



# Deep Learning Based Gait Recognition Using Convolutional Neural Network in the COVID-19 Pandemic

Md Shohel Sayeed <sup>1\*</sup>, Pa Pa Min <sup>1</sup>, Md Ahsanul Bari <sup>1</sup>

<sup>1</sup> Faculty of Information Science and Technology, Multimedia University, 75450 Melaka, Malaysia.

## Abstract

Gait recognition is the behavioral biometric trait that tracks humans based on their walking motion. It has gained attention because of its non-invasive and unobtrusive behaviors and applicable to the different application area. In this paper, we target model-free gait recognition with the deep learning approach for the Muslim community in the COVID-19 pandemic. The different convolutional neural network architectures (CNN) are examined by using the spatio-temporal gait representation called Gait Energy Images (GEI). We explored both the identification and verification problems to determine the suitability of the proposed CNN frameworks. In gait recognition, the intraclass variation is larger than the inter-class variation because of the shooting view, the walking speed, the wearing condition, and so on. To tackle this challenge, the verification framework is more suitable for the 1:1 association of gait recognition. As for the verification problem, we implemented the Siamese network with the parallel CNN architecture. All the proposed methods are tested against the public gait datasets called OUISIR-LP and OUISIR-MVLP to determine the identification and verification performance in terms of recognition accuracy and error rate.

## Keywords:

Deep Learning;  
Convolutional Neural Network;  
Gait Recognition;  
COVID-19 Pandemic.

## Article History:

<b>Received:</b>	18	January	2022
<b>Revised:</b>	27	July	2022
<b>Accepted:</b>	11	August	2022
<b>Available online:</b>	16	August	2022

## 1- Introduction

Biometrics refers to the study of recognizing the identity of a human based on their physiological characteristics or behavioral characteristics. Among the behavioral biometrics, gait has drawn the attention of computer vision researchers because of its non-invasive and unobtrusive nature. Gait is the behavioral biometric that identifies and recognizes people based on their walking style. Since 2020, people have been suffering from the COVID-19 pandemic, which was caused by the deadly coronavirus. The coronavirus is transmitted from human to human or human to animal via airborne droplets [1]. People can get the infection through close contact with a person who has symptoms of the virus, including coughing and sneezing. According to WHO, some general guidelines were published, such as separating the infected patient from other family members into a single room, implementing contact and droplet precautions, airborne precautions, etc. Patients should be asked to wear a simple surgical mask and practice cough hygiene. Caregivers should be asked to wear a surgical mask when in the same room as the patient and use hand hygiene every 15–20 min [2]. Other than patients and medical professionals, normal people are suggested to practice social distancing and wearing face masks. As a result, there is a difficult way to identify people in public areas. For security and surveillance concerns, face-recognition is the most popular and efficient biometric application in society. Face recognition is not a useful practice in the new normal life style with the wearing of face masks.

In 1964, Murray et al. [3] concluded that the walking motion of each individual is unique and difficult to conceal. Early studies have shown that gait biometrics can be used to identify people and also their gender and age [4, 5]. Compared to other biometrics, gait can be easily captured and collected from a distance and without the subject's

\* **CONTACT:** shohel.sayeed@mmu.edu.my

**DOI:** <http://dx.doi.org/10.28991/ESJ-2022-06-05-012>

© 2022 by the authors. Licensee ESJ, Italy. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<https://creativecommons.org/licenses/by/4.0/>).

cooperation. Because of its applications in different research problems such as criminal investigation, surveillance monitoring, clinical analysis, gender and age estimation, and action recognition, gait recognition is perceived as the most promising biometric system for the next generation. The main objectives of this project are to create deep-learning based gait recognition methods for Ummah, which is the Arabic word for community. In the current situation with the COVID-19, the people of the Muslim community are practicing social distancing and wearing face masks. In this situation, the inability to identify the individual based on their face became a social concern. So, the proposed research will be able to tackle these problems using the gait biometric traits, which use human motion. The gait can be able to collect and detect the individual from the large distance and without the cooperation from the people. So, the main benefit of the proposed gait recognition system is that it can be used in public areas such as shopping malls, airports, banks, and so on. The proposed system will use the deep-learning framework that can be able to perform the Ummah recognition tasks effectively and efficiently.

The gait biometrics can be collected through a vision-based approach with a camera or sensor-based approach. For the vision-based approach, the camera can capture the walking subject from a distance and at different angles. And the sensor approach can collect the gait information by wearing the sensor on the subject or by placing the sensor on the floor where people walk on it [6, 7]. In this paper, we will discuss vision-based gait recognition. There are two approaches to vision-based gait recognition: model-based and model-free or appearance-based. The model-based approach focuses on body information (such as joints, limbs, and arms) to construct or model the human body or motion. The gait features are extracted from the human model (structural or 3D model). The structural-based focus is on stride parameters or motion parameters of the human body structure. The stride parameters, in this case, mean the ratios of different body parts' sizes, which can be called static body parameters(joints) or dynamic parameters (the stride length and speed) [8]. On the other hand, the model-free approach mainly focuses on the static and motion information of the walking person. This approach obtains the motion information from the gait sequences and silhouette of the gait. Normally, the gait descriptor and gait patterns are extracted from the silhouette [9]. The advantage of the model-based approach is that it is robust against covariate factors such as view, occlusion, and noise. However, it also has the limitation of only working well with high-resolution images and being computationally expensive, even though gait recognition is supposed to work and be effective at low-resolution and real-time [10]. Compared to the model-based approach, the model-free approach extracts the features from the video sequences or silhouettes regardless of the underlying structure. So, it is more suitable for outdoor applications when the parameters of the body structure are difficult to precisely estimate. Since this approach is less sensitive to the quality of the images, the computation cost is lower compared to the model-based approach, which requires high-quality images to construct the model structure [11, 12].

For the appearance-based approach, the researchers used both traditional hand-craft features extraction and deep-learned approach to learn the distinctive features from the input gait. The hard-craft features extraction needs two-stages to extract the features first and then perform the matching or classification meanwhile the deep learned approach can combine both feature extraction and classifier into one. And also, hand-craft features approaches have limitations on diverse and large data. They can't easily scalable to diverse datasets [10] and also can't handle indistinctive interclass different and large intra-class variation [13]. The deep learning approach gains attention in past few years due to their ability to perform well in a complex problem such as image recognition, signal processing and neural language processing and so on. The deep learned features extraction approach also performed well in gait recognition task [14, 15] especially the convolutional neural network [16-19]. The ability of CNN can extract high-level features from the input. The learned features from the CNN are well-captured information and robust for recognition task so, we adopted the convolutional neural network for our gait recognition task.

