimated by employing the autocorrelation method described in the [10]. The gait cycle represents the time interval between repetitive walking events, typically commencing when one foot makes contact with the ground. As the walking pattern of an individual is periodic, it is adequate to consider just a single gait cycle from the entire gait sequence.

After obtaining a gait cycle consisting of size-normalized and centre-aligned silhouettes, the next step involves computing a gait entropy image (GEnI). This is achieved by calculating the Shannon entropy for each pixel within the silhouette images throughout the entire gait cycle, as outlined in [1]:

$$GEnI = I(x, y) = \sum_{k=1}^{K} p_k(x, y) \log_2 p_k(x, y), \tag{1}$$

where x and y represent the pixel coordinates, while $p_k(x, y)$ denotes the probability of the pixel having the k^{th} value within a complete gait cycle. For our

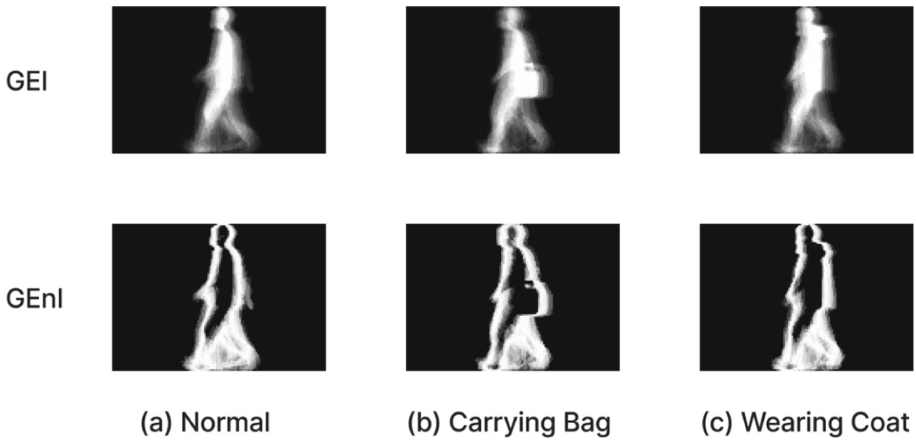

Fig. 1. Examples of Gait Energy Images (GEI) and Gait Entropy Images (GEnI) from the CASIA-B dataset [24] with different walking conditions. Columns (a) Normal Walking, (b) Carrying a bag, and (c) Wearing a coat.

specific case, since the silhouettes are binary images, we have $K = 2$. Examples of Gait Entropy Images (GEnIs) and Gait Energy Images (GEIs) extracted from the CASIA-B dataset are depicted in Fig. 1. The GEnIs exhibit higher intensity values in dynamic areas such as the legs and arms, while the static regions like the head and torso demonstrate lower values. This discrepancy arises due to the greater uncertainty and information content of silhouette pixels in the dynamic areas, resulting in higher entropy values. Additionally, the impact of appearance changes caused by carrying a bag or wearing a coat is more pronounced in the GEIs, whereas in the GEnIs, it is substantially diminished and mainly observable in the outer contour of the human body. In our proposed approach, we utilize GEnIs as input to our CNN architecture, allowing our deep model to learn gait features that are invariant to both view and appearance variations which results in effective recognition performance in challenging scenarios.

3.2 Convolutional Neural Network Architecture

Our CNN architecture is comprised of eight layers, with the initial six layers consisting of two sets of convolutions, batch normalization, and pooling layers. The final two layers are fully connected (FC) layers, where the first FC layer contains 1024 units, followed by another FC layer with M number of output units. At each output unit, the SoftMax function is applied. Assuming there are M subjects in the training set, each subject is represented by an integer number ranging from 1 to M.

More precisely, the i^{th} unit in the final layer is ideally designed to output 1 when the input belongs to subject i, while it outputs 0 otherwise. The architecture of our CNN is depicted in Fig. 2. Regularization is implemented using

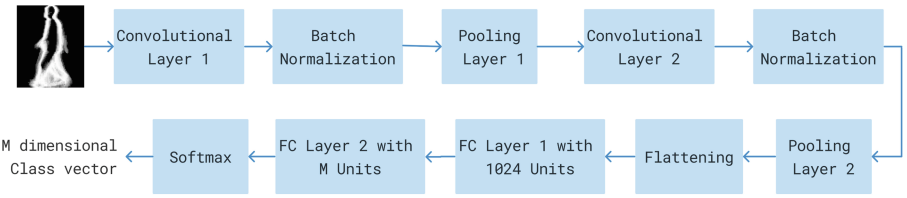

Fig. 2. The CNN Architecture.

