Learning Effective Gait Features For Gait Recognition Using Deep Convolutional Autoencoder

Abstract: Gait recognition is important for identifying individuals at a distance but becomes challenging by covariates such as clothing and carrying conditions. Recent methods, often reliant on supervised learning and extensive labeled data, may not be feasible in applications with large datasets. To address this, a new method using a deep convolutional autoencoder—an unsupervised learning technique—has been developed to extract distinctive gait features resilient to these variables. This technique reduces the dimensionality of feature space representing the gait data and these features are classified using softmax classifier. Experimented on the CASIA-B dataset, this approach demonstrates superior performance in gait recognition.

Keywords: Gait Recognition, Unsupervised Feature Learning, Convolutional Autoencoder.

1 Introduction

Gait, a behavioral biometric, involves analyzing an individual's walking style to uniquely identify them from a distance, even in low-resolution or uncontrolled environments and does not require subject cooperation [Mu67]. The most popular techniques for gait recognition are *model-free* which this paper utilizes, are suitable for lower-resolution gait videos and are less computationally intensive [HB06, BXG09]. These model-free techniques typically use silhouettes to represent human gait but face challenges due to variations in appearance from different carrying styles, clothing, and viewing conditions. They often depend on manually designed features, limiting their ability to learn intrinsic patterns from the data effectively [Wa10]. Recent advancements in deep learning, particularly through convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have enhanced gait recognition but rely heavily on supervised learning and extensive labeled data, which can be biased and impractical in scenarios like surveillance. Unsupervised learning techniques offer a promising alternative, excelling at identifying complex patterns and dependencies for more robust gait representation.

This paper introduces an unsupervised learning approach using a deep convolutional autoencoder to develop efficient gait representations from unlabeled gait data for gait recognition where the subject's identity in the gait sequence is unknown. The model considers spatial proximity within the gait image to extract key features in a latent space with significantly reduced dimensions, ensuring these features are unaffected by variations in clothing and carrying conditions. Our model is better suited for biometric and real-time gait recognition systems, effectively learning gait features from unlabeled data in an unsupervised manner. It represents these features in very low dimensions, optimizing space usage and improving efficiency in terms of processing time and recognition accuracy. To the best of

our knowledge, this is the first work which uses convolutional autoencoder to learn low dimensional gait features for gait recognition and our experimental results show the superior recognition performance using the CASIA-B dataset [YTT06].

2 Related Work

In model-free gait recognition, gait silhouettes are transformed to produce distinctive features, enhancing robustness against varying conditions. Common such gait representations include gait energy image [YTT06], gait entropy image [BXG09], masked GEnI [BXG10], frequency domain gait entropy [Ro15], gait flow image [LCL11], and chrono gait image [Wa12]. However, these manually crafted features have limited ability to detect intrinsic data patterns. Techniques like PCA, LDA, and MvDA are used to reduce dimensionality and achieve feature invariance, but they often overlook spatial proximity in gait images, leading to potential overfitting issues [HB06, Ma14].

Over the past decade, deep learning has significantly advanced gait recognition, achieving state-of-the-art results [SME23]. For, instance, Shiraga et al. [Sh16] developed GEI-Net, an eight-layer CNN that uses gait energy images for feature extraction. Chunfeng et al.'s GaitNet [So19] integrates gait segmentation and recognition within a unified CNN architecture. GaitSet [Ch18] analyzes gait as a set using CNNs to extract temporal data from silhouettes, while GaitPart [Fa20] focuses on partial spatio-temporal features. Additionally, LSTM based methods are used to maintain temporal information within gait sequences [FLL16, BP19]. Despite their advancements, these approaches rely on supervised learning and extensive labeled data, posing challenges in practical applications due to the high cost and effort required to obtain this data.