Like other biometric traits, gait recognition also used the pattern recognition approach which compares the gait pattern from different users. Given the query (Probe) sequence, the gait recognition system tries to match up with the highest similarity 3 score sequence within the system. To evaluate the performance of the gait biometric system, we need to consider both identification and verification. Most of the current gait researches are presented identification or classification problem. The identification applied 1: $N$  association which means the system tries to recognize the individual against all the registered gaits within the system. Since the gait recognition system applied the pattern recognition approach, it considered both inter-class (entities belong to different class/subject) and intra-class (entities belong to the same class/subject). Since the intra-class variations with different covariate factors (such as views or clothing conditions) are greater than the inter-class variation, gait recognition became a challenging problem [13]. To tackle this problem, the researchers suggested using the verification task that applied the 1:1 association [20-22]. For the verification task, the given pair of gait features (Probe and Gallery) is compared whether they belong to the same subject or not based on their similarity score with a certain threshold. The verification-based gait recognition is suitable for several applications such as criminal investigation: matching between a suspect and perpetrator or detecting the wanted person. So, this paper will present both identification and verification framework for the gait recognition system. If your paper is intended for a conference, please contact your conference editor concerning acceptable word processor formats for your particular conference. To summarize, this paper will present the following contribution.

- The gait identification framework with deep convolutional features. The convolutional neural network presented by Min et al. [24] used for the classification task to analyze the recognition accuracy.
- The gait verification framework with Siamese Architecture. The end-to-end Siamese network is proposed to use the verification task to determine the suitability and usability of deep learnt features by CNN.
- Evaluate the proposed framework with public gait benchmark datasets. We evaluate the proposed methods with OU-ISIR Large population dataset and OU-MVLP multi-view large population datasets.

## 2- Related Work

### 2-1- Gait Representation

For the appearance-based gait recognition, the accuracy and robustness of the recognition model are vitally depending on the representation of the gait sequences. Last decades, the researchers are using different ways to represent the walking sequences such as direct silhouette approach, motion-based (optical flow) approach, and binary silhouette-based gait representation. In 2005, Sarkar et al. [25] firstly presented the model-free gait recognition approach that uses the direct silhouette as the gait features. They converted the video sequences into the silhouette and represented as gait features from the generated silhouette sequences. Due to the limitation of data, the binary silhouette-based gait representation such as Gait Energy Image [26], Gait Flow Image [27], Chrono Gait Image, Gait Entropy Image [28] are introduced. The Gait 4 Energy Image (GEI) can be obtained by extracting the silhouette of the human and averaging the sequence of the silhouette [26]. GEIs become popular as the gait representation since they can capture both the temporal and spatial information and also present the human motion in a single image compare to other gait representations.

### 2-2- Gait Classification

The Appearance-based approach normally includes the phase of silhouette extraction or gait sequence representation, features learning (Feature extraction and selection), and lastly classification. The two-stage hand-craft features extraction use dimension reduction method for feature selection and classifier to do the matching. The researchers transformed different gait representations using a classical linear approach called Principle Component Analysis (PCA) and Multiple Discriminant Analysis (MDA) [26] or other dimension reduction approach such as Linear Discriminant Analysis (LDA) [28]. Lishani et al. [29] proposed to use a supervised feature selection approach Gabor filter bank-based feature extraction with Spectral Regression Kernel Discriminant Analysis (SRKDA) to reduce computational cost and allows them to handle large kernel matrices. For the classification, they presented the templet matching or distance metric classifier such as the Euclidean distance metric. Also other classifiers such as K Nearest Neighbor (KNN), Support Vector Machine (SVM), Joint Bayesian classifier, and so on [11, 26, 27]. The deep-learned approach supports the end-to-end system by combining both feature extraction and classification.

Das and Chakrabarty [14] presented deep pipeline architecture with Deep Stacked AutoEncoders (DSA) with the body points as the gait features. The gait features pass through the two layers of an autoencoder and Softmax classifier. Feng et al. [15] also proposed another deep-learned approach called "Long short-term memory" (LSTM) by using the body joint heat map with the pose estimation method. The early use of the convolutional neural network in gait recognition was described in Castro et al. [16] with the inspiration of the general action recognition in the video which was introduced by Siomonyan and Zisserman [30]. The CNN approach is later proposed by many researchers with different kind of architectures like VGG-16 [17], 3DCNN [18], ResNet [19] with different numbers of layers. Hou et al. [31] introduced the novel gait recognition architecture called Gait Lateral Network which allowed to capture both discriminative and compact features from the gait silhouette. They used the gait silhouette sequence as the unordered set and maximize the inherent features in deep convolution neural network to learn the discriminative features. In Huang et al. [32], the authors presented the 3D Local Convolution Neural network to tackle the temporal changing pattern from body parts while the subject walking. They used the backbone CNNs to improve the gait recognition model by extracting local features from local neighbourhood. So, the novel gait recognition is introduced called 3D local CNNs using the local information.

### 2-3- Gait Verification

Iwama et al. [33] introduced one-to-one gait matching using the Direct Matching (DM) approach. They used six different gait representations as gait features and compare the similarity between the two gait features (Probe and Gallery) using the Euclidean distance is adopted as the distance measurement. Rather than using the single-image gait representation as to the gait features, Shiraga et al. [20] presented deep-learned gait features extracted from CNN (convolutional neural network) as the input gait signatures. The pairs of GEIs images fed into the GEINet to extract the gait features and compute the dissimilarity score L2 distance between the Probe and Gallery GEIs. Zhang et al. [13]

proposed a novel Siamese neural network framework for verification 5 gait tasks to decrease the gap between the classification and verification problem. Inspired by the Siamese network, Tong et al. [21] proposed the verification-based pairwise gait recognition with a coupled deep neural network. Based on the extracted gait features with the CNN, the system decided on the similarity between the input GEIs pair. The gait features are normalized with the L2 norm and sent it to contrastive loss to determine the distance between the input pairs. Other verification-based gait recognition such as '2in' and '3in' with two parallel CNN and three-parallel networks are also proposed [22].

Recently Tong et al. [23] proposed the triplet-based network with CNNs for gait recognition called restrictive triplet network (RTN) with optimized restrictive triplet loss. The triplet loss and restrictive loss are combined in the loss layers as triplet restrictive loss to optimize the network. The authors argued that the performance of the RTN network is promising due to their novel optimizing strategy. Masood et al. [34] presented the gait verification model with dynamic gait features (DGF) using the subpixel motion estimation technique. The verification process is performed with the Cross-Correlation Score as a feature vector and the SVM (Support Vector machine) classifier.

### 3- Proposed Architecture

#### 3-1- Input Data

Gait Energy Image (GEI) is used as a gait representation for both identification and verification gait framework. Gait Energy Image is the Spatio-temporal gait representation that represents gait features into the single image. GEIs converted the sequence of gait silhouettes into the two-dimensional image which preserves human motion. GEIs are firstly introduced by Han and Bhanu [26] to reduce the burden of limited gait training templates. Since GEI can capture both temporal and spatial information, it became the most popular gait representation. Besides, GEIs also include information on both of the silhouette shapes and dynamic walking motion. GEIs can be obtained by extracting the silhouette of the human and averaging the sequence of silhouettes. The detail of computing the gait energy image can be found in Han and Bhanu [26].