dropout, and non-linearity is introduced using the ReLU activation function at every layer except the last layer. To optimize our CNN model, we consider several hyperparameters, including the number of filters, size of filters, number of epochs, dropout rate, and batch size. We determine the optimal values for these hyperparameters through cross-validation. The configuration of each convolutional and pooling layer, along with their corresponding optimal hyperparameters, is presented in Table 1. We employ this configuration for gait feature learning and recognition in an end-to-end manner.

Table 1. CNN Configuration with Optimal Hyperparameters.

Hyperparameters		
Conv Layer 1	# Filters	40
	Size of Filters	3×3
	Stride	2
Max Pooling Layer 1	Size of Filters	2×2
	Stride	2
Conv Layer 2	# Filters	32
	Size of Filters	5×5
	Stride	3
Max Pooling Layer 2	Size of Filters	3×3
	Stride	2
Number of Epochs	100	
Learning Rate	0.001	
Dropout Rate	20%	
Batch Size	64	

Despite being relatively shallow, our CNN architecture effectively learns robust gait representations and achieves good recognition accuracy. This is primarily because gait data, such as silhouettes (e.g., GEnI), do not possess significant complexity in terms of texture information. Therefore, a shallow CNN architecture is sufficient for capturing and encoding gait characteristics. This

stands in contrast to other domains like face recognition [20] or activity recognition [22], where deep networks are commonly employed to learn highly discriminative features. Additionally, our preliminary experiments have confirmed that incorporating additional convolutional layers after the existing sets of convolutions, batch normalization, and pooling in our network do not lead to a significant improvement in gait recognition accuracy.

3.3 Training the CNN

Let the training set consist of N gait sequences belonging to M subjects. We start by computing a gait entropy image (GEnI) for each gait sequence, resulting in a set of training GEnIs denoted as $\{I_1, I_2, \ldots, I_N\}$. These GEnIs are accompanied by their corresponding ground truth label vectors $\{y_1, y_2, \ldots, y_N\}$. Each input GEnI, I_i belonging to subject j, is associated with a label denoted by a M-dimensional vector, $y_i = [y_{i1}, y_{i2}, \ldots, y_{iM}]$, where y_{ij} is equal to 1 and the remaining entries are set to 0.

Given an input image I_i and set of weighting parameters w, we define the output z_i of last layer as a function of I_i and w using forward propagation as:

$$z_i = f(I_i, w), \tag{2}$$

where z_i is denoted as a M dimensional vector. We subsequently apply the SoftMax function on every element of z_i and obtain the final class probability vector $\hat{y}_i$ where each j^{th} element in $\hat{y}_i$ is computed using the SoftMax function as follows,

$$\hat{y}_{ij} = \frac{e^{z_{ij}}}{\sum_{j=1}^{M} e^{z_{ij}}}, \quad for\ j = 1, 2, \ldots, M \tag{3}$$

Given a training set: $\{I_1, I_2, \ldots, I_N\}$, we train our CNN network by minimizing the following cross-entropy loss function L using the stochastic gradient descent algorithm:

$$L(w) = \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} \log \hat{y}_{ij}. \tag{4}$$

In this training phase, we obtain an optimal set of weighting parameters $\hat{w}$ as:

$$\hat{w} = \arg\min_{w} \quad L(w). \tag{5}$$

3.4 Gait Recognition

Given a gait sequence in the test set, we begin by calculating its gait entropy image. This gait entropy image serves as the input to the trained CNN model. At the intermediate layers, the CNN model learns higher-level features, and at the final layer, it performs classification using the SoftMax function. In other words, by utilizing the optimal set of weighting parameters $\hat{w}$, we obtain an M-dimensional class probability vector $\hat{y}_t$ for the test GEnI I_t using the Eq. 2 and 3. The class label or subject ID is determined by selecting the index corresponding to the highest value in $\hat{y}_t$.

4 Experimental Results

4.1 Datasets and Test Protocol

Our experiments utilize the CASIA-B dataset [24]. CASIA-B is a highly utilized and openly accessible gait database, comprising gait sequences of 124 individuals captured from 11 distinct viewpoints spanning from 0° to 180° (with 18° increments). The dataset encompasses three variations of walking conditions namely normal walking (NM), walking with a coat (CL), and walking with a bag (BG), respectively with 6, 2, and 2 gait sequences per subject per view.

