Computer Systems
Science and Engineering

Tech Science Press

DOI: 10.32604/csse.2023.040327
Article

A Triplet-Branch Convolutional Neural Network for Part-Based Gait Recognition

Sang-Soo Yeo[1], Seungmin Rho[2,*], Hyungjoon Kim[3], Jibran Safdar[4], Umar Zia[5] and Mehr Yahya Durrani[5]

[1]Department of Computer Engineering, Mokwon University, Daejeon, Korea
[2]Department of Industrial Security, Chung-Ang University, Seoul, 06974, Korea
[3]Department of Computer Engineering, Semyung University, Jechun-si, Korea
[4]Department of Computer Science, UET Taxila, Taxila, Pakistan
[5]Department of Computer Science, COMSATS University Islamabad, Attock Campus, Attock, Pakistan
*Corresponding Author: Seungmin Rho. Email: smrho@cau.ac.kr
Received: 14 March 2023; Accepted: 15 May 2023; Published: 28 July 2023

Abstract: Intelligent vision-based surveillance systems are designed to deal with the gigantic volume of videos captured in a particular environment to perform the interpretation of scenes in form of detection, tracking, monitoring, behavioral analysis, and retrievals. In addition to that, another evolving way of surveillance systems in a particular environment is human gait-based surveillance. In the existing research, several methodological frameworks are designed to use deep learning and traditional methods, nevertheless, the accuracies of these methods drop substantially when they are subjected to covariate conditions. These covariate variables disrupt the gait features and hence the recognition of subjects becomes difficult. To handle these issues, a region-based triplet-branch Convolutional Neural Network (CNN) is proposed in this research that is focused on different parts of the human Gait Energy Image (GEI) including the head, legs, and body separately to classify the subjects, and later on, the final identification of subjects is decided by probability-based majority voting criteria. Moreover, to enhance the feature extraction and draw the discriminative features, we have added soft attention layers on each branch to generate the soft attention maps. The proposed model is validated on the CASIA-B database and findings indicate that part-based learning through triplet-branch CNN shows good performance of 72.98% under covariate conditions as well as also outperforms single-branch CNN models.

Keywords: Vision-based surveillance systems; deep learning; triplet-branch CNN; gait recognition; covariate conditions

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

Security and surveillance of different areas seem to be quite critical in today's modern days of the digital realm to guarantee the safety of workplaces [1]. To implement this, surveillance cameras are designated at different locations for recording activities of different places, locations, and people as a monitoring system. Such multi-camera networks record abundant volumes of videos or footage at the moment, hence manually monitoring and analyzing such enormous amounts of video content takes a great deal of time and effort. To automate such processes, vision-based surveillance is designed using state-of-the-art methods. Such vision-based surveillance systems can be designed in different ways, but establishing surveillance systems based on biometrics is getting increasingly common. These biometric-based surveillance systems exploit the physical and behavioral attributes of different individuals in determining their identities [2]. More explicitly, the examples of physical attributes involve the face, iris, fingerprints, etc. On the other hand, behavioral features include voice, signature, human gait, etc. The interpretation or examination of such biological features is come under the study of biometrics to determine human identities [1].

In the aforementioned biometric traits, gait recognition is becoming more prevalent, advantageous, and conceivable [3]. Human gait recognition is an example of a biometric system whose primary goal is to identify humans based on their walking patterns [4,5]. In comparison with other biometric modalities, gait patterns can be retrieved from a larger distance and underlying persons can be recognized without their cooperation with the system [6,7]. On the other hand, face, iris, fingerprints, etc. necessitates a person to cooperate with the system, for example, the face should be at a minimal distance from the camera to be properly recognized. In addition, gait recognition continues to perform well whenever other biometric traits such as faces as well as fingerprints stay obscured. Due to such advances and strengths, human recognition through gait patterns is an innovative technology and can be more feasible to be utilized as a vision-based surveillance system deployed at different important spots including shopping malls and military regions, etc. In previous research studies, several approaches were designed by researchers to carry out recognition which was mostly categorized as model-based as well as model-free methods. More precisely, in model-based approaches, gait features are retrieved by designing the model of the human body using different geometrical shapes [8,9]. These gait features are drawn by using different parameters involving speed, step size, and stride of individuals. Likewise, model-free techniques also referred to as appearance-based approaches employ the statistical as well as spatiotemporal features from silhouette-based frames of videos. These silhouettes are acquired by performing segmentation or background subtraction from frames of videos. Such kind of appearance-based techniques is popular under the umbrella of a vision-based surveillance system designed by exploiting human gait [10].

Presently, in the era of Deep Learning (DL), different problems have been solved in several domains including surveillance systems such as gait recognition and anomaly detection [11,12], tumor detection, and skin cancer detection [13,14], planetscope nanosatellites classification [15], an intrusion detection system for edge computing [16], botnet detection and classification [17], automated weed detection algorithms utilizing UAV imagery [18,19], insider threat detection using NLP [20,21], etc. Such deep learning model applications can be integrated with intelligent systems to perform several tasks [22]. Among all of them, one of the biometric-based surveillance applications of deep learning models includes human gait-based identification frameworks. There exist different gait representations that are utilized as input of these models such as Gait Energy Image (GEI) [23], Boundary Energy Images (BEI) [24], Gait Entropy Image (GeNI) [25], Motion Silhouette Images (MSI) [26], Enhanced Gait Energy Image (EGEI) [27], Gait Flow Image (GFI) [28], Gait Information Image (GII) [29], Dynamic Gait Energy Image (DGEI) [30], and Color-Mapped Contour Gait images (CCGI) [31]. Of

all these, GEI-based [23] gait representations are the most commonly used. The advantages of GEI over silhouettes are that it offers less computational complexity and exhibits good performance. The steps involved in designing the deep learning frameworks include retrieving silhouettes from images and then calculating gait representation from silhouettes that ultimately becomes the input of the deep learning model. In existing research, different architectural configuration-assisted models are designed to enhance performance [31,32]. Moreover, the concept of transfer learning is also leveraged to perform gait recognition employing AlexNet, VGG19, and DenseNet models [33,34].

In the aforementioned research studies, the suggested deep learning models show good performance, nevertheless, when it comes to covariate variables, the performance reduces substantially. It is discussed in [12] that there exist different types of covariate variables due to which performance is affected. These covariate variables involve carrying scenarios, clothing variations, walk speed conditions, as well as occlusion (both static and dynamic), and viewpoint variations. To more clearly explained the problem, if the underlying deep learning model is trained on GEI images of different persons in the normal walk, afterward whenever the model is put to test mode, then if the person comes under different clothing and carrying conditions such as wearing long coats or carrying heavy suitcases, their gait patterns are affected resulting in a more challenging scenario for the model to properly recognize the person.

More precisely, clothing variations cause difficulties in obtaining good accuracy because they impact the feature set [35]. For example, some important gait patterns are hidden in a long coat worn by a person since the coat hides the regions of body parts [3]. Likewise, speed and occlusion conditions are also challenging factors to consider while designing the gait recognition frameworks [1,36,37]. It is observed from the above discussion that in appearance-based approaches to human gait recognition, covariate factors remain an important challenge in the gait-based identification system. In a nutshell, the major goal of this work is to develop an improved deep-learning model in terms of performance under covariate situations. One important research question is what if that deep learning model is focused on certain regions to determine the identity since it might be the case that all of the regions of GEI images are not affected due to covariate factors. Secondly, in some recent studies [12,38–42] in which deep learning models are designed and are trained, and tested on the same walking conditions but in real circumstances, the walking condition during testing is not known in advance. Hence, there is a need to investigate model performance when the covariate conditions of test time are not known in advance. In addition, several deep learning models have been proposed in past research as well, however when it comes to covariate conditions then the performance of the model drops [12]. Therefore, it is logically deduced from the results of existing studies that covariate condition is the major challenge or problem that limits the performance of gait recognition. Hence, to handle this problem, a triplet branch CNN model is proposed in this research study for human identification based on gait patterns. More precisely, in this research there is extraction of three selective regions i.e., head, body, and legs, and utilize them as input to separate branches of the CNN model. The rationale of selecting three parts separately is to handle covariate conditions e.g., these covariate conditions affect certain regions of images, for instance, if a subject wears a hat, then in this case head region is more affected, but identification is possible through the body and legs regions. Moreover, each branch operates on separate regions of interest and extracts the features which are also further refined using soft-attention layers to generate the attention maps from activation or feature maps. Each branch first separately identifies the subject and later on final subject identification is done by probability-based criteria, i.e., the label of subjects decided by each branch with the highest probability becomes the final label of that particular subject. In this research, the evaluation of the proposed model is performed

on the CASIA-B gait dataset and the findings show good improvement under clothing and carrying conditions.

- A triplet-branch CNN model is suggested for region-wise identification of humans based on gait features to handle covariate conditions.
- To empower feature learning, soft attention layers are added to extract more refined features.
- Experiments indicate that the proposed model shows good results in comparison with baseline methods.