Using unsupervised deep learning models is advantageous for learning effective gait representations, as they focus on capturing intrinsic gait structures without predefined labels. However, research in this area is limited. Notable efforts include Shiqi et al.'s use of stacked autoencoders for view-invariant features [Yu17b], and the application of deep autoencoders to differentiate identity and covariate traits in gait sequences [Li20]. Additionally, generative adversarial networks [Yu17a] were employed in other studies for generating feature maps without covariates. While these methods have shown impressive performance, they often involve complex network structures and require large amounts of unlabeled data, leading to high computational complexity.

3 Proposed Work

The workflow of our proposed method is displayed in Figure 1.

3.1 Convolutional Autoencoder for Feature Learning

The convolutional autoencoder (CAE) has become a key model in unsupervised learning, designed to extract highly effective, low-dimensional features from images known as

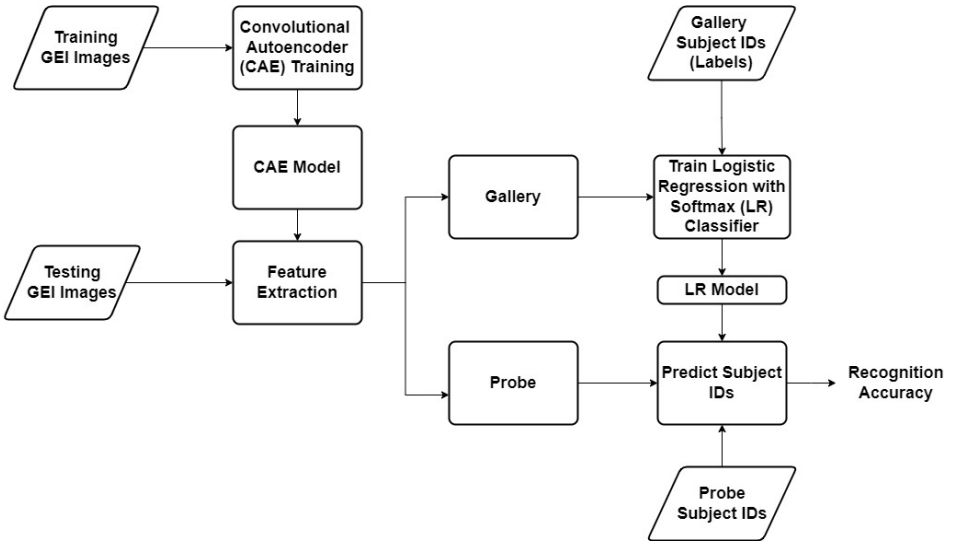


Fig. 1: Diagram illustrating the workflow of the proposed approach.

bottleneck features, and to reconstruct the original image in the output layer. The CAE consists of an encoder and a decoder, both built with convolutional layers. The encoder compresses the image into a compact feature representation in a lower-dimensional latent space, while the decoder reconstructs the image from this compressed form. The architecture of our convolutional autoencoder is shown in Figure 2. Our CAE's encoder has three convolutional layers followed by two fully connected layers, where the second acts as the bottleneck layer, significantly reducing the dimensionality compared to the original image. The decoder reverses this process with two fully connected layers and three convolutional layers that mirror the encoder's structure. This setup allows the network to maintain spatial details and leverage translational invariance of images. The model's hyperparameters, such as the number of filters, filter size, stride, and padding, are optimized and detailed in Figure 2. We use gait energy image (GEI) as input for our convolutional autoencoder because GEI is a popular feature used in both traditional and deep learning gait analysis. GEI is computed by averaging the pixel values of the silhouettes over one gait cycle.