To evaluate the effectiveness of our proposed method, we employ the subject-dependent testing protocol. Under this protocol, both the training and testing datasets comprise gait sequences from all subjects in our dataset. For the cross-walking experimental setup, we exclusively utilize four normal walking sequences (NM) of each subject, incorporating each view, for training purposes. Consequently, our training set encompasses a total of 44 gait sequences per subject (11 views multiplied by 4 sequences), resulting in a grand total of $44 * 124 = 5456$ sequences. Conversely, the test set comprises the remaining sequences, including two normal (test subset-NM), two coat (test subset-CL), and two bag (test subset-BG) sequences for each subject and each view. In the case of the cross-view scenario, we assess the test accuracy for each view independently during the cross-walking experiment.

To initiate the training process of the CNN, we begin by computing the gait entropy image (GEnI) for each gait sequence within the training set. These GEnIs have a fixed size of 88×128 pixels. An example of these extracted GEnIs can be observed in Fig. 1. Since our training set is of moderate size, we take precautions to mitigate the risk of overfitting. We accomplish this by employing cross-validation, where we randomly split the training set into a 70% training subset and a 30% validation subset. By selecting hyperparameters that minimize variance error, we aim to identify a set of hyperparameters in which both the training and validation errors are low and their discrepancy is also minimal. Utilizing the training GEnIs and the optimal set of hyperparameters (refer to Table 1), we proceed to train our CNN model as illustrated in Fig. 2, allowing us to learn the weights ($\hat{w}$) of the model. Since our model is trained using GEnIs of normal walking gait sequences captured with different view angles, it learns the reliable gait features invariant to view angles, clothing and carrying conditions. It is important to note that TensorFlow was utilized for training the CNN, and all experiments were conducted on a machine equipped with an 11th Gen Intel (R) Core (TM) i7-1165G7 processor running at a frequency of 2.80 GHz. The machine also featured an SSD with 512 MB storage capacity, 8 GB of RAM, and operated on a 64-bit operating system.

4.2 Results

Rank-1 recognition accuracy is utilized to measure and present our results. Table 2 displays the outcomes for the cross-view scenario across three test sub-

sets. The first row corresponds to the case when the test subset consists of normal walking sequences, while the second and third rows demonstrate the results when the test subsets comprise bag and coat sequences, respectively. In order to incorporate the cross-view setting within the cross-walking scenario, we assess the recognition accuracy for each individual angle, as illustrated in Table 2. The task of cross-view and cross-walking gait recognition is particularly challenging. However, our method has yielded highly promising outcomes on the NM subset. Despite the fact that silhouettes captured from frontal and back viewpoints, such as 0°, 18°, 162°, and 180°, contain limited gait information, our method achieves a remarkable accuracy rate of over 90% on the NM subset. Regarding the BG subset, our method continues to perform well, particularly with test angles of 72°, 90°, 108°, and 126°, as these viewpoints offer the most visible gait information. However, the performance declines for the CL test subset, even for the 90° view, as indicated in Table 2. The reason behind these outcomes is that carrying a bag only impacts a small portion of the gait silhouette, whereas wearing a coat can significantly alter one's appearance. Another potential factor contributing to the performance degradation could be the scarcity of training data. Given the greater appearance variations, the CL subset proves to be more challenging than the NM and BG subsets. Insufficient training data may result in networks being prone to overfitting.

Table 2. Recognition Accuracies in Cross Walking and Cross View Scenarios using CASIA-B dataset. First Row: Results for NM test subset. Second Row: Results for BG test subset. Third Row: Results for CL test subset. Here, the training set contains 4 normal walking sequences of every subject captured with all 18 view angles.