The rest of the paper is organized as Section II provides the literature review, section III provides the proposed methodology, and Section IV describes the results with analysis followed by a conclusion and references. Fig. 1 shows several example frames of video and their associated silhouettes from the CASIA-B gait dataset.

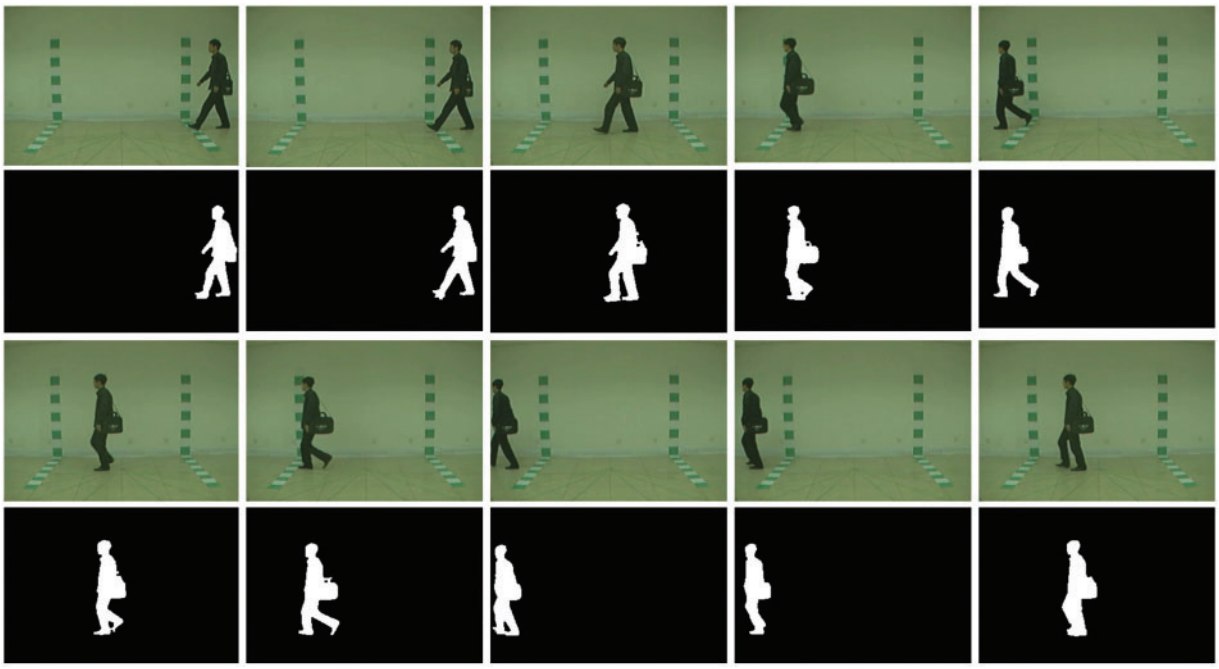

Figure 1: Frames and silhouettes of videos from the CASIA-B dataset

2 Literature Review

To effectively recognize persons based on their gait patterns, different researchers suggested several methods including model-based and model-free techniques. In each kind, distinct machine and deep learning algorithms are proposed. The differences in existing studies also exist in terms of the approach that is used to acquire the gait data, for example, floor sensors, accelerometers, radar, or image-based data using cameras. All of such research studies attempt to handle and reduce the impact of covariate variables which is a major issue in human gait recognition. Particularly, machine learning methods as well as rule-based techniques are suggested in the work of Kececi et al. [43] to identify humans based on their gait patterns. The wearable accelerometer and gyroscope are employed to retrieve the gait patterns which then become the input of Random forests, Decision trees, and multi-layer perceptron. Their suggested technique provides good results of about 99% accuracy. Following on, the Artificial

Neural Networks (ANN) based method is also designed in the work of Bari et al. [44]. In their work, the gait pattern of individuals is extracted by employing Kinect sensors followed by designing a novel feature set. That feature includes joint relative cosine dissimilarity and joint relative triangle area and attained good results in form of recall and F1-score values. Such sensor-based techniques are good in terms of performance, nevertheless, they are not feasible and practically deployable due to the high cost of sensors. More specifically, such methods necessitate an individual to either wear a sensor or require a floor sensor to obtain human gait patterns for their identification [3].

In contrast to these approaches, vision-based surveillance approaches are more feasible and commonly used. These methods do not require human cooperation as well as less costly in comparison with sensors-based methods. It is reported in existing research that human gait can be obtained from low-resolution cameras even if they are at a larger distance from humans [3]. In this category, i.e., vision-based gait recognition, several methods are suggested. For instance, a model-based framework designed by Liao et al. [45] by exploiting the 3D human pose features followed by utilizing convolutional neural networks (CNNs). Under the challenging situations of camera viewpoint and clothing variations, their suggested technique exhibits excellent results. To enhance the performance of human gait recognition, graph convolutional neural networks are also suggested in the work of Teepe et al. [46]. Their proposed method is referred to as skeleton-based gait recognition and the results are obtained by performing the experiments on CASIA-B and OUMVLP-Pose gait datasets and achieving benchmarking results. An et al. [47] also suggested the 3D pose features since they are less influenced by covariate variables, particularly viewpoint variations. Moreover, they have also extracted the spatiotemporal features by fusion of CNN-LSTM deep learning models to increase the recognition rate. The findings of the study also show that 3D pose features work better than 2D pose features. In addition, the graph neural networks are suggested in the work of Zheng et al. [48] with a modification of adding dual branches. The objective of their method is to lessen the camera view interferences in conjunction with enabling spatiotemporal relationships. In order to predict the skeleton's camera viewpoints, a separate model namely an angle estimator is proposed. Furthermore, they have tested their technique on a large gait database i.e., the CASIA-B gait dataset. Such model-based techniques show good performance and are effective enough to overcome the challenging scenarios of covariate variables. But on the other hand, they need better-resolution videos to appropriately compute the geometric model of individuals, hence they are computationally demanding [1].

Furthermore, there also exists some set of approaches referred to as appearance-based techniques, in which the algorithms are designed to work over the human silhouette images, i.e., binary images to obtain their gait features. These images are also less dedicated to texture and color variations. It is observed from studies that algorithms that work on such silhouette images are less complex than those of model-based techniques. But such silhouette-based methods also have some limitations i.e., the performance of the model substantially reduces when it comes to different covariate conditions such as camera viewpoints, clothing, as well as carrying and speed conditions. To overcome such limitations, and to build cost-effective as well as good performance gait recognition systems for automated surveillance, different frameworks are designed. For example, a 3D CNN is designed in the work of Gul et al. [49] whose input is the GEI rather than binary silhouette images. This GEI-based gait representation is more compact and extracts body motion and spatial shape-based features. OULP and CASIA-B are the two datasets on which the experimentation is performed in their research. The findings show that their proposed method is effective in handling covariate conditions. Likewise, a novel deep-learning model for human gait recognition is also suggested by Arshad et al. [42] in which VGG19 and AlexNet are combined to draw the features and then feature selection is performed using entropy and skewness methods. In the end, the final feature set is utilized

to perform the recognition of humans using refined gait features. AVAMVG gait, CASIA A, B, and C are the datasets on which they have tested their technique to indicate their findings and it is observed that the fusion of both models leads to good outcomes. Subsequently, to overcome the problem of covariate conditions particularly clothing and carrying scenarios, Alsaggaf et al. [50] suggested the use of Generative Adversarial networks (GANs) type namely CGANs i.e., cycle consistent GANs. They have performed the translation of GEI images that is disrupted from covariate factors to normal GEIs. The working of CCGANs involved the cycle and they have trained the model in an unsupervised manner to reconstruct the GEIs. The experimentation on the CASIA-B gait dataset shows that their method attains good accuracy under challenging circumstances. Similarly, a novel CNN model capable of handling different variations is designed in the work of Alotaibi et al. [32]. Their deep learning model is more generalizable and has been trained on GEI images without performing augmentations and has attained competitive accuracies. A joint-learning-based deep learning model involving both the identification and verification process is proposed by Li et al. [51]. Their proposed model builds upon joint intensity transformer networks that are good enough to handle covariate factors. For both verification and identification, distinct loss functions are utilized including triplet and contrastive loss. The findings of experimentation show that their suggested method generates good results whenever it comes to challenging clothing and carrying conditions. Moreover, the Histogram of oriented gradients (HOG) as well as the Zernike moment with random transform-based methods are used in the work of Semwal et al. [52] to draw the gait features followed by the classification of approaches using machine learning methods. The features are extracted from the GEIs which are then utilized to perform human recognition through gait.