#### 3-2- Gait Classification Framework with Convolutional Neural Network

Our proposed gait classification framework used the convolutional neural network to extract the gait features. The input data, Gait Energy Image will pass through the CNN to perform the feature learning and classification. Convolutional Neural Network (CNN) is comprised of convolutional layers, pooling layers, normalization layers as the hidden layers, fully connected layers, and also input and output layers as well. The convolutional layer is the most important layer which performs convolution operation to detect the features patterns of the input image. Pooling layers perform down-sampling the dimension of the input and fully connected layer and the Softmax layer takes in charge of classification. Figure 1 shows the proposed gait classification framework as for the architecture, we adopted the simple ten-layers CNN architecture presented by Min et al. [24]. We use all three architectures with different activation functions (ReLU, LeakyReLU, and PReLU) as documented in Min et al. [24]. The details of the network architecture are shown in Table 1.

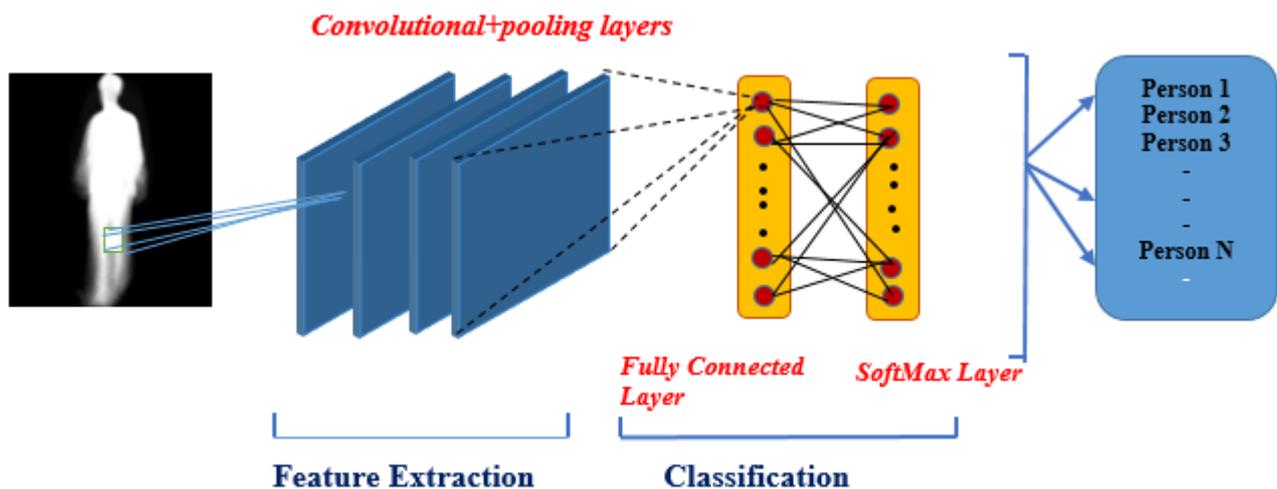


Figure 1. Gait Classification Framework

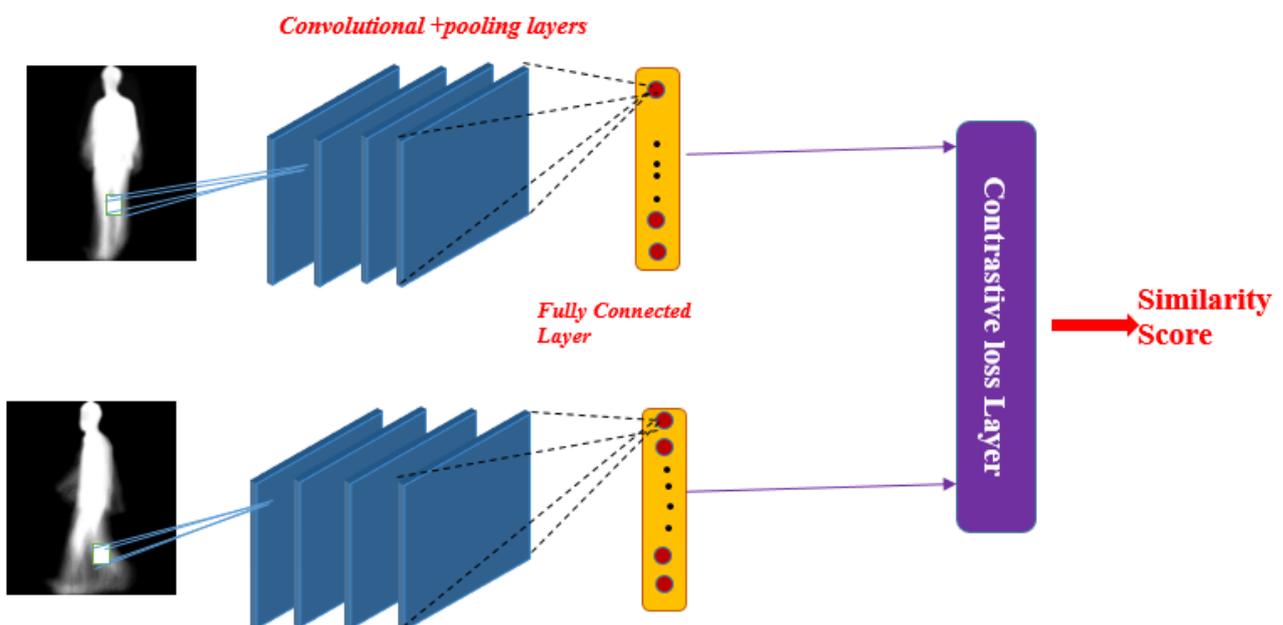
**Table 1. Details of CNN Architectures for GEI Size (240 × 240)**

Network Layer	Filter Size	No of Filters	Stride	Size of output volume
<i>Input Image</i>				
Conv_1	3 × 3	16	1	240 × 240 × 1
Pool_1 (Max)	2 × 2		2	238 × 238 × 16
Conv_2	3 × 3	32	1	119 × 119 × 16
Pool_2 (Max)	2 × 2		2	117 × 117 × 32
Conv_3	3 × 3	64	1	58 × 58 × 32
Pool_3 (Max)	2 × 2		2	56 × 56 × 64
Conv_4	3 × 3	124	1	28 × 28 × 64
Pool_4 (Max)	2 × 2		2	26 × 26 × 124
FC_1	-	-	-	13 × 13 × 124
FC_2	-	-	-	1 × 1 × 1048
(Softmax)	-	-	-	1 × 1 × 124

### 3-3- Gait Verification Framework with Siamese Networks

This section presents the verification-based gait recognition using Siamese architecture. In gait recognition, the intra-class variation is larger than the inter-class variation because of the shooting view, the walking speed, the wearing condition, and so on. To tackle this challenge, the verification framework is more suitable for the 1:1 association of gait recognition.

The Siamese architecture is firstly introduced [35] for the signature verification problem and Chopra et al. [36] later adopted Siamese CNNs to solve the face verification problem. The idea of the Siamese network is to learn the similarity between the input pairs using distance metric learning and decide whether they belong to the same subject or not. The proposed method (shown in Figure 2) is to construct the two-parallel convolutional neural network with the pairs of gait inputs. As for the gait representation, we used the Gait Energy Image (GEI) as mentioned in section (3.1). We randomly selected the pairs of GEI with the same identity as the positive pair and the GEIs with the different identities as the negative pairs. During training, the input GEIs pairs are fed into the identical CNN branches to extract the distinct gait features. The proposed Siamese architecture has the ability to pull the inputs pairs of the same identity to closer and pushed away to negative pair based on learned features representation.