Training Set: NM #1-4 with 0°–180°												
Test Subset	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Mean
NM #5-6	91.42	92.59	93.44	95.34	95.57	96.90	95.56	94.77	95.72	93.31	92.49	**94.46**
BG #1-2	27.77	37.12	37.12	46.32	50.76	53.15	51.18	41.48	34.01	33.13	30.10	**40.19**
CL #1-2	14.11	18.13	19.40	22.22	25.28	22.29	22.01	20.72	18.61	17.66	16.15	**19.69**

Table 3 presents the comparison of our proposed method with existing approaches in the literature. These methods were selected as they are the most recent ones evaluated under cross-view and cross-walking scenarios, and their scores were directly obtained from the original papers. For a fair comparison, we also include two traditional methods, namely GEI and GEnI, which utilize gait energy image (GEI) and gait entropy image (GEnI) as features, respectively. These traditional methods perform gait recognition in cross-walking and cross-view scenarios by employing direct template matching (Euclidean distance)

between the training and test sets. The available results for cross-walking scenarios involving the BG and CL test sets are limited. However, our proposed method outperforms traditional methods such as GEI and GEnI by significant margins. This highlights the advantage of using learned features from CNN compared to handcrafted features in gait recognition. In the experimental setting where both the training and test sets consist of NM sequences, our method demonstrates superior performance compared to state-of-the-art methods such as CMCC [8] and GEINET [17], and comparable performance with recent deep methods including GaitSet [4], CNN-LB [23], and GaitNet [19]. Moreover, for the bag and coat test sets, our proposed method shows comparable performance. These results underscore the effectiveness of our end-to-end recognition model based on CNN, which outperforms methods that separate feature extraction and recognition into distinct phases.

Table 3. Comparison with state-of-the art gait recognition methods in terms of Rank-1 accuracy (%) using CASIA-B dataset. Here, '-' denotes that results are not reported.

Test		Training (NM #1-4): 0° - 180°											
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Mean
NM #5-6	GEI	9.14	14.87	15.03	16.07	23.75	22.5	23.91	17.2	12.44	13.49	9.02	16.13
	GEnI	12.45	16.12	13.1	12.38	15.02	23.62	23.21	21.10	15.20	15.57	16.18	16.72
	CMCC [8]	46.3	-	-	52.4	-	48.3	-	56.9	-	-	-	-
	GEINet [17]	45.8	57.6	67.1	66.9	56.3	48.3	58.3	68.4	69.4	59	46.5	58.5
	CNN-LB [23]	79.1	88.4	95.7	92.8	89.1	87	89.3	92.1	94.4	89.4	75.4	88.4
	GaitSet [4]	90.8	97.9	99.4	96.9	93.6	91.7	95	97.8	98.9	96.8	85.8	95
	GaitNet [19]	75.6	91.3	91.2	92.9	92.5	91.0	91.8	93.8	92.9	94.1	81.9	89.9
	Proposed	91.42	92.59	93.44	95.34	95.57	96.90	95.56	94.77	95.72	93.31	92.49	94.46
BG #1-2	GEI	6.44	9.91	11.61	9.32	15.32	12.21	11.65	11.97	8.26	8.94	6.37	10.18
	GEnI	7.64	9.82	7.31	7.08	9.81	13.63	12.37	12.59	7.27	7.26	7.97	9.34
	CMCC [8]	-	-	-	-	-	-	-	-	-	-	-	-
	GEINet [17]	-	-	-	-	-	-	-	-	-	-	-	-
	CNN-LB [23]	64.2	80.6	82.7	76.9	64.8	63.1	68	76.9	82.2	75.4	61.3	72.4
	GaitSet [4]	83.8	91.2	91.8	88.8	83.3	81	84.1	90	92.2	94.4	79	87.2
	GaitNet [19]	-	-	-	-	-	-	-	-	-	-	-	-
	Proposed	27.77	37.12	37.12	46.32	50.76	53.15	51.18	41.48	34.01	33.13	30.10	40.19
CL #1-2	GEI	2.57	4.27	6.23	6.32	6.28	7.01	6.6	6.97	5.18	4.62	2.54	5.33
	GEnI	3.02	5.01	4.61	4.52	6.44	8.24	8.06	8.64	6.36	5.33	4.42	5.88
	CMCC [8]	-	-	-	-	-	-	-	-	-	-	-	-
	GEINet [17]	-	-	-	-	-	-	-	-	-	-	-	-
	CNN-LB [23]	37.7	57.2	66.6	61.1	55.2	54.6	55.2	59.1	58.9	48.8	39.4	54
	GaitSet [4]	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50	70.4
	GaitNet [19]	-	-	-	-	-	-	-	-	-	-	-	-
	Proposed	14.11	18.13	19.40	22.22	25.28	22.29	22.01	20.72	18.61	17.66	16.15	19.69

5 Conclusion

In this paper, we have introduced a unified CNN model that combines gait feature learning and recognition into a single end-to-end network. This model has the capability to automatically discover discriminative representations for gait recognition, ensuring invariance to variations in view angle, clothing, and carrying conditions. By integrating the learning process of each component, our unified model simplifies the traditional step-by-step procedure and yields noticeable performance improvements compared to separate learning approaches. To evaluate the effectiveness of our proposed method, we have conducted extensive experiments using the CASIA-B gait dataset. The experimental results have clearly demonstrated that our method surpasses the performance of state-of-the-art traditional methods and achieves comparable results to recent deep learning-based gait recognition methods in both cross-view and cross-walking scenarios.