Furthermore, there also exists some research studies in which particular gait representations are suggested to reduce the impact of covariate variables which ultimately helps in improving performance. In this context, a particular kind of image referred to as Skeleton Gait Energy Image (SGEI) is suggested in the work of Yao et al. [53] which is robust to challenging scenarios. In addition, a multi-stage deep learning model is also designed along with a fusion of two types of gait representation i.e., GEI and SGEI. This integration is accomplished to minimize the limitation of appearance-based gait patterns. In a challenging clothing scenario, their suggested technique exhibits superior results. Similarly, another special category of gait representation is suggested and referred to as Gait Entropy Image (GeNI) in the research study of Bashir et al. [54]. These images show the pixel's randomness during the complete cycle of an individual's gait. Their proposed gait representation also works well in improving the performance of gait recognition under covariate conditions. Likewise, to handle occlusion conditions which are also considered as one of the covariate factors several methods have been devised [55–58].

Moreover, some recent literature on gait recognition frameworks also involves the use of deep learning models such as Mogan et al. [38] suggested a VGG16-MLP-based deep learning model by employing GEI-based gait representation to perform gait recognition. Their approach is based on the transfer learning concept in which the first pre-trained VGG16 model is used to extract features and later on fined tuned to perform classification and achieved good results. Similarly, in another work by Mogan et al. [39] they employed a DenseNet variant namely DenseNet-201 with multi-layer perceptron. This pre-trained DenseNet model draws the features from GEI and later on the associations among these latent features and associated class is learned using the MLP model. Ambika et al. [40] also employ DenseNet to handle view-based conditions in which the final layer is modified to include softmax activation with 10 hidden units to perform 10 subject identification. Mehmood et al. [41] employs DenseNet-201 for feature extraction followed by feature selection to

reduce the feature vector sizes, and in the last classification is performed using One against All Multi Support Vector Machine (OAMSVM).

A hybrid model by combining both VGG-19 and AlexNet is also designed in [42] to perform feature extraction for efficient gait recognition in addition to entropy and skewness-based features. In all of these studies, the training and testing scenarios are built on the same covariate condition, hence, the performance of the model is hidden when both sets are built under different covariate conditions. Therefore, the main goal of this study is to build a deep learning model that is good to identify the subject's conditions when it is trained on normal walk scenarios and validated on the subject when they appear with coats or while carrying bags. Over these recent studies, this study attempts to answer a question, what if the model is learned to recognize subjects when they appear with a normal walk and put on test mode to recognize persons when they appear with different covariate conditions. Secondly, these covariate conditions are not the only coats and bags, the person can come with any condition in real-time scenarios, hence, when the model is trained on all these conditions then it fails to recognize a person when it comes with new conditions. As an intriguing scenario, this research construct a more challenging setting in which the model is evaluated with unknown covariate conditions. The suggested approach is created as a region-based method to deal with this problem.

3 Methodology

In this section, there is description of the proposed work in detail. More precisely, the proposed model starts from preprocessing stages that is computing the gait representations namely GEI images followed by a triplet-branch CNN model with soft attention layers to perform human gait recognition as an identification task. Fig. 2 depicts the overall pictorial representation of the proposed model and details of each step is described below:

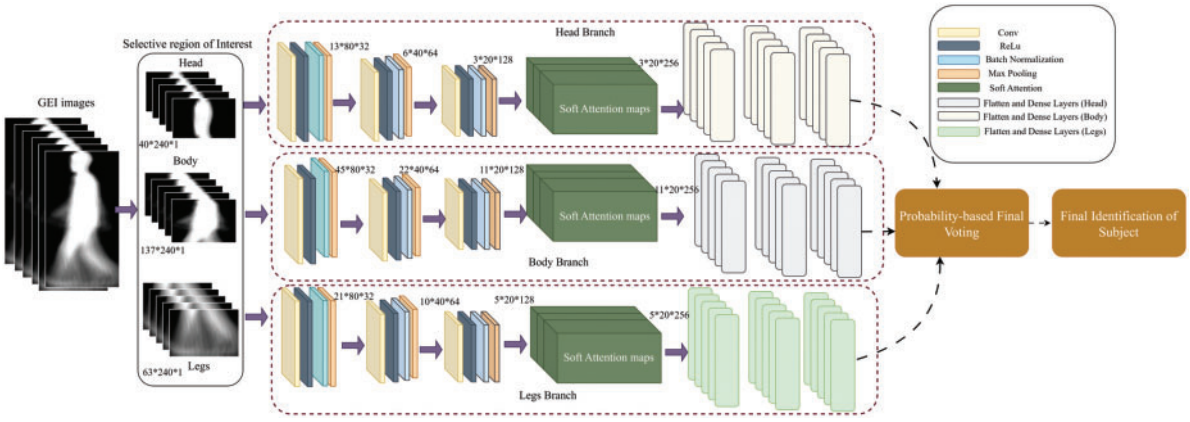

Figure 2: Architecture of suggested Triplet-branch CNN model

3.1 Gait Representations

In existing studies, there exist several gait representations in appearance-based approaches, however, the human GEI is one of the most commonly used gait representations. In this research, first step is the collection of the human gait videos under different walking conditions such as normal walking, walking while carrying heavy bags, and walking while wearing coats. The videos are processed to compute binary silhouette images frame by frame followed by computing the GEI-based gait

representation. More precisely, the mathematical equation used to compute the GEI images is given below:

$$GEI_{image} = G(x, y) = \frac{1}{T} \sum_{t=1}^{T} I(x, y, t) \tag{1}$$

In the above Eq. (1), T is the total number of aligned binary silhouettes at time t having x and y coordinates denoted as$I(x, y, t)$. The result of this equation is the GEI image denoted as GEI_{image} and is computed for every subject in the database. The high pixel values in the GEI denote the static part of the human body while the bottom regions where the values of pixels are lower represent the motion-reliant information about the subject. These gait representations are more compact than silhouette images and can be utilized as input to the deep learning model.

3.2 Region Extraction

In the problem of gait recognition, challenges of covariate conditions are the main factor to hinder the performance of the deep learning model. For instance, as shown in Fig. 3a, the GEI image in which a person carries a bag is distinct from those of normal walking. However, it is also observed that only some of the regions are disrupted due to covariate conditions. Hence, to make learning from different regions this article proposed triple-branch CNN in which each branch specifically focused on a separate part. To enable this kind of learning, the original GEI images is divided into three parts i.e., head, body, and legs. More precisely, a normalized and center-aligned GEI image of dimension $240 \times 240 \times 1$ in which the head part is from 0 to 40 in height, the body is from 40 to 177 and the legs are from 177 to 240 as shown in Fig. 3b. After extracting regions, input is given in form of these regions of interest to separate branches of the CNN model.

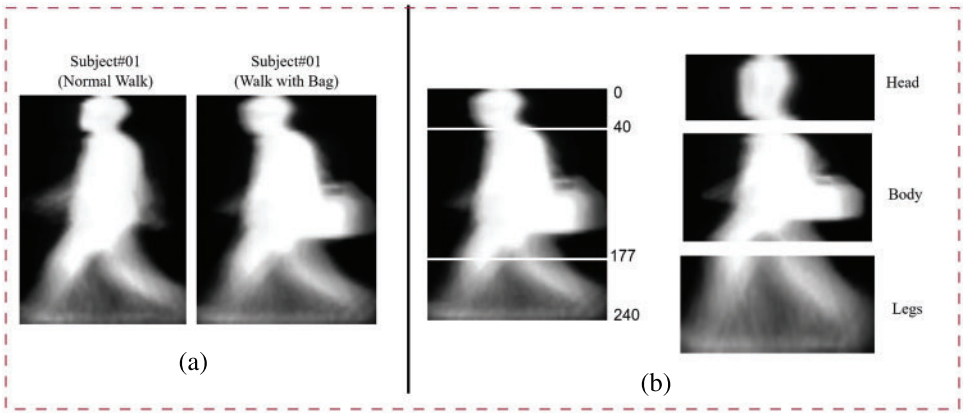

Figure 3: Region Extraction from GEI images

3.3 Proposed Triplet-Branch Convolutional Neural Network

After preparing the gait representation and extracting regions of interest, a triplet-branch CNN model is deisgned. Convolution Neural Networks (CNNs) are one of the most popular and state-of-the-art algorithms for solving various computer vision problems. CNN model starts learning in a hierarchy by performing two operations in a repeated manner namely convolution and pooling in a manner just like the working of basic cells in the visual cortex of the human brain. This kind of automated feature extraction made them applicable to various domains and saves the overload of traditional feature extraction approaches from images such as Histogram of Oriented Gradients

(HOG), Local Binary Patterns (LBP), etc. CNN models are considered to be the most popular algorithms to solve computer vision challenges in a variety of ways including image classification [59], object detection and localization [60], as well as semantic, instance, and panoptic segmentation [61]. CNN-based feature extraction is more generalizable and is end-to-end than conventional machine learning approaches, particularly in the case of vision data. In this research, the approach of CNNs is suggested, however, with a modification to handle the challenge of covariate conditions. More precisely, a different region of interests of GEIs of size 40 × 240 (Head), 63 × 240 (legs), and 177 × 240 (body) are inputted to each branch of triplet-CNN. Following on, each branch of triplet-CNN comprises convolutional and max-pool layers to extract the features of each part.