**Figure 2. Gait Verification Framework**

Let, the  $x_i$  and  $x_j$  is the pair of GEIs and  $Y \in [0,1]$  is the binary label of pair  $(x_i, x_j)$ . If  $x_i$  and  $x_j$  belong to the same identity (positive pair),  $Y=1$ , and  $Y=0$  for the negative pair with different identities.

The positive pairs or negative pairs of GEIs are fed into the parallel CNN (convolutional neural network) with shared parameters. Here, the convolutional neural network has almost the same architecture as mentioned in Table 1. However, there will be some adjustments to suit Siamese architecture since the previous network designed for the classification problem. The first four convolutional layers; Conv1, Conv2, Conv3, and Conv4 will still have the same filter size and numbers of filters. Also, the pooling layers and activation function will be the same as previously mentioned. The output from the convolutional layers will be flattened and acts as the input into the fully connected layer (FC1). The features produced from the (FC1) layers will be the last features vectors for this architecture as we removed the Softmax layer (FC2) which use the number of classes as the neurons for classification. The features vectors produced from each CNN network will be computed with the Euclidean distance for the semantic similarity between the input GEIs. And then, the contrastive loss function is used to optimize the semantically similar pair to close together and dissimilar pair will be pushed away.

Each pair of GEI ( $x_i$  and  $x_j$ ) passes through the twin convolutional network with the same parameters and  $w$  denotes as the shared weight matrix that needs to be learned throughout the architecture. The features vectors encoded from the last fully connected layer;  $F_w(x_i)$  from the input  $x_i$  and  $F_w(x_j)$  from  $x_j$  is computed the Euclidean distance between them. The distance between  $x_i$  and  $x_j$  can be calculated as follow:

$$D_w(x_i, x_j) = \left\| F_w(x_i) - F_w(x_j) \right\|_2 \quad (1)$$

As we calculated the distance between the features vectors from the input pair, the distance of input pairs is compared with the predetermined threshold. If  $\left\| F_w(x_i) - F_w(x_j) \right\|_2$  is small, the  $x_i$  and  $x_j$  belongs to the same identity and  $x_i$  and  $x_j$  are different identities if  $\left\| F_w(x_i) - F_w(x_j) \right\|_2$  is large. To achieve this conclusion, the contrastive loss is used for the optimization of distance vectors. The contrastive loss can be defined as follows:

$$L(x_i, x_j, Y) = (1 - Y) \frac{1}{2} \{ \max(m - D_w, 0) \}^2 + Y \frac{1}{2} D_w^2 \quad (2)$$

where the positive number  $m$  is the margin for minimum loss distance for GEIs pairs with different identities. During the training, the output from the pairwise GEIs will be combined by the contrastive loss layer to produce the contrastive loss function. And then, the model is trained using the back-propagating with contrastive loss.

## 4- Experimental Methods

### 4-1- Data Sets

**OU-ISIR Large Population Dataset:** The OU-ISIR LP dataset is one of the largest gait databases with 4007 subjects. The details for the subjects are 2135 males and 1872 females with ages ranging from 1 to 94 years. The walking sequence for each subject is recorded twice for Probe and Gallery and captured from the four different views ( $55^\circ, 65^\circ, 75^\circ, 85^\circ$ ) [33]. The proposed method is evaluated in both the same-view and cross-view setting. Even though the OULP dataset has a smaller view variation compare to other datasets, the number of subjects makes it suitable for the evaluation of our proposed method. For both same-view setting and the cross-view setting, we followed the same protocol as Min et al. [24] to meet the comparison benchmarks. This evaluation setting used the subset of OULP datasets with 1912 subjects and 4 different views (55, 65, 75, 85). The Probe and Gallery data sets are divided according to the datasets. Equally with 956 subjects each. So each subject has eight GEI images (4 views  $\times$  2 sequences) and normalized the size into (128  $\times$  88)

**OU-ISIR MVLP:** The dataset includes 10307 subjects with 5114 males and 5193 females and ages ranging from 2 to 87 years. The OU-MVLP is not only the largest population dataset but also captured from the 14 viewing angles. Each subject is captured from different viewpoints ranging from  $0^\circ - 90^\circ$  (forward) and  $180^\circ - 270^\circ$  (backward). So, there are 14 views angles with 15-degree interval ( $0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ, 90^\circ, 180^\circ, 195^\circ, 210^\circ, 225^\circ, 240^\circ, 255^\circ, 270^\circ$ ) and two walking sequences of each subject. To summarize, the OU-MVLP dataset has the 28 walking sequences (14  $\times$  2) for the individual subject. For the experiments in OU-MVLP dataset, we followed the protocol setting of Takemura et al. [37], the dataset is divided into a nearly equal group with 5153 subjects for the training set and 5154 subjects for the testing data set. To evaluated our proposed method, we only used ( $0^\circ, 30^\circ, 60^\circ, 90^\circ$ ) since the GEIs with  $180^\circ$  view different are flipped version of the images and considers same-view pair based on perspective projection assumption. So, we 10 will only focus on the four typical views ( $0^\circ, 30^\circ, 60^\circ, 90^\circ$ ) as mentioned in the authors of the OU-MVLP datasets [37].

## 4-2- Evaluation Criteria

### 4-2-1- Gait Identification/Classification

The proposed CNN model performed the classification tasks (1: N) for the individual gait recognition. We planned to evaluate our proposed CNN architectures with all activation functions to compare the results in terms of accuracy and training speed. The three proposed CNN architectures with different non-linear activation functions namely, 1) CNN with ReLU, 2) CNN with LeakyReLU and 3) CNN with PReLU. Here, we denoted as CNN-I, CNN-II, and CNN-III for a simpler version.

To evaluate the identification performance of the proposed method, we used the correct classification rates (CCRs) which is Rank 1 identification rate with:

$$CCR = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

where, TP is True Positive, TN is True Negative, FP is False Positive, FN is False Negative.

### 4-2-2- Gait Verification

The proposed Siamese architecture performed 1:1 verification task using the convolutional neural network. We evaluated the verification performance based on different proposed CNN architectures (CNN with ReLU, CNN with LeakyReLU and CNN with PReLU) to determine the suitability of architectures in verification tasks. The details of data preparation for each dataset and evaluation results will be presented in the following section. So, to analyze the performance of the verification task (one-to-one matching), we calculated the False Acceptance Rate (FAR) and False Rejection Rate (FRR).

- FAR (False Acceptance Rate):
  - FAR also known as Type-I error denotes that the error rate when the imposter is falsely accepted as the genuine person.
  - The threshold value needs to increase to reduce the error of FAR.
- False Rejection Rate (FRR):
  - FRR is the Type-II error which is the error rate when the genuine person is falsely rejected. The increase threshold value in the FAR will make the increase in FRR error.
- Equal error rate (EER):
  - To choose the best threshold value, we need to use EER (equal error rate) which is the rate at point FAR and FRR is equal.
  - The ideal EER rate is zero and the lower the rate the better for the efficient performance in the gait verification system.
  - The proposed Siamese system is evaluated against the EER rates for all datasets.