Let the training set comprise N gait sequences associated with M subjects. Our initial step involves computing a gait energy image for each gait sequence. Consequently, we compute a collection of training GEIs denoted as $\{I_1, I_2, \ldots, I_N\}$, which serves as the input to our CAE. Given an input I_i and the encoder weights $\mathbf{w}$, we compute the encoder's output via an encoder function f:

$$z_i = f(I_i; \mathbf{w}). \tag{1}$$

Here, z_i is the encoded feature of the input I_i. The decoder then reverts this encoded feature to reconstruct the input image as follows:

$$I_i' = g(z_i; \mathbf{w}') = g(f(I_i; \mathbf{w}); \mathbf{w}'), \tag{2}$$

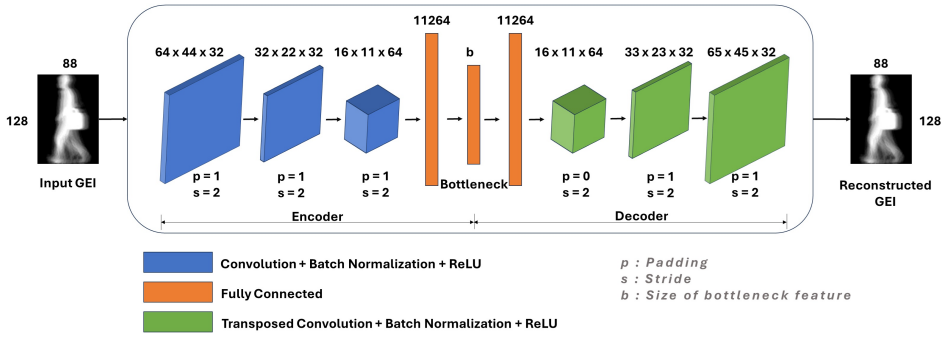


Fig. 2: Convolutional autoencoder architecture for learning gait features.

where I_i' is the reconstructed image, g is the decoder function and $\mathbf{w}'$ are the decoder weights. Given a training set $\{I_1, I_2, \ldots, I_N\}$, we train the CAE by minimizing the mean squared difference between the input and the reconstructed image using the stochastic gradient descent algorithm:

$$\begin{aligned}[\mathbf{w}, \mathbf{w}'] &= \min \frac{1}{N} \sum_{i=1}^{N} ||I_i - I_i'||^2 \\ &= \min \frac{1}{N} \sum_{i=1}^{N} ||I_i - g(f(I_i; \mathbf{w}); \mathbf{w}')||^2.\end{aligned} \tag{3}$$

During this learning phase, we identify an optimized set of weighting parameters, denoted as $\mathbf{w}$ and $\mathbf{w}'$. After training the model, the bottleneck features are obtained using equation (1) and are utilized for gait recognition. As depicted in Figure 2, the bottleneck layer, encompassing b number of neurons, yields a b-dimensional feature vector. The bottleneck is used as a gait feature for the gait recognition task.

3.2 Gait Recognition

For a given set of gallery and probe GEIs, we derive b-dimensional gait features using our trained CAE model. Subsequently, we employ logistic regression with a softmax classifier to perform gait recognition. The softmax classifier is trained by utilizing the gait features in the gallery set, leading to the identification of subjects/classes for each gait sequence present in the probe set. To boost recognition accuracy, we use LDA on the CAE features to reduce dimensions further, say b' (where $b' < b$). The LDA is trained on gallery features, and then applied to transform both gallery and probe features into a lower dimensional space.

Let the gallery set $\{x_1, x_2, \ldots, x_n\}$ contains n gait features from m subjects, each accompanied by its ground truth label vectors $\{y_1, y_2, \ldots, y_n\}$. Each input feature, $x_i \in \mathbb{R}^{b'}$ for subject j is associated with a label denoted by an m-dimensional vector $y_i = [y_{i1}, y_{i2}, \ldots, y_{im}]$,

where y_{ij}=1 and all other entries are 0. Note that without LDA, $x_i \in \mathbb{R}^b$. We train a softmax classifier on the gallery set by minimizing cross-entropy loss using stochastic gradient descent, resulting in the weights $\boldsymbol{w}$ given by:

$$\boldsymbol{w} = \arg\min \Big(\sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} \log \hat{y}_{ij} \Big), \tag{4}$$

where $\hat{y}_{ij}$ denotes the predicted probability for the j^{th} class (subject) when given an input feature vector x_i and weights $\boldsymbol{w} \in \mathbb{R}^{b' \times m}$. It is computed using the softmax function as:

$$\hat{y}_{ij} = \frac{e^{x_i^T \boldsymbol{w}_j}}{\sum_{k=1}^{m} e^{x_i^T \boldsymbol{w}_k}}, \quad \text{for } j = 1, 2, \ldots, m \tag{5}$$