More precisely, a collection of convolutional filters is convolved over the image denoted as $\{W_k\}_{k\in K}$ to produce another stack of tensors namely the activation maps denoted as $\{H_j\}_{j\in J}$. A tracking table CT that maintains the associations among a collection of GEI image-parts i, a kernel K, and the outcome j. On the resulting activation maps, an activation function is applied. In this study, "ReLu" is employed as an activation function. Moreover, the relationship built through convolutional layers is given in Eq. (2):

$$h_j(x) = \sum_{i,k\in CT_{i,j,k}} (f_i * w_k)(x) \tag{2}$$

In the above Eq. (2), a convolution layer is denoted by $*$ that is applied over the GEI image parts i.e., legs, head, and body parts. In the proposed model, there is a total of three convolution layers having a total number of filters 32, 64, and 128 and a kernel size of dimension 3 × 3 having padding parameter set to "same". After every convolution layer followed by the activation function "ReLu", batch normalization and max-pool layers are added. With the help of this batch normalization layer, training becomes faster due to normalizing input utilizing mean as well as variance. More precisely, on each branch of CNN, a batch of samples of dimension D in the form of $X \in R^{N*D}$ i.e., matrix is inputted to the batch normalization layer wherein each instance is denoted by x_i. Eq. (3) is employed to mathematically present this concept.

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{3}$$

In the above Eq. (3), the symbols $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$ represent the mean and variance that are computed using Eqs. (4) and (5).

$$\mu = \frac{1}{N}\sum_i x_i \tag{4}$$

$$\sigma^2 = \frac{1}{N}\sum_i (x_i - \mu)^2 \tag{5}$$

Following, these batch normalization layers, a max-pool layer of window size 2 × 2 is added to downsample each GEI part. The outcome of this layer is to select the highest value from a given window and it is mathematically defined below:

$$y_{k,w}{}^i = max_{0\le a,b<p}(x_{i_{k\times p+a,w\times p+b}}) \tag{6}$$

In the above Eq. (6), the highest value is selected from a region $p \times p$, and provided to a neuron $y_{k,w}{}^i$ that is present at the location (k, w) on i^{th} activation maps. These consecutive sets of layers i.e., *Convolution* → *Relu* → *BatchNormalization* → *Maxpool* are added three times to extract the features from each GEI part. Following on, soft attention layers are added to refine the features.

3.3.1 Soft Attention

After extracting the features from each branch concentrated on a specific part of the human GEI is refined using soft attention layers [62,63]. The input of this layer is the feature extracted from preceding branches comprising convolution and max-pool layers. As described in some research studies, soft attention is inputted with the feature tensor ($\boldsymbol{t}$) resulting from previous layers as given in the below equation:

$$f_{sa} = \gamma t\left(\sum_{k=1}^{K} softmax(W_k * t)\right) \tag{7}$$

In the above Eq. (7), feature tensor $t \in \mathbb{R}^{h\times w\times d}$ is provided to the 3D convolutions [64] having weights $W_k \in \mathbb{R}^{h\times w\times d\times K}$ in which K indicates a number of weights i.e., 3D. By employing softmax the outcome of this operation of convolution is normalized to produce K attention maps. These attention maps are fused to generate standalone activation maps. This resulting attention map is acted as a weighting function α. To attentively preserve the important features, this α is multiplied with the input tensor t followed by scaling with a learnable scaling parameter γ. In the end, the attentively scaled features f_{sa} is fused with the original input tensor (t) to generate the final features. When training of the triplet-CNN starts, a value of γ is set to 0.01, so that the model progressively learns the amount of attention necessary by triplet CNN. A pictorial representation of this module is depicted in Fig. 4. From Fig. 4, it is clear that feature or activation maps resulting from different regions of each GEI that are inputted to separate branches of CNN are then passed to soft attention layers to generate attention maps. These attention maps comprise the refined information or features from every part of GEI. This layer refined the activations to certain important features that are more discriminable to identify the subject of the person using gait patterns from every specific region. Hence, the proposed approach focuses on more precise and refined features. It first attempts to extract features from separate parts and then refines those particular parts' features to generate a more streamlined feature.

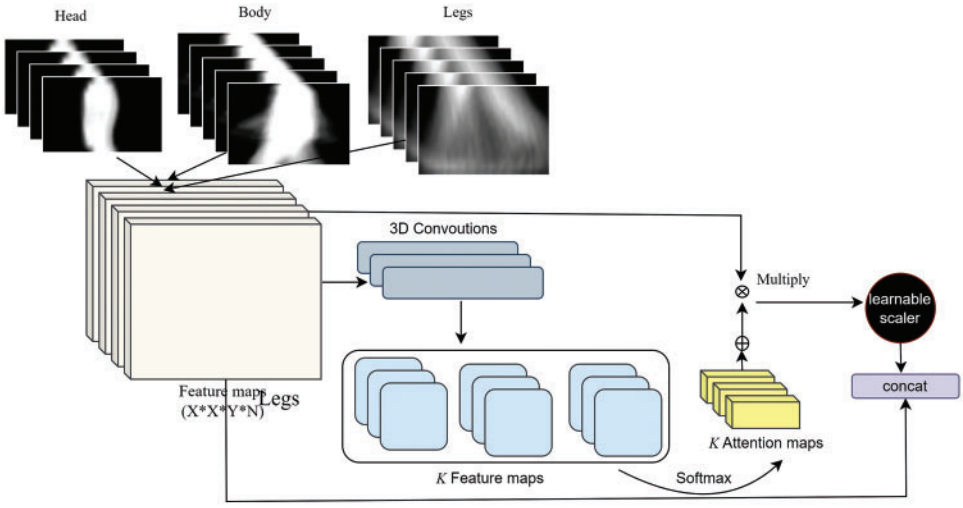

Figure 4: Soft attention layer in triplet CNN model

3.3.2 Classification Layers

As mentioned in the above sections each branch is focused on a separate part of GEI to extract features and later on each branch has separate classification layers to classify each part of the GEI.

For instance, each GEI part (i.e., head, legs, and body) is focused to extract features in a separate branch of triplet-CNN, and later on, each branch classifies each part belongs to which subject. For each branch, initially flatten layers are added to flatten the feature vectors, and later on, fully connected layers are added with hidden units equal to the number of subjects available in the database along with soft-max activations. After classification is done on each branch part by part, then the final label of the subject is decided based on prediction probabilities from each branch. The classification of a human subject with high confidence is finally selected as the label of a person. More precisely, these probabilities are accessed from the softmax activation of each branch. Each branch assesses each part of GEI in terms of how confident it is that this part belongs to which subject. The branch having high confidence in classifying that subject is utilized to decide the final label and identification of a subject. Furthermore, during training the main model is added combining all three branches to build up triplet CNN. This main model combines the loss and accuracies of each branch. Moreover, each branch has its loss function, optimizers, and all other parameter settings and trains separately. More precisely, each branch employs the Adam optimizer, the loss function is "categorical cross-entropy", and the learning rate is initially set to 0.001 which is then decayed by *learning_rate/epochs*, where epochs are set to 200. These hyperparameters are tuned to their best values after performing simulations with different settings of these parameters. Moreover, the computational complexity in terms of total trainable parameters is about 15.5 million and it takes about 30 mins to train on the CASIA-B gait dataset on the Google Colab with 12 GB NVIDIA Tesla K80 GPU.

4 Experiments and Discussions

In this section, there is discussion of the dataset which used in this research as well as the results with proper analysis. Moreover, the evaluation metrics used to evaluate the performance of the proposed model are also discussed.

4.1 Dataset

To assess the performance of the proposed model, CASIA gait dataset [65] is employed which is designed by the Chinese Academy of Sciences. This dataset is comprised of a total of three parts such as CASIA-A, B, and CASIA-C datasets. There are a total of ten video sequences is available in the dataset out of which six videos belong to normal walk sequences, two videos belong to walking sequences in which a person carries a bag, and two videos belong to walking sequences in which a person walks while wearing coats. Particularly, the normal walk sequences are denoted by "nm", carrying conditions i.e., bags denoted by "bg", while clothing conditions i.e., coats are denoted by "cl". Moreover, data from 124 subjects are included in the database.

4.2 Evaluation Metrics

The evaluation metrics used to evaluate the performance of the proposed model as an identification model includes accuracy, precision, recall, and F1-score. The following is the equation used to compute these metrics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{11}$$

In the above Eqs. (8)–(11), TN denotes the true negative, TP denotes the true positives, FN denotes the false negatives while the FP denotes the false positives.