## 5- Results and Discussion

### 5-1- Gait Classification

We evaluated our classification performance on OUISIR and OUMVLP dataset with the Correct Classification Rate as the evaluation criteria. Table 2 showed that the CCR of the OU-ISIR LP dataset in cross-view and same-view setting in three different CNN architecture. As a result, the CNN-II (CNN with LeakyReLU) outperformed than the other two. The walking sequences for the OU-ISIR LP dataset are captured from the side-view with the only 10 degrees different from each other. The accuracy of the different view with smaller angles different achieved better than the larger view. For the cross-view evaluation, we performed the analysis with the different gallery and probe views. The walking sequences for the OU-ISIR dataset are captured from the side-view with the only 10 degrees different from each other. The accuracy for the smaller angles difference is achieved better compared to larger angles. As shown in Table, the angle difference with 10 to 20 degrees obtained the accuracy of above 90 % on average and there is a slight drop in accuracy with the 30 degrees difference. However, CNN-II and CNN-III able to perform well with an average 90% CCR even in the cross-view setting.

**Table 2. Correct Classification Rate for OUISIR-LP Dataset**

Gallery View	Methods	Probe View			
		55	65	75	85
55	CNN-I	94.8	93.0	89.1	82.0
	CNN-II	96.7	94.6	92.6	88.0
	CNN-III	95.1	92.0	90.1	84.0
65	CNN-I	92.3	96.1	93.6	91.0
	CNN-II	93.6	97.4	94.7	92.8
	CNN-III	93.0	96.4	92.6	90.8
75	CNN-I	89.0	90.7	95.5	91.6
	CNN-II	92.6	93.6	97.2	93.7
	CNN-III	90.0	92.4	96.8	92.4
85	CNN-I	85.9	89.6	92.6	97.7
	CNN-III	89.2	92.2	94.5	98.1
	CNN-III	87.7	91.4	93.2	97.9

Table 3 showed the classification accuracy with the OUMVLP dataset. The OU-MVLP dataset is the larger dataset with the bigger view angles. As mentioned in the previous section, the evaluation is carried out on only four views pairs for both the same-view and cross-view settings. As shown in Table 3, the same view gait pair are performed better because of the proposed CNN network which is robust and able to learn the discriminative features for gait recognition. However, gait recognition is challenging for covariate factors such as capturing views angle differences. So, the performance for the cross-view setting with different probe 11 and gallery pairs is dropping compare to the same-view. According to the evaluation result, the gallery view with (0°) and different probe views (i.e 30°,60°,90°) is the lowest because the capturing from the (0°) is the frontal view and there is a lack of the forward and backward motion in the gait features. Since the forward and backward motions are mostly observable in 90° (side view), the accuracy for the 0-90 view pairs obtained the worse accuracy out of all pairs. The other gallery-probe pairs are angle differences with 30 to 60 degrees and it also suffers a slight drop in classification performance.

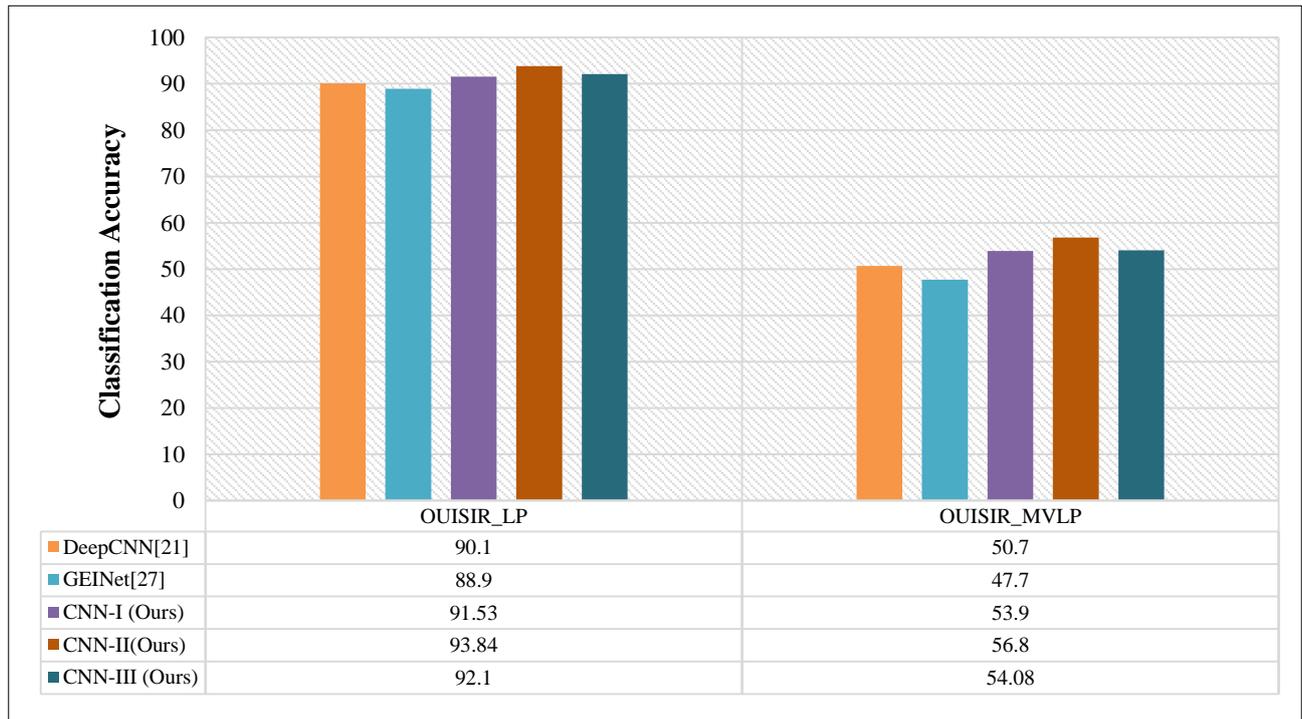
**Table 3. Correct Classification Rate for OUISIR-MVLP Dataset**

Gallery View	Methods	Probe View			
		0	30	60	90
0	CNN-I	84.8	49.0	32.1	22.0
	CNN-II	87.7	50.6	36.6	24.0
	CNN-III	85.1	47.0	33.1	21.0
30	CNN-I	56.1	90.1	58.6	31.0
	CNN-II	58.6	92.1	60.7	34.8
	CNN-III	58.0	89.1	50.6	31.8
60	CNN-I	35.4	58.1	89.5	51.2
	CNN-II	37.8	60.1	91.2	53.8
	CNN-III	36.2	57.4	90.8	52.4
90	CNN-I	20.9	30.4	60.6	92.7
	CNN-II	23.6	32.1	62.2	93.1
	CNN-III	21.8	31.8	61.2	91.6

The CCR of the three different architectures (CNN-I, CNN-II and CNN-III) is compare for each public dataset. The results show the CNN-II obtained the highest average CCR for the classification rate. The CNN-II used the Leaky ReLU activation function which is the advance activation function to solve the vanishing gradient problem and provide efficient training network. The predetermined slope in Leaky ReLU helps to avoid the dropping out of the neuron of negative values and it can obtain useful information from the features while training the network. These effects relatively improve the overall accuracy and perform better than others. The PReLU (Parametric ReLU) which was used in CNN-III obtained the second best average accuracy for all three datasets. To compare the performance of the proposed architectures, the CNN-I recorded the lowest average CCR and the CNN-II recorded the highest CCR. However, all

three architectures has shown desirable results because of the suitable convolutional neural network for gait classification tasks. The next section will present the comparison of current gait classification performance in terms of correct classification rate.