References

1. Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. In: 3rd International Conference on Imaging for Crime Detection and Prevention, ICDP 2009, pp. 1–6 (2009)
2. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. Pattern Recogn. Lett. **31**(13), 2052–2060 (2010)
3. Cao, K., Jain, A.K.: Automated latent fingerprint recognition. IEEE Trans. Pattern Anal. Mach. Intell. **41**(4), 788–800 (2019)
4. Chao, H., He, Y., Zhang, J., Feng, J.: GaitSet: regarding gait as a set for cross-view gait recognition. CoRR abs/1811.06186 (2018)
5. Chen, X., Xu, J.: Uncooperative gait recognition. Pattern Recogn. **53**(C), 116–129 (2016)
6. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2006)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, NIPS 2012, vol. 1, pp. 1097–1105. Curran Associates Inc., Red Hook, NY, USA (2012)
8. Kusakunniran, W., Wu, Q., Zhang, J., Li, H., Wang, L.: Recognizing gaits across views through correlated motion co-clustering. IEEE Trans. Image Process. **23**(2), 696–709 (2014)
9. Lam, T., Cheung, K., Liu, J.: Gait flow image: a silhouette-based gait representation for human identification. Pattern Recogn. **44**, 973–987 (2011)
10. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait recognition using a view transformation model in the frequency domain. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 151–163. Springer, Heidelberg (2006). https://doi.org/10.1007/11744078_12
11. Mansur, A., Makihara, Y., Muramatsu, D., Yagi, Y.: Cross-view gait recognition using view-dependent discriminative analysis. In: IEEE International Joint Conference on Biometrics, pp. 1–8 (2014)
12. Murray., M.: Gait as a total pattern of movement (1967)

13. Nguyen, K., Fookes, C., Jillela, R., Sridharan, S., Ross, A.: Long range iris recognition: a survey. Pattern Recogn. **72**, 123–143 (2017)
14. Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., Bowyer, K.: The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans. Pattern Anal. Mach. Intell. **27**(2), 162–177 (2005)
15. Sepas-Moghaddam, A.: Face recognition: a novel multi-level taxonomy based survey. IET Biometrics **9**, 58–67 (2020)
16. Sepas-Moghaddam, A., Etemad, A.: Deep gait recognition: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **45**(1), 264–284 (2023)
17. Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., Yagi, Y.: GEINet: view-invariant gait recognition using a convolutional neural network. In: 2016 International Conference on Biometrics (ICB), pp. 1–8 (2016)
18. Singh, J.P., Jain, S., Arora, S., Singh, U.P.: Vision-based gait recognition: a survey. IEEE Access **6**, 70497–70527 (2018)
19. Song, C., Huang, Y., Huang, Y., Jia, N., Wang, L.: GaitNet: an end-to-end network for gait based human identification. Pattern Recogn. **96**(C), 106988 (2019)
20. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
21. Wang, C., Zhang, J., Wang, L., Pu, J., Yuan, X.: Human identification using temporal information preserving gait template. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2164–2176 (2012)
22. Wang, J., Chen, Y., Hao, S., Peng, X., Hu, L.: Deep learning for sensor-based activity recognition: a survey. CoRR abs/1707.03502 (2017)
23. Wu, Z., Huang, Y., Wang, L., Wang, X., Tan, T.: A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Trans. Pattern Anal. Mach. Intell. **39**(2), 209–226 (2017)
24. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: 18th International Conference on Pattern Recognition, ICPR 2006, vol. 4, pp. 441–444 (2006)
25. Zhang, R., Vogler, C., Metaxas, D.: Human gait recognition at sagittal plane. Image Vis. Comput. **25**(3), 321–330 (2007)
26. Zhao, G., Liu, G., Li, H., Pietikainen, M.: 3D gait recognition using multiple cameras. In: 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, pp. 529–534 (2006)