4.3 Experiments, Discussions, and Comparisons

To evaluate the performance of the proposed model, an experimentation is performed on the CASIA-B gait dataset. More precisely, in the first step videos of different subjects are preprocessed to acquire gait representation namely GEI images. Following on, the given GEI images is partitioned into different regions i.e., Head, body, and legs. These regions eventually become the input of the triplet CNN model to perform the gait recognition. Each branch of the model will result in predictions; however, a probability-based voting is performed to decide the final label of the subject. The model is evaluated by dividing the dataset according to different walking conditions. The training set includes part-based GEI images of normal walking sequences (nm-01 to nm-04), while the test set includes the remaining normal walk sequences (nm-05 to nm-06), carrying sequences of bags (bg-01 to bg-02), and clothing sequences of coats (cl-01 to cl-02). The results of the proposed triplet CNN in terms of accuracy, recall, precision, and F1 score are given in Table 1. More precisely, the first row of Table 1 shows results with normal walk conditions, while the second and third row shows results with bags and coats-based covariate conditions. Simulations are ran several times and then report the average results along with standard deviations as shown in Table 1. Moreover, if analysis of the results is performed in terms of different walking conditions, then as shown in Table 1, it is observed that when train and test scenarios are similar i.e., normal walking then the accuracy of the gait recognition model is high which is about 96.290%. Similarly, the precision, recall, and F1-score are about 95.908%, 96.290%, and 93.544% respectively. On the other hand, when experiments are performed in different walking scenarios in which covariate variables are involved i.e., carrying condition of bags or clothing conditions of coats, then the proposed model still performs well with accuracy, precision, recall, and F-Score is about 57.408%,56.44%,57.258%, and 57.258% respectively in carrying conditions of bags.

Table 1: Results of triplet CNN in different walking scenarios

No	Train set	Test set	Accuracy	Precision	Recall	F-score
01	Normal walk	Normal walk	96.290 ± 1.255	95.908 ± 1.6093	96.290 ± 1.2558	93.544 ± 1.533
02	Normal walk	Walk with bags	57.408 ± 3.0485	56.44 ± 2.7200	57.258 ± 3.149	57.258 ± 3.149
03	Normal walk	Walk with Coats	65.263 ± 1.47092	66.4098 ± 0.6461	65.080 ± 1.5248	62.77 ± 0.7764
		Mean	72.987 ± 1.9248	72.987 ± 1.9248	72.919 ± 1.6584	71.19067 ± 1.819

Likewise, if the performance of the model is analyzed in terms of clothing conditions of coats, then the proposed model achieves accuracy, precision, recall and F1-score is about 65.263%, 66.4098%, 65.080%, and 62.77%. It is observed from the results that in the presence of strict challenges of covariate conditions the performance of the proposed model is encouraging. Since the presence of these covariate variables disrupts the gait patterns or features. However, the motivation behind the proposed triplet CNN is that it looks into different regions of GEI which might or might not be affected by covariate factors. In addition, the average results (mean results of all walking conditions) is also reported. Hence, overall, the proposed triplet-CNN model achieves a mean accuracy, precision, recall, and F1-score of 72.987%, 72.987%, 72.919%, and 71.19067%. Furthermore, the training accuracy and loss values are also plotted for each branch of CNN separately. More precisely, the head, legs, and body branches of CNN's model loss and accuracy plots for normal walk conditions are depicted in Fig. 5. It is observed

from the Figures that the proposed model shows good results in terms of convergence performance over a set of epochs. Moreover, simulations ran several times whose results are also depicted in Figs. 5–8 respectively. These graphs indicate that the proposed model shows good convergence and reaches optimal values over several simulations.

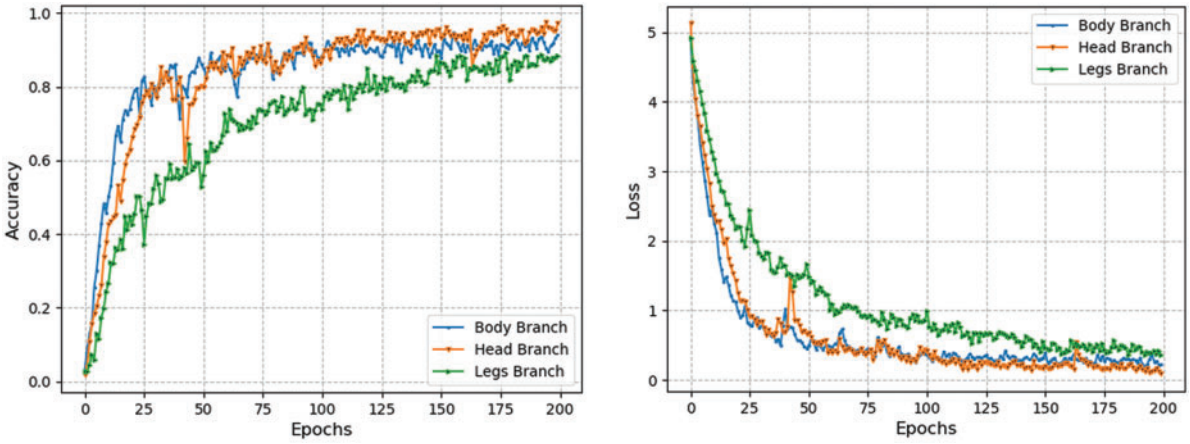

Figure 5: Accuracy and loss curves of different branches of triplet-CNN under normal walk conditions

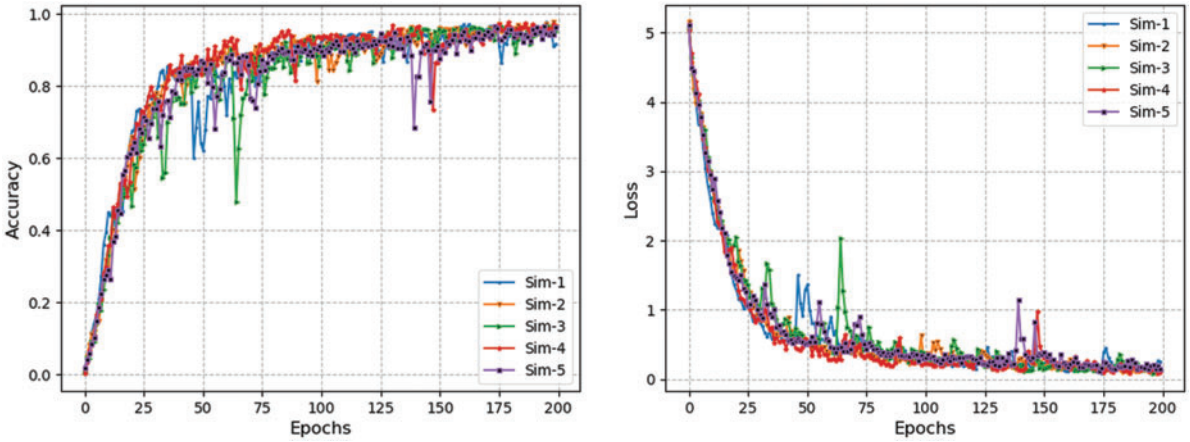

Figure 6: Accuracy and loss curves of "head branch" of triplet-CNN over different simulations

In the existing studies, several models and approaches are designed to perform gait recognition. The main objective of such methods is to increase and improve the performance of gait recognition frameworks when they are subjected to several covariate conditions including carrying, clothing, occlusions, and speed-related challenges. Because this covariate variable affects the image-specific regions resulting in the loss of discriminative subject-related features. Hence, when there is less discrimination in features and if they are not sufficient to properly indicate the subject then the performance of the model drops. In this research study, a triplet-branch CNN is proposed which focused on each region of GEI separately to classify the underlying subjects.

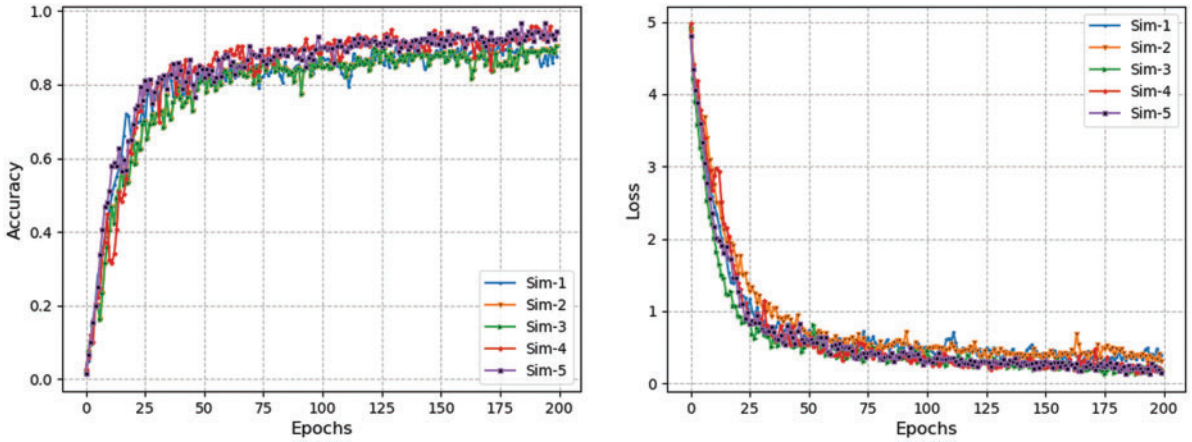

Figure 7: Accuracy and loss curves of "body branch" of triplet-CNN over different simulations