Figure 3 shown the comparison result of the proposed gait classification framework with existing methods. For the fair comparison, the three proposed CNN architectures with different activation functions (namely CNN-I, CNN-II, CNN-III) are analyzed against the current existing gait recognition methods. All the existing methods are evaluated with the different benchmark datasets; OU-ISIR LP dataset and OU-ISIR MVLP dataset. The average classification accuracy on OUISIR-LP performance is slightly dropping because of the cross-view evaluation setting even though the accuracy on the same-view recorded the desired results. In the proposed method, CNN-II is able to obtain the highest accuracy with 93.84% and, followed by the CNN-II i.e. 92.1%. The comparison method GEINet is the lowest among them, and all of the proposed architectures are able to perform better on classification tasks.



**Figure 3. Comparison of CCR with Other Methods**

As the OU-ISIR MVLP is the largest dataset with highest view angles differences. So, the average performance for this dataset is quite low compared to other datasets. The mean accuracy of the classification is in the range of 40-60 % for all the comparison methods. According to the Figure 3, the average accuracy for our proposed architectures (CNN-I, CNN-II, CNN-III) are 53.9%, 56.8% and 54.8 respectively. Even in the challenging OU-ISIR dataset, our proposed method able to achieve the highest accuracy compared to other existing methods. Our proposed methods achieved a better result because of the learnability of the convolutional neural network. The architectures and depth of our CNN able to extract the robust gait features and makes it suitable for the gait recognition tasks. And also, the advance activation function like LeakyReLU and PReLU can learn the important information on the feature without dropping the negative features like ReLU.

### 5-2- Gait Verification

For the verification tasks, we evaluated EER (equal error rate) on both datasets. The results of both datasets are documented in Tables 4 and 5. We compared our Siamese Architecture with different convolutional neural networks to analyze the verification performance and suitability of deep network in feature extraction. Since gait recognition relies on the quality of gait features, we explore how the different convolutional neural network work on feature extraction tasks. Both OUISIR-LP and OUISIR- MVLP are large multi-view datasets and so, the performance based on the cross-view setting is also considered. The OUISIR-LP datasets have the 4 different views with 10-degree angles difference and all are captured in lateral view (55°, 65°, 75°, 85°). The walking motion for four angles are considered similar and it is easier to perform the cross-view testing. According to Table 4 the EER of all three architectures is relatively small and able to achieve good verification performance. Even though the EER is increased as the angles difference becomes bigger, it is still able to obtain the desired outcome especially for CNN-II which has the least EER value compare to others.

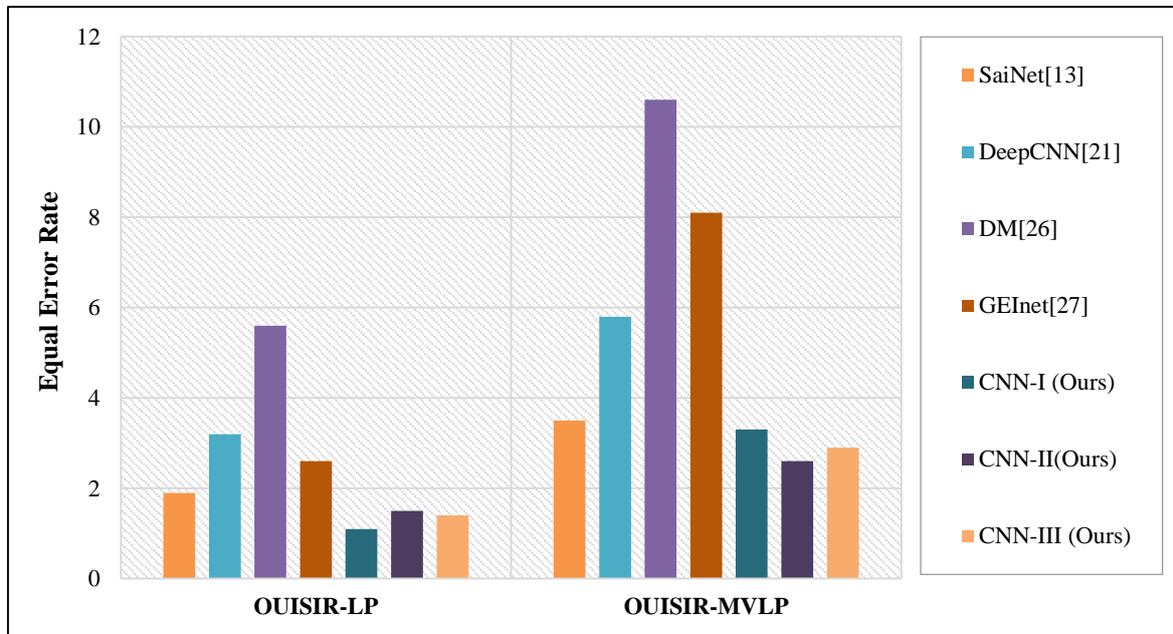
**Table 4. Equal Error Rate OUISIR-LP dataset**

CNN-I					
Probe					
Gallery	0	30	60	90	Mean
0	<b>3.1</b>	3.2	6.5	5.8	4.4
30	3.8	<b>1.6</b>	2.5	2.9	3.7
60	6.1	2.3	<b>1.3</b>	2.3	3.0
90	6.5	2.4	1.6	<b>1.3</b>	2.9
Mean	4.6	2.3	3.0	3.1	3.3
CNN-II					
Probe					
Gallery	0	30	60	90	Mean
0	<b>1.6</b>	2.9	5.6	4.8	3.7
30	3.2	<b>1.0</b>	2.0	2.4	2.1
60	5.3	1.8	<b>0.9</b>	1.8	2.3
90	5.0	2.0	1.6	<b>0.7</b>	2.3
Mean	3.7	1.9	2.5	2.4	<b>2.6</b>
CNN-III					
Probe					
Gallery	0	30	60	90	Mean
0	<b>1.7</b>	3.0	6.0	5.2	3.9
30	3.7	<b>1.2</b>	2.2	2.4	2.3
60	5.7	2.0	<b>1.1</b>	2.0	2.7
90	5.4	2.1	1.7	<b>0.9</b>	2.5
Mean	4.1	2.0	2.7	2.7	<b>2.9</b>