Here, $\boldsymbol{w}_j$ corresponds to the j^{th} column in $\boldsymbol{w}$. Given a probe feature x_p and learned weights $\boldsymbol{w}$, the class probabilities are calculated using equation (5). The class label or subject ID is inferred by selecting the index corresponding to the highest probability:

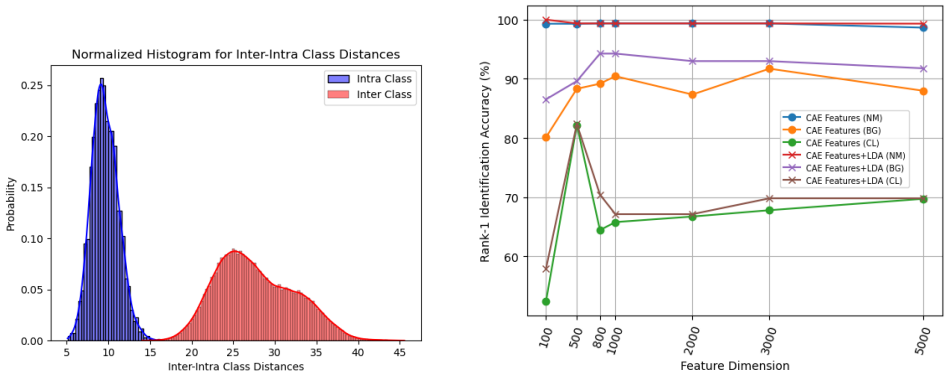
$$\arg\max(\hat{y}_{p1}, \hat{y}_{p2}, \ldots, \hat{y}_{pm}). \tag{6}$$

4 Experimental Results

4.1 Experimental Setup

We conducted our experiments using the CASIA-B dataset [YTT06], which includes gait sequences from 124 subjects across three different walking scenarios: normal walking (NM), walking with a bag (BG), and walking in a coat (CL). Each subject is represented by 6 sequences for normal walking, and 2 sequences each for the bag and coat conditions.

We assessed our method using a subject-independent testing protocol, where the training and test sets contain non-overlapping subjects. We selected 74 subjects for training and reserved 50 for testing. Our training involved a convolutional autoencoder (CAE) model using GEIs sized at 128×88 pixels derived from every gait sequence across all views, totalling $20,000$ GEIs for training. The CAE was trained with a learning rate of 0.0005, a batch size of 32, and over 500 epochs, exploring various bottleneck feature dimensions (denoted as b. We evaluated our gait recognition method using a test set of 50 subjects, divided into a gallery set with the first four normal walking (NM) sequences and a probe set including the remaining 2 NM, 2 BG, and 2 CL sequences for each subject. This creates a cross-walking scenario, where gallery subjects are observed in normal conditions, while probe sequences involve carrying items or clothing. We focused on the optimal 90° camera angle for our experiments because it provides most gait information. We did not account for variations in viewing angles. Gait features from the gallery were trained using logistic regression with a softmax function, regularization parameter of 0.05, and up to 1000 iterations.