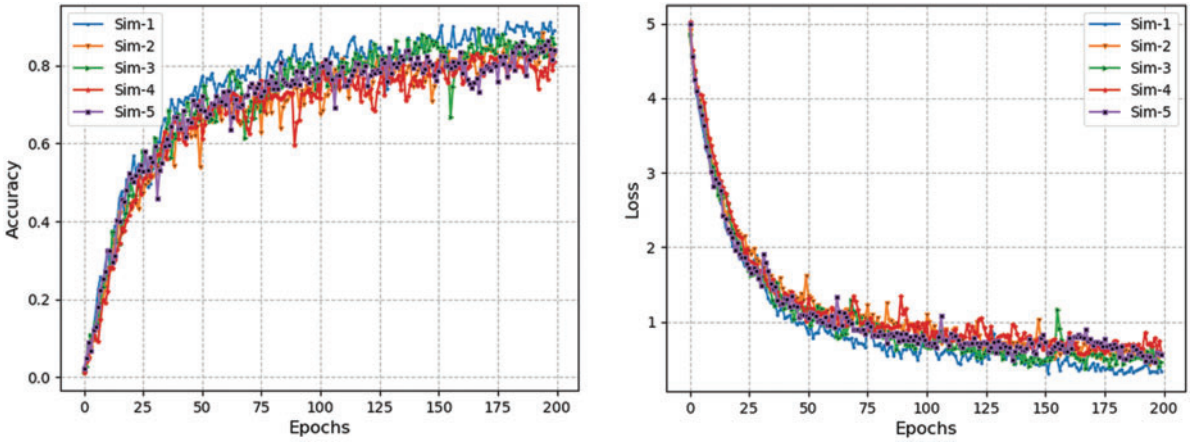

Figure 8: Accuracy and loss curves of "leg branch" of triplet-CNN over different simulations

The motivation behind this approach is to correctly classify the subject from the perspective of different parts of GEI even if some regions of GEI are affected due to covariate variables. It is observed from the above results and experiments that the proposed model shows good results in comparison with the single branch CNN model, as well as some other baseline methods including transfer learning approaches e.g., MobileNet and Vgg16 model. It is observed from Tables 2–4 that the proposed triplet CNN model outperforms the baseline models. The main reason behind these improvements is part-based learning which helps in learning features from different parts and the parts that are not disrupted by covariate factors will ultimately help in extracting more discriminative features. More precisely, the single-branch CNN model achieves an average accuracy of 48.238% while the proposed triplet CNN model achieves mean accuracy of 72.987%. Hence, there is an improvement of 24.749% is observed in this case. Similarly, 22.117% performance improvement is observed in the case of comparative analysis with transfer learning assisted VGG16 model. VGG refers to the Visual Geometry Group and is made up of units, each of which is made up of 2D Convolution and Max Pooling layers. It is available in two variants, VGG16 and VGG19, having 16 and 19 layers, respectively [66].

Table 2: Comparison results of proposed triplet CNN with single branch CNN in different walking conditions

No	Train set	Test set	Accuracy	Precision	Recall	F-score
			Proposed triplet branch CNN model			
01	Normal walk	Normal walk	96.290 ± 1.255	95.908 ± 1.6093	96.290 ± 1.2558	93.544 ± 1.533
02	Normal walk	Walk with bags	57.408 ± 3.0485	56.44 ± 2.7200	57.258 ± 3.149	57.258 ± 3.149
03	Normal walk	Walk with Coats	65.263 ± 1.47092	66.4098 ± 0.6461	65.080 ± 1.5248	62.77 ± 0.7764
		Mean	72.987 ± 1.9248	72.987 ± 1.9248	72.919 ± 1.6584	71.19067 ± 1.819
			Baseline single branch CNN model			
01	Normal walk	Normal walk	95 ± 1.118	95.81 ± 1.095	94.67 ± 1.118	94.264 ± 1.0868
02	Normal walk	Walk with bags	30.20 ± 2.64	29.04 ± 3.147	30.32 ± 2.54	27.34 ± 2.664
03	Normal walk	Walk with Coats	19.516 ± 4.20	10.61 ± 3.227	19.5164 ± 4.20	13.92 ± 3.313
		Mean	48.238 ± 2.65	45.513 ± 2.48	48.168 ± 2.619	45.17 ± 3.5319

Table 3: Comparison results of proposed triplet CNN with VGGNet (Transfer learning approach) in different walking conditions

No	Train set	Test set	Accuracy	Precision	Recall	F-score
			Proposed triplet branch CNN model			
01	Normal walk	Normal walk	96.290 ± 1.255	95.908 ± 1.6093	96.290 ± 1.2558	93.544 ± 1.533
02	Normal walk	Walk with bags	57.408 ± 3.0485	56.44 ± 2.7200	57.258 ± 3.149	57.258 ± 3.149
03	Normal walk	Walk with Coats	65.263 ± 1.47092	66.4098 ± 0.6461	65.080 ± 1.5248	62.77 ± 0.7764
		Mean	72.987 ± 1.9248	72.987 ± 1.9248	72.919 ± 1.6584	71.19067 ± 1.819
			Baseline transfer learning assisted VGG16			
01	Normal walk	Normal walk	90 ± 3.39	90.13±3.99	89.51 ± 3.40	88.369 ± 3.95
02	Normal walk	Walk with bags	44.53 ± 4.77	44.60 ±3.11	44.70 ± 4.75	41.64 ± 3.66
03	Normal walk	Walk with Coats	18.08 ± 0.618	16.11 ± 1.627	18.279 ± 0.839	15.001 ± 0.836
		Mean	50.87 ± 2.92	50.28 ± 2.909	50.829 ± 2.99	48.33 ± 2.815

Table 4: Comparison results of proposed triplet CNN with MobileNet (Transfer learning approach) in different walking conditions

No	Train set	Test set	Accuracy	Precision	Recall	F1-score
			Proposed triplet branch CNN model			
01	Normal walk	Normal walk	96.290 ± 1.255	95.908 ± 1.6093	96.290 ± 1.2558	93.544 ± 1.533
02	Normal walk	Walk with bags	57.408 ± 3.0485	56.44 ± 2.7200	57.258 ± 3.149	57.258 ± 3.149
03	Normal walk	Walk with Coats	65.263 ± 1.47092	66.4098 ± 0.6461	65.080 ± 1.5248	62.77 ± 0.7764
		Mean	72.987 ± 1.9248	72.987 ± 1.9248	72.919 ± 1.6584	71.19067 ± 1.819
			Baseline transfer learning assisted MobileNet			
01	Normal walk	Normal walk	68.15 ± 0.23	69.75 ± 0.64	68.27 ± 0.232	65.24 ± 0.011
02	Normal walk	Walk with bags	28.20 ± 3.29	25.58 ± 4.818	28.49 ± 3.28	29.88 ± 3.513
03	Normal walk	Walk with Coats	19.70 ± 1.82	19.813 ± 2.41	20.826 ± 1.81	17.18 ± 1.557
		Mean	38.68 ± 1.78	38.381 ± 2.622	39.19 ± 1.774	37.43 ± 1.693

Another model, i.e., MobileNet [67] is also employed as a baseline for comparative analysis. MobileNet is built on a simplified design that builds low-weight deep neural networks using depth-wise separable convolutions. It is observed from the results in Table 3 that the proposed triplet CNN also shows good improvement with great margins in comparison with the MobileNet model. In addition, it is also observed when no covariate conditions are involved then the performance of the proposed, including the baseline model shows good results e.g., in the case of the Normal walk. However, when covariate factors are involved then the performance is affected. For example, in the case of VGG16, the performance of the model is dropped to 44.53% in the carrying condition of bags, but the proposed triplet CNN better handle this problem of covariate factors and shows an accuracy of 57.068%. Similarly, with clothing conditions, the performance of the baselines model is very less, but the proposed triplet CNN achieves an accuracy of 65.2635% respectively. Similarly, the other metrics which include precision, recall, and F1-score have also encouraged values. For instance, a proposed triplet-CNN achieves a precision of about 72.987% which is higher than the baseline methods i.e., the baseline single-branch CNN model achieves a mean precision of 45.513%. On the other hand, the other two baseline architectures in which the concept of transfer learning is involved achieve a precision of 50.28% and 38.381% which are comparatively low values. In addition, if recall values are analyzed then the proposed triplet-CNN model also has a high recall of about 72.919% which is comparatively higher than baseline models including single-branch CNN architecture VGG16 and MobileNet as they have attained the recall values of 48.168%, 50.829%, and 39.19% respectively. Furthermore, the F1-score provides a combined evaluation of both precision and recall values. In the case of F1-Score, the values achieve up to 71.190% which is again higher than base-lines models of deep learning which exhibit the F1-Score of about 45.17%, 48.33%, and 37.43% respectively.