**Table 5. Equal Error Rate for OUISIR-MVLP dataset**

CNN-I					
Probe					
Gallery	55	65	75	85	Mean
55	<b>1.4</b>	1.5	1.7	1.9	1.6
65	1.7	<b>1.5</b>	1.6	1.8	1.6
75	1.7	1.4	<b>1.2</b>	1.3	1.4
85	2.0	1.6	1.3	<b>1.1</b>	1.5
Mean	1.7	1.5	1.4	1.5	1.5
CNN-II					
Probe					
Gallery	55	65	75	85	Mean
55	<b>0.9</b>	1.1	1.3	1.4	1.1
65	0.9	<b>1.0</b>	1.2	1.3	1.1
75	1.1	1.0	<b>0.8</b>	1.0	0.9
85	1.3	1.2	0.9	<b>0.7</b>	1.0
Mean	1.0	1.1	1.0	1.1	<b>1.1</b>
CNN-III					
Probe					
Gallery	55	65	75	85	Mean
55	<b>1.1</b>	1.3	1.4	1.6	1.4
65	1.4	<b>1.2</b>	1.3	1.5	1.4
75	1.6	1.4	<b>1.1</b>	1.2	1.3
85	1.8	1.5	1.3	<b>1.0</b>	1.4
Mean	1.5	1.4	1.3	1.3	1.4

The evaluation set for our OUISIR-MVLP datasets has the largest subjects and four different views. The evaluation set for our OUISIR-MVLP datasets has the largest subjects and 4 different views with the largest angular difference ( $0^\circ$ ,  $30^\circ$ ,  $60^\circ$ ,  $90^\circ$ ). So, it is very difficult to evaluate in the cross-view setting. As shown in Table 5, the EER comparatively becomes larger at the 90-degree angle difference. If we compare the three different CNN architectures, the CNN-II has the smallest EER values compared to the others. For a fair comparison, we also analyzed our proposed methods against other existing methods, as shown in Figure 4. The other gait verification frameworks are examined in terms of EER. According to the result, our proposed methods are outperformed for both OUISIR-LP and OUISIR-MVLP since Siamese architectures are suitable for verification tasks and the CNN inside the architectures makes it more powerful to the features extraction and learnability of the gait signatures.



**Figure 4. Comparison of EER with other methods**

For a fair comparison, the proposed verification framework will be compared against the current state-of-the-art existing methods. The methods are direct matching [33], SiaNet [13], and GEInet [20]. The direct matching (DM) used the L2 distance to compare the probe and gallery pair in the original GEI feature vectors. The SiaNet, which was proposed by Zhang et al. [13], is relatively similar to the proposed verification framework. However, the difference is the architecture of the convolutional neural network inside the Siamese Network; they used the Nearest Neighbour classifier. Finally, the GEInet is used because of the simple CNN to extract the gait features and use the Euclidean distance to compare the similarity of the probe and gallery gait features. The other gait verification frameworks are examined in terms of EER. The results of the comparison are illustrated in Figure 4.5. The results show that the direct matching approach has the highest EER rates for all the datasets. Therefore, it can be calculated that the DM approach is not suitable for verification tasks. The SiaNet is able to perform better than DM and GEInet because of the Siamese network. The proposed methods are outperformed for all datasets, even for the OUISIR-LP and OUISIR-MVLP with a large cross-view setting. The reason is that the Siamese architectures are suitable for verification tasks, which are able to pull the distance between the similar subjects closer and push the dissimilar subjects further. Another notable discovery is that the convolutional neural network in the architectures makes it more powerful for features extraction and learnability of the gait signatures because it is a suitable architecture for gait recognition and makes use of advanced activation functions, such as Leaky ReLU and PReLU.

## 6- Conclusion

Gait recognition has recently become popular research, particularly in the field of biometrics, because unlike fingerprints, retinas, and hand geometry, it does not need any direct contact with a subject. Our main objective is to ease and help the current situation where the face mask is still mandatory for the community. The gait-based features extraction and recognition efficient and real-time model in a real-time situation will be of great value to civilian applications such as banking security, airport access control, customs and immigration procedures, as well as different government and private services in a real-time environment. Thus, the project has the potential to have value and impact for Muslims and non-Muslims on society as well as the economy and nation. This study designed the gait recognition model using deep convolutional features. Due to the limitations of the hard-craft feature extraction approach in diverse datasets, deep learning has gained attention for the past few years. Deep learning can perform well with larger and more

complex datasets. Hence, the CNN was used to extract the gait features. The GEI was used for the gait representation, which captured both the temporal and spatial information and also presented the human motion in a single image.

To attain the objective of applying the deep learning technique for gait recognition, a CNN is proposed for gait feature modelling. The advantage of the deep learning approach is that it can perform feature extraction and classification in one learning feature. Hence, the simple ten-layer of CNN was proposed for the gait feature extraction and classification of individual gait. The proposed network utilized the power of different activation functions and compared three activation functions (ReLU, LeakyReLU, and PReLU) within the same networks and training settings. Gait recognition can be divided into gait classification and gait verification. The gait classification is a one-to-many problem that aims to compare the query gait signature with the gaits in the entire dataset. The gait verification deals with the one-to-one association to decide if a given pair of gaits (Probe and Gallery) belong to the same identity or otherwise. Hence, the verification framework was also presented using the Siamese architecture with a built-in convolutional neural network. According to the results, our proposed methods outperformed the current state-of-the-art methods on both classification and verification tasks. In the future, we plan to explore other gait problems such as gait retrieval or re-identification.

## 7- Declarations

### 7-1- Author Contributions

S.S., P.P.M., and A.B. contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

### 7-2- Data Availability Statement

The data presented in this study are available in the article.

### 7-3- Funding

This work was supported by Fisabilillah R&D Grant Scheme (FRDGS).

### 7-4- Institutional Review Board Statement

Not applicable.

### 7-5- Informed Consent Statement

Not applicable.

### 7-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

## 8- References

- [1] Kumar, D. (2020). Corona Virus: A Review of COVID-19. *Eurasian Journal of Medicine and Oncology*, 4(1), 8–25. doi:10.14744/ejmo.2020.51418.
- [2] Singhal, T. (2020). A review of coronavirus disease-2019 (COVID-19). *The Indian journal of paediatrics*, 87(4), 281-286. doi:10.1007/s12098-020-03263-6
- [3] Murray, M. P., Drought, A. B., & Kory, R. C. (1964). Walking patterns of normal men. *The Journal of Bone & Joint Surgery*, 46(2), 335-360. doi:10.2106/00004623-196446020-00009.
- [4] Kandel, I., Castelli, M., & Manzoni, L. (2022). Brightness as an Augmentation Technique for Image Classification. *Emerging Science Journal*, 6(4), 881-892. doi:10.28991/ESJ-2022-06-04-015.
- [5] Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9(5), 353–356. doi:10.3758/BF03337021.
- [6] Zhang, Z., Hu, M., Wang, Y. (2011). A Survey of Advances in Biometric Gait Recognition. *Biometric Recognition*. CCBR 2011. Lecture Notes in Computer Science, 7098. Springer, Berlin, Germany. [https://doi.org/10.1007/978-3-642-25449-9\\_19](https://doi.org/10.1007/978-3-642-25449-9_19).
- [7] Lv, Z., Xing, X., Wang, K., & Guan, D. (2015). Class energy image analysis for video sensor-based gait recognition: A review. *Sensors (Switzerland)*, 15(1), 932–964. doi:10.3390/s150100932.
- [8] Deng, M., Wang, C., Cheng, F., & Zeng, W. (2017). Fusion of spatial-temporal and kinematic features for gait recognition with deterministic learning. *Pattern Recognition*, 67(C), 186–200. doi:10.1016/j.patcog.2017.02.014.