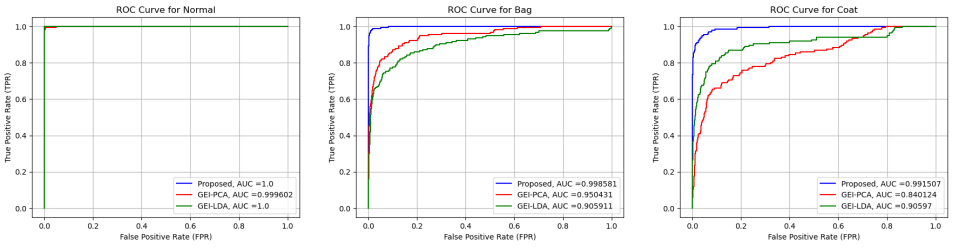


(a) (b)

Fig. 3: (a) Inter and intra class variation histogram. (b) Plot depicting the relationship between "Feature Dimension" and "Recognition Accuracy" across three cross-walking scenarios: normal gait sequences (NM), gait sequences with a bag (BG), and gait sequences with a coat (CL).

4.2 Effects of Features and Dimensionality

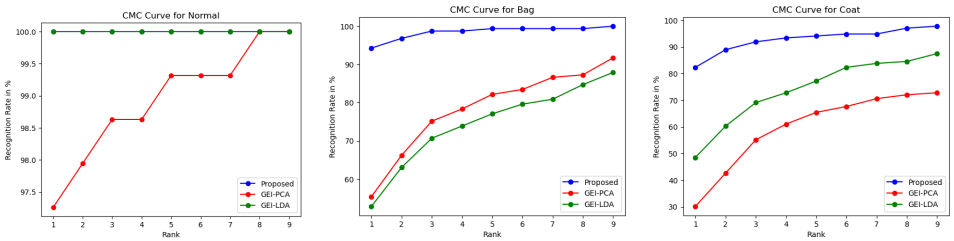
We assess the efficacy of our gait features by visualizing the intra-class and inter-class variation histograms as depicted in Figure 3a. To achieve this, we utilize the *normal* gait features derived from the trained CAE model, followed by LDA transformation. The resulting plot shows a clear separation between intra-class and inter-class histograms. This shows the informativeness and discriminative power of our features, consequently enhancing the efficiency of the classifier and its robustness to covariates. We investigate how



(a) (b) (c)

Fig. 4: ROC Curves. Comparison among GEI-PCA, GEI-LDA and Proposed methods. Here, AUC stands for "'area under the curve"'.

the dimensionality of features, denoted by b, affects gait recognition performance in two scenarios: using bottleneck features from a Convolutional Autoencoder (CAE) and enhancing these features with Linear Discriminant Analysis (LDA). Our results, displayed in Figure 3b, show that lower-dimensional features are effective in both scenarios, with LDA further improving recognition accuracy. Specifically, for normal walking conditions, we achieved 100% accuracy using 100-dimensional features processed with LDA. Recognition in more challenging conditions like carrying a bag or wearing a coat also showed



(a) (b) (c)

Fig. 5: CMC Curves. Comparison among GEI-PCA, GEI-LDA and Proposed methods.

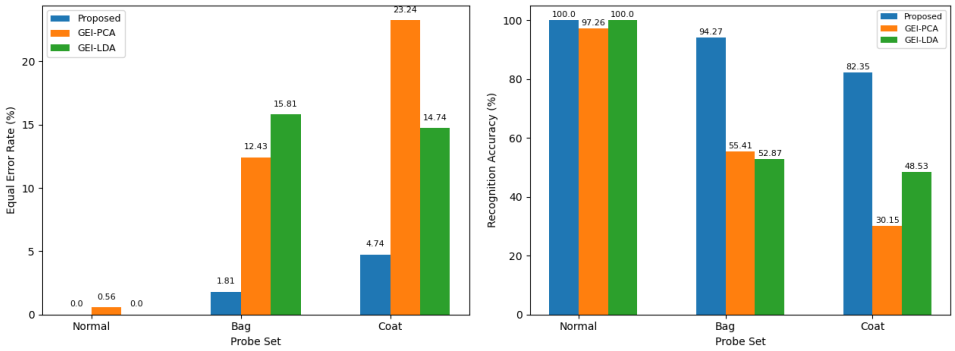
robust performance, with accuracies reaching up to 94.6% and 82.35% respectively when using LDA-enhanced features. These results confirm the effectiveness of our approach in handling variations due to clothing and carrying items, benefiting from comprehensive training across different gait scenarios. Our method also demonstrates the practical advantages of lower-dimensional feature representations, which require less storage and facilitate faster processing, ideal for real-time recognition applications.