Moreover, to compare the proposed method with state-of-the-art methods, a summarized comparison is performed as shown in Table 5 in which those methods and their results are given. It is evident from Table 5 that in existing methods, the training and testing covariates are when same then the results are higher, and this is the most widely chosen setting as shown in Table 5 (i.e., where "no" is written in column 4). Yet, when models are subjected to diverse covariate circumstances, the performance of these approaches is veiled. Because the subject under real-time surveillance might be viewed under diverse settings at test time, i.e., the test situations such as a person coming with either bags and jackets or any other scenario are not known in advance. As a result, there is a requirement for a model that can be trained on a person's usual walking scenarios and then evaluated on a person when they encounter any walking situations. When analysis of the model is carried out with this experimental setting, the results are relatively poor as shown in Table 5 (i.e., where "yes" is written in column 4). This is because deep learning models are incapable of dealing with unknown covariate circumstances. To address this, architecture is modified and learning of the CNN model to triple-branch CNN which is good enough to deal with unknown walking scenarios as given in Table 5 and shows 72.987% accuracy.

To conclude, the proposed model was observed to be more advantageous in comparison with single-branch and transfer learning methods as shown in Tables 2–5. The reason for this improved performance is the region-based learning because due to covariate conditions, the GEI image is disrupted to some extent such as when a person wears a coat then a body part is more affected. To counter this, the triplet-CNN model tries to identify a person from several regions such as the head, body, and legs with the possibility that certain specific parts are not influenced and a person may be readily identified, resulting in the strength of the triplet-CNN. Furthermore, it is necessary to acknowledge the limitation of the proposed work, hence, one possible limitation is that when covariate conditions disrupted almost all regions of GEI then this will ultimately hinder the process of discriminative feature learning and performance drops. For example, a person wearing a cap along

with coats, then in this case both head and body part is affected. To overcome this in the future, GANs-based models can be built to reconstruct the affected regions to normal walk GEIs. Furthermore, the proposed work can also be extended to a one-shot learning model as a verification task in which a similarity learning-based deep learning model is extended to regions-based similarity learning.

Table 5: Comparison results of proposed triplet CNN with state-of-the-art methods

Authors	Models	Dataset	Different covariate conditions in Train/Test	Performance (%)
Mogan et al. [38]	VGG16-MLP	CASIA-B	No	100%
Min et al. [68]	Deep CNN	CASIA-B	No	98.75%
Aung et al. [69]	CNN	CASIA-B	No	92.94%
Howard et al. [67]	MobileNet*	CASIA-B	Yes	38.68%
Simonyan et al. [66]	VGG16*	CASIA-B	Yes	50.87%
Min et al. [68]	Deep CNN*	CASIA-B	Yes	48.23%
Proposed	Triplet-CNN	CASIA-B	**Yes**	**72.987%**

Note: **reproduced*.

5 Conclusion

Human gait-based surveillance systems are one of the evolving biometric technologies used to enable surveillance at different places. As a vision-based surveillance system, it has potential advantages as it allows to recognize people even when they are uncooperative. In the existing literature, several approaches comprising deep learning and traditional machine learning methods are designed, nevertheless, covariate variables are one of the challenging problems that hinder the accuracy of the underlying system in identifying different subjects. These covariate variables affect certain regions of gait representation images, hence, to cope with this problem, a triplet branch CNN is proposed that is able to deal with each region of the GEI image separately. Later on, probability-based majority voting criteria are designed to finally decide the label of subjects. Moreover, soft attention layers are also added to refine the extracted features from each region. The experiments have been conducted on the CASIA-B gait dataset and triplet-CNN achieve good results. The findings of the research indicate that the region-based CNN model has an improvement with great margins over single-branch models. In the future, the proposed model can also be extended to include the region-proposal network as a layer in the model to extract regions automatically instead of considering only three regions. Additionally, the triplet-branch CNN model is expanded and trained as a transfer learning strategy, with each branch based on a pre-trained deep learning model and fine-tuned to perform classification.

Acknowledgement: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1F1A1063134) and also the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) Support Program (IITP-2022-2018-0-01799) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Funding Statement: This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1F1A1063134) and also the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center)

Support Program (IITP-2022-2018-0-01799) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] J. P. Singh, S. Jain, S. Arora and U. P. Singh, "Vision-based gait recognition: A survey," *IEEE Access*, vol. 6, pp. 70497–70527, 2018.

[2] M. Kumar, N. Singh, R. Kumar, S. Goel and K. Kumar, "Gait recognition based on vision systems: A systematic survey," *Journal of Visual Communication and Image Representation*, vol. 75, no. 6, pp. 103052, 2021.

[3] C. Wan, L. Wang and V. V. Phoha, "A survey on gait recognition," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–35, 2018.

[4] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother *et al.,* "The humanID gait challenge problem: Data sets, performance, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162–177, 2005.

[5] M. S. Nixon, J. N. Carter, D. Cunado, P. S. Huang and S. Stevenage, "Automatic gait recognition," in *Biometrics: Personal Identification in Networked Society*, New York, NY: Springer, pp. 231–249, 1996.

[6] M. S. Nixon, T. Tan and R. Chellappa, *Human identification based on gait*, vol. 4. Berlin, Germany: Springer Science & Business Media, 2010.

[7] C. Shen, S. Yu, J. Wang, G. Q. Huang and L. Wang, "A comprehensive survey on deep gait recognition: Algorithms, datasets and challenges," *arXiv preprint arXiv*:2206.13732, 2022.

[8] S. X. Yang, P. K. Larsen, T. Alkjær, E. B. Simonsen and N. Lynnerup, "Variability and similarity of gait as evaluated by joint angles: Implications for forensic gait analysis," *Journal of Forensic Sciences*, vol. 59, no. 2, pp. 494–504, 2014.

[9] C. BenAbdelkader, R. Cutler and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification," in *Proc. of Fifth IEEE Int. Conf. on Automatic Face Gesture Recognition*, Washington, DC, USA, pp. 372–377, 2002.

[10] S. K. Gupta, "Reduction of covariate factors from Silhouette image for robust gait recognition," *Multimedia Tools and Applications*, vol. 80, no. 28–29, pp. 36033–36058, 2021.

[11] R. Nawaratne, D. Alahakoon, D. De Silva and X. Yu, "Spatiotemporal anomaly detection using deep learning for real-time video surveillance," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 393–402, 2019.

[12] M. Bukhari, K. B. Bajwa, S. Gillani, M. Maqsood, M. Y. Durrani *et al.,* "An efficient gait recognition method for known and unknown covariate conditions," *IEEE Access*, vol. 9, pp. 6465–6477, 2020.

[13] A. Ari and D. Hanbay, "Deep learning based brain tumor classification and detection system," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 26, no. 5, pp. 2275–2286, 2018.

[14] R. Ashraf, S. Afzal, A. U. Rehman, S. Gul, J. Baber *et al.,* "Region-of-interest based transfer learning assisted framework for skin cancer detection," *IEEE Access*, vol. 8, pp. 147858–147871, 2020.

[15] M. A. Haq, "Planetscope nanosatellites image classification using machine learning," *Computer Systems Science & Engineering*, vol. 42, no. 3, pp. 1031–1046, 2022.

[16] M. A. Haq, M. A. R. Khan and A. Talal, "Development of PCCNN-based network intrusion detection system for EDGE computing," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 1769–1788, 2022.

[17] M. A. Haq and M. A. R. Khan, "DNNBoT: Deep neural network-based botnet detection and classification," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 1729–1750, 2022.

[18] N. Razfar, J. True, R. Bassiouny, V. Venkatesh and R. Kashef, "Weed detection in soybean crops using custom lightweight deep learning models," *Journal of Agriculture and Food Research*, vol. 8, no. 3, pp. 100308, 2022.

[19] M. A. Haq, "CNN based automated weed detection system using UAV imagery," *Computer Systems Science & Engineering*, vol. 42, no. 2, pp. 837–849, 2022.

[20] A. Saaudi, Z. Al-Ibadi, Y. Tong and C. Farkas, "Insider threats detection using CNN-LSTM model," in *Int. Conf. on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, pp. 94–99, 2018.

[21] A. S. M. Miah, M. A. M. Hasan and J. Shin, "Dynamic hand gesture recognition using multi-branch attention based graph and general deep learning model," *IEEE Access*, vol. 11, pp. 4703–4716, 2023.

[22] M. Zhang, W. Wang, G. Xia, L. Wang and K. Wang, "Self-powered electronic skin for remote human-machine synchronization," *ACS Applied Electronic Materials*, vol. 5, no. 1, pp. 498–508, 2023.

[23] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316–322, 2005.

[24] S. K. Gupta, G. M. Sultaniya and P. Chattopadhyay, "An efficient descriptor for gait recognition using spatio-temporal cues," in *Emerging Technology in Modelling and Graphics*, Berlin, Germany: Springer, pp. 85–97, 2020.

[25] K. Bashir, T. Xiang and S. Gong, "Gait recognition without subject cooperation," *Pattern Recognition Letters*, vol. 31, no. 13, pp. 2052–2060, 2010.