- [9] Zhang, X., Sun, S., Li, C., Zhao, X., Hu, Y. (2017). DeepGait: A Learning Deep Convolutional Representation for Gait Recognition. *Biometric Recognition. CCBP 2017. Lecture Notes in Computer Science*, 10568. Springer, Cham, Switzerland. doi:10.1007/978-3-319-69923-3\_48.
- [10] Castro, F. M., Marin-Jimenez, M. J., Guil, N., Lopez-Tapia, S., & Perez De La Blanca, N. (2017). Evaluation of CNN Architectures for Gait Recognition Based on Optical Flow Maps. *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*. doi:10.23919/BIOSIG.2017.8053503.
- [11] Habib, S., Khan, I., Aladhadh, S., Islam, M., & Khan, S. (2022). External Features-Based Approach to Date Grading and Analysis with Image Processing. *Emerging Science Journal*, 6(4), 694-704. doi:10.28991/ESJ-2022-06-04-03.
- [12] Jeevan, M., Jain, N., Hanmandlu, M., & Chetty, G. (2013). Gait recognition based on gait pal and pal entropy image. *2013 IEEE International Conference on Image Processing*. doi:10.1109/ICIP.2013.6738864.
- [13] Zhang, C., Liu, W., Ma, H., & Fu, H. (2016). Siamese neural network based gait recognition for human identification. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. doi:10.1109/ICASSP.2016.7472194.
- [14] Das, D., & Chakrabarty, A. (2016). Human gait recognition using deep neural networks. *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies - ICTCS*. doi:10.1145/2905055.2905192.
- [15] Feng, Y., Li, Y., & Luo, J. (2016). Learning effective Gait features using LSTM. *2016 23<sup>rd</sup> International Conference on Pattern Recognition (ICPR)*. doi:10.1109/ICPR.2016.7899654.
- [16] Castro, F.M., Marín-Jiménez, M.J., Guil, N., Pérez de la Blanca, N. (2017). Automatic Learning of Gait Signatures for People Identification. *Advances in Computational Intelligence. IWANN 2017. Lecture Notes in Computer Science*, 10306. Springer, Cham, Switzerland. doi:10.1007/978-3-319-59147-6\_23.
- [17] Wolf, T., Babae, M., & Rigoll, G. (2016). Multi-view gait recognition using 3D convolutional neural networks. *2016 IEEE International Conference on Image Processing (ICIP)*. doi:10.1109/icip.2016.7533144.
- [18] Thapar, D., Nigam, A., Aggarwal, D., & Agarwal, P. (2018). VGR-net: A view invariant gait recognition network. *2018 IEEE 4<sup>th</sup> International Conference on Identity, Security, and Behavior Analysis (ISBA)*. doi:10.1109/ISBA.2018.8311475.
- [19] Wu, Z., Huang, Y., Wang, L., Wang, X., & Tan, T. (2017). A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2), 209–226. doi:10.1109/TPAMI.2016.2545669.
- [20] Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2016). GEINet: View-invariant gait recognition using a convolutional neural network. *2016 International Conference on Biometrics (ICB)*. doi:10.1109/ICB.2016.7550060.
- [21] Tong, S., Fu, Y., & Ling, H. (2017). Verification-based pairwise gait identification. *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. doi:10.1109/ICMEW.2017.8026299.
- [22] Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2019). On Input/Output Architectures for Convolutional Neural Network-Based Cross-View Gait Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(2), 2708–2719. doi:10.1109/TCSVT.2017.2760835.
- [23] Tong, S. bing, Fu, Y. zhuo, & Ling, H. fei. (2019). Cross-view gait recognition based on a restrictive triplet network. *Pattern Recognition Letters*, 125, 212–219. doi:10.1016/j.patrec.2019.04.010.
- [24] Min, P. P., Sayeed, S., & Ong, T. S. (2019). Gait recognition using deep convolutional features. *2019 7<sup>th</sup> International Conference on Information and Communication Technology (ICoICT)*. doi:10.1109/ICoICT.2019.8835194.
- [25] Sarkar, S., Phillips, P. J., Liu, Z., Vega, I. R., Grother, P., & Bowyer, K. W. (2005). The humanID gait challenge problem: data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2), 162–177. doi:10.1109/tpami.2005.39.
- [26] Han, J., & Bhanu, B. (2006). Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2), 316–322. doi:10.1109/TPAMI.2006.38.
- [27] Lam, T. H. W., Cheung, K. H., & Liu, J. N. K. (2011). Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition*, 44(4), 973–987. doi:10.1016/j.patcog.2010.10.011.
- [28] Bashir, K., Tao Xiang, & Shaogang Gong. (2009). Gait recognition using gait entropy image. *3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009)*. doi:10.1049/ic.2009.0230.
- [29] Lishani, A. O., Boubchir, L., Khalifa, E., & Bouridane, A. (2016). Gabor filter bank-based GEI features for human Gait recognition. *39th International Conference on Telecommunications and Signal Processing (TSP)*. doi:10.1109/TSP.2016.7760962.
- [30] Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. *Advances in neural information processing systems*, 27.

- [31] Hou, S., Cao, C., Liu, X., Huang, Y. (2020). Gait Lateral Network: Learning Discriminative and Compact Representations for Gait Recognition. *Computer Vision – ECCV 2020, Lecture Notes in Computer Science*, 12354. Springer, Cham, Switzerland. doi:10.1007/978-3-030-58545-7\_22.
- [32] Huang, Z., Xue, D., Shen, X., Tian, X., Li, H., Huang, J., & Hua, X. S. (2021). 3D Local Convolutional Neural Networks for Gait Recognition. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. doi:10.1109/ICCV48922.2021.01465.
- [33] Iwama, H., Okumura, M., Makihara, Y., & Yagi, Y. (2012). The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Transactions on Information Forensics and Security*, 7(5), 1511–1521. doi:10.1109/TIFS.2012.2204253.
- [34] Masood, H., & Farooq, H. (2021). An Appearance Invariant Gait Recognition Technique Using Dynamic Gait Features. *International Journal of Optics*, 2021, 1-15. doi:10.1155/2021/5591728.
- [35] Bromley, J., bentz, J. W., bottou, L., guyon, I., lecun, Y., moore, C., ... shah, R. (1993). Signature verification using a “siamese” time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 07(04), 669–688. doi:10.1142/s0218001493000339.
- [36] Chopra, S., Hadsell, R., & LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. doi:10.1109/CVPR.2005.202.
- [37] Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T., & Yagi, Y. (2018). Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSI Transactions on Computer Vision and Applications*, 10(1), 1-14. doi:10.1186/s41074-018-0039-6.