4.3 Comparison with GEI

Given that GEIs serve as input to derive effective low-dimensional gait features from our convolutional autoencoder model, we initially compare our approach with the GEI-PCA and GEI-LDA methods. The GEI-PCA and GEI-LDA methods utilize GEIs as gait features and apply PCA and LDA, respectively for dimensionality reduction. For comparison, we consider our method utilizing CAE features followed by LDA for gait recognition as it yields superior performance. The comparison between these methodologies is depicted through ROC curves (Figure 4), CMC curves (Figure 5), EERs (Figure 6a) and rank-1 recognition accuracy (Figure 6b). We can see that our proposed method significantly outperforms GEI-PCA and GEI-LDA across scenarios involving normal, bag, and coat sequences. These results demonstrate the capability of our proposed method to extract superior gait features in lower dimensions that exhibit robustness to variations in clothing and carrying conditions.

4.4 Comparison with State-of-the-Art

To demonstrate our proposed method's efficacy, we compare it with both traditional and modern state-of-the-art gait recognition methods as outlined in Table 1. All methods utilize the CASIA-B dataset and focus on the 90° view. Our approach not only outperform traditional methods significantly but also shows superior results compared to newer deep learning techniques, particularly in challenging conditions such as wearing a coat. This underscores the effectiveness of using convolutional autoencoder-derived features over handcrafted ones and highlights the advantages of dimensionality reduction in improving gait recognition performance.



(a) (b)

Fig. 6: Comparison among GEI-PCA, GEI-LDA and Proposed methods in terms of (a) Equal error rates (%). (b) Recognition Accuracy (%). The blue bars correspond to the proposed method.

Method	**Gallery (NM #1-4)**		
	Probe: NM #5-6	**Probe: BG #1-2**	**Probe: CL #1-2**
GEI [YTT06]	97.6	52.0	32.7
GEI-PCA [HB06]	92.99	40.26	23.53
GEnI [BXG09]	98.3	80.0	33.5
Masked-GEnI [BXG10]	**100**	77.8	43.1
EnDFT [Ro15]	97.61	83.87	51.61
CGI [Wa12]	**100**	71.76	46.7
GEINet [Sh16]	93.2	59.2	55.8
GaitNet [So19]	96.9	89.1	60.2
SAE [Yu17b]	95.97	65.32	42.74
GaitGAN [Yu17a]	98.39	64.52	48.39
LSTM [Li20]	92.3	88.9	62.3
Proposed	**100**	**94.27**	**82.35**

Tab. 1: Comparison with state-of-the art gait recognition methods in terms of Rank-1 accuracy (%) using CASIA-B dataset [YTT06].

5 Conclusion

This paper presents an unsupervised learning method using a deep convolutional autoencoder trained on unlabeled GEIs to extract lower-dimensional gait features effective across different covariate conditions and utilizes softmax classifier for final recognition task. To the best of our knowledge, this is the first time, that such a CAE based unsupervised deep learning method is employed for gait recognition with excellent and promising results. The method's success is attributed to its ability to represent gait features in a lower-dimensional space that captures the most relevant gait information.

References

[BP19] Battistone, Francesco; Petrosino, Alfredo: TGLSTM: A time based graph deep learning approach to gait recognition. Pattern Recognition Letters, 126:132–138, 2019.

[BXG09] Bashir, Khalid; Xiang, Tao; Gong, Shaogang: Gait recognition using Gait Entropy Image. In: 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009). pp. 1–6, 2009.

[BXG10] Bashir, Khalid; Xiang, Tao; Gong, Shaogang: Gait Recognition without Subject Cooperation. Pattern Recogn. Lett., 31(13):2052–2060, oct 2010.

[Ch18] Chao, Hanqing; He, Yiwei; Zhang, Junping; Feng, Jianfeng: GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition. CoRR, abs/1811.06186, 2018.

[Fa20] Fan, Chao; Peng, Yunjie; Cao, Chunshui; Liu, Xu; Hou, Saihui; Chi, Jiannan; Huang, Yongzhen; Li, Qing; He, Zhiqiang: GaitPart: Temporal Part-Based Model for Gait Recognition. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14213–14221, 2020.

[FLL16] Feng, Yang; Li, Yuncheng; Luo, Jiebo: Learning effective Gait features using LSTM. In: 2016 23rd International Conference on Pattern Recognition (ICPR). pp. 325–330, 2016.

[HB06] Han, J.; Bhanu, Bir: Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(2):316–322, 2006.

[LCL11] Lam, Toby; Cheung, K.H.; Liu, James: Gait Flow Image: A Silhouette-based Gait Representation for Human Identification. Pattern Recognition, 44:973–987, 04 2011.

[Li20] Li, Xiang; Makihara, Yasushi; Xu, Chi; Yagi, Yasushi; Ren, Mingwu: Gait Recognition via Semi-supervised Disentangled Representation Learning to Identity and Covariate Features. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13306–13316, 2020.

[Ma14] Mansur, Al; Makihara, Yasushi; Muramatsu, Daigo; Yagi, Yasushi: Cross-view gait recognition using view-dependent discriminative analysis. In: IEEE International Joint Conference on Biometrics. pp. 1–8, 2014.

[Mu67] Murray., M.: Gait as a total pattern of movement. American Journal of Physical Medicine, 1(46):290–333, 1967.

[Ro15] Rokanujjaman, Md.; Islam, Md. Shariful; Hossain, Md. Altab; Islam, Md. Rezaul; Makihara, Yashushi; Yagi, Yasushi: Effective part-based gait identification using frequency-domain gait entropy features. Multimedia Tools Appl., 74(9):3099–3120, may 2015.

[Sh16] Shiraga, Kohei; Makihara, Yasushi; Muramatsu, Daigo; Echigo, Tomio; Yagi, Yasushi: GEINet: View-invariant gait recognition using a convolutional neural network. In: 2016 International Conference on Biometrics (ICB). pp. 1–8, 2016.

[SME23] Sepas-Moghaddam, Alireza; Etemad, Ali: Deep Gait Recognition: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1):264–284, 2023.

[So19] Song, Chunfeng; Huang, Yongzhen; Huang, Yan; Jia, Ning; Wang, Liang: GaitNet: An End-to-End Network for Gait Based Human Identification. Pattern Recogn., 96(C), dec 2019.

[Wa10] Wang, Jin; She, Mary; Nahavandi, Saeid; Kouzani, Abbas: A Review of Vision-Based Gait Recognition Methods for Human Identification. In: 2010 International Conference on Digital Image Computing: Techniques and Applications. pp. 320–327, 2010.

[Wa12] Wang, Chen; Zhang, Junping; Wang, Liang; Pu, Jian; Yuan, Xiaoru: Human Identification Using Temporal Information Preserving Gait Template. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11):2164–2176, 2012.

[YTT06] Yu, Shiqi; Tan, Daoliang; Tan, Tieniu: A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. In: 18th International Conference on Pattern Recognition (ICPR'06). volume 4, pp. 441–444, 2006.

[Yu17a] Yu, Shiqi; Chen, Haifeng; Reyes, Edel B. García; Poh, Norman: GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 532–539, 2017.

[Yu17b] Yu, Shiqi; Chen, Haifeng; Wang, Qing; Shen, Linlin; Huang, Yongzhen: Invariant Feature Extraction for Gait Recognition Using Only One Uniform Model. Neurocomputing, 239, 02 2017.