[26] T. H. Lam and R. S. Lee, "A new representation for human gait recognition: Motion silhouettes image (MSI)," in *Int. Conf. on Biometrics*, Hong Kong, China, pp. 612–618, 2006.

[27] N. Liu, J. Lu, Y. -P. Tan and Z. Chen, "Enhanced gait recognition based on weighted dynamic feature," in *2009 16th IEEE Int. Conf. on Image Processing (ICIP)*, Cairo, Egypt, pp. 3581–3584, 2009.

[28] K. Bashir, T. Xiang and S. Gong, "Cross view gait recognition using correlation strength," in *British Machine Vision Conf.*, Aberystwyth, UK, pp. 1–11, 2010.

[29] P. Arora, M. Hanmandlu and S. Srivastava, "Gait based authentication using gait information image features," *Pattern Recognition Letters*, vol. 68, no. 2, pp. 336–342, 2015.

[30] X. Yang, Y. Zhou, T. Zhang, G. Shu and J. Yang, "Gait recognition based on dynamic region analysis," *Signal Processing*, vol. 88, no. 9, pp. 2350–2356, 2008.

[31] G. M. Linda, G. Themozhi and S. R. Bandi, "Color-mapped contour gait image for cross-view gait recognition using deep convolutional neural network," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 18, no. 1, pp. 1941012, 2020.

[32] M. Alotaibi and A. Mahmood, "Improved gait recognition based on specialized deep convolutional neural network," *Computer Vision and Image Understanding*, vol. 164, no. 13, pp. 103–110, 2017.

[33] H. Arshad, M. A. Khan, M. I. Sharif, M. Yasmin, J. M. R. Tavares *et al.,* "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition," *Expert Systems*, vol. 39, no. 7, pp. e12541, 2020.

[34] X. Wu, T. Yang and Z. Xia, "Gait recognition based on densenet transfer learning," *International Journal of Science and Environment*, vol. 9, pp. 1–14, 2020.

[35] Y. Guan, C. T. Li and Y. Hu, "Robust clothing-invariant gait recognition," in *Eighth Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing*, Piraeus-Athens, Greece, pp. 321–324, 2012.

[36] M. R. Aqmar, K. Shinoda and S. Furui, "Robust gait recognition against speed variation," in *20th Int. Conf. on Pattern Recognition*, Washington, DC, United States, pp. 2190–2193, 2010.

[37] S. Zheng, J. Zhang, K. Huang, R. He and T. Tan, "Robust view transformation model for gait recognition," in *18th IEEE Int. Conf. on Image Processing*, Brussels, Belgium, pp. 2073–2076, 2011.

[38] J. N. Mogan, C. P. Lee, K. M. Lim and K. S. Muthu, "VGG16-MLP: Gait recognition with fine-tuned VGG-16 and multilayer perceptron," *Applied Sciences*, vol. 12, no. 15, pp. 7639, 2022.

[39] J. N. Mogan, C. P. Lee, K. S. M. Anbananthen and K. M. Lim, "Gait-DenseNet: A hybrid convolutional neural network for gait recognition," *IAENG International Journal of Computer Science*, vol. 49, no. 2, pp. 393–400, 2022.

[40] K. Ambika and K. R. Radhika, "View invariant gait authentication using transfer learning," in *Proc. of the Int. Conf. on Innovative Computing & Communication (ICICC) 2021*, 2021. https://doi.org/10.2139/ssrn.3835056

[41] A. Mehmood, M. A. Khan, M. Sharif, S. A. Khan, M. Shaheen *et al.,* "Prosperous human gait recognition: An end-to-end system based on pre-trained CNN features selection," *Multimedia Tools and Applications*, vol. 68, pp. 1–21, 2020.

[42] H. Arshad, M. A. Khan, M. I. Sharif, M. Yasmin, J. M. R. Tavares *et al.,* "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition," *Expert Systems*, vol. 39, no. 7, pp. e12541, 2022.

[43] A. Kececi, A. Yildirak, K. Ozyazici, G. Ayluctarhan, O. Agbulut *et al.,* "Implementation of machine learning algorithms for gait recognition," *Engineering Science and Technology, an International Journal*, vol. 23, no. 4, pp. 931–937, 2020.

[44] A. H. Bari and M. L. Gavrilova, "Artificial neural network based gait recognition using kinect sensor," *IEEE Access*, vol. 7, pp. 162708–162722, 2019.

[45] R. Liao, S. Yu, W. An and Y. Huang, "A model-based gait recognition method with body pose and human prior knowledge," *Pattern Recognition*, vol. 98, no. 2, pp. 107069, 2020.

[46] T. Teepe, J. Gilg, F. Herzog, S. Hörmann and G. Rigoll, "Towards a deeper understanding of skeleton-based gait recognition," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, pp. 1569–1577, 2022.

[47] W. An, R. Liao, S. Yu, Y. Huang and P. C. Yuen, "Improving gait recognition with 3D pose estimation," in *Chinese Conf. on Biometric Recognition*, Urumchi, China, pp. 137–147, 2018.

[48] L. Zheng, Y. Zha, D. Kong, H. Yang and Y. Zhang, "Multi-branch angle aware spatial temporal graph convolutional neural network for model-based gait recognition," *IET Cyber-Systems and Robotics*, vol. 4, no. 2, pp. 97–106, 2022.

[49] S. Gul, M. I. Malik, G. M. Khan and F. Shafait, "Multi-view gait recognition system using spatio-temporal features and deep learning," *Expert Systems with Applications*, vol. 179, no. 1109/34, pp. 115057, 2021.

[50] W. A. Alsaggaf, I. Mehmood, E. F. Khairullah, S. Alhuraiji, M. F. S. Sabir *et al.,* "A smart surveillance system for uncooperative gait recognition using cycle consistent generative adversarial networks (CCGANs)," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–12, 2021.

[51] X. Li, Y. Makihara, C. Xu, Y. Yagi and M. Ren, "Joint intensity transformer network for gait recognition robust against clothing and carrying status," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 12, pp. 3102–3115, 2019.

[52] V. B. Semwal, A. Mazumdar, A. Jha, N. Gaud and V. Bijalwan, "Speed, cloth and pose invariant gait recognition-based person identification," in *Machine Learning: Theoretical Foundations and Practical Applications*, Berlin, Germany: Springer, pp. 39–56, 2021.

[53] L. Yao, W. Kusakunniran, Q. Wu, J. Zhang, Z. Tang *et al.,* "Robust gait recognition using hybrid descriptors based on skeleton gait energy image," *Pattern Recognition Letters*, vol. 150, no. 8, pp. 289–296, 2021.

[54] K. Bashir, T. Xiang and S. Gong, "Gait recognition using gait entropy image," in *3rd Int. Conf. on Imaging for Crime Detection and Prevention (ICDP 2009)*, London, UK, 2009.

[55] M. Uddin, D. Muramatsu, N. Takemura, M. Ahad, A. Rahman *et al.,* "Spatio-temporal silhouette sequence reconstruction for gait recognition against occlusion," *IPSJ Transactions on Computer Vision and Applications*, vol. 11, no. 1, pp. 1–18, 2019.

[56] A. Roy, S. Sural, J. Mukherjee and G. Rigoll, "Occlusion detection and gait silhouette reconstruction from degraded scenes," *Signal, Image and Video Processing*, vol. 5, no. 4, pp. 415–430, 2011.

[57] M. Hofmann, D. Wolf and G. Rigoll, "Identification and reconstruction of complete gait cycles for person identification in crowded scenes," in *Proc. Intern. Conf. on Computer Vision Theory and Applications (VISAPP)*, Algarve, Portugal, 2011.

[58] D. Muramatsu, Y. Makihara and Y. Yagi, "Gait regeneration for recognition," in *Int. Conf. on Biometrics (ICB)*, Phuket, Thailand, pp. 169–176, 2015.

[59] C. Affonso, A. L. D. Rossi, F. H. A. Vieira and A. C. P. de Leon Ferreira, "Deep learning for biological image classification," *Expert Systems with Applications*, vol. 85, no. 24, pp. 114–122, 2017.

[60] A. Mauri, R. Khemmar, B. Decoux, M. Haddad and R. Boutteau, "Real-time 3D multi-object detection and localization based on deep learning for road and railway smart mobility," *Journal of Imaging*, vol. 7, no. 8, pp. 145, 2021.

[61] A. Kirillov, K. He, R. Girshick, C. Rother and P. Dollár, "Panoptic segmentation," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, CA, USA, pp. 9404–9413, 2019.

[62] M. A. Shaikh, T. Duan, M. Chauhan and S. N. Srihari, "Attention based writer independent verification," in *17th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR)*, Dortmund, Germany, pp. 373–379, 2020.

[63] N. Tomita, B. Abdollahi, J. Wei, B. Ren, A. Suriawinata *et al.,* "Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides," *JAMA Network Open*, vol. 2, no. 11, pp. e1914645, 2019.

[64] D. Tran, L. Bourdev, R. Fergus, L. Torresani and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Santiago, Chile, pp. 4489–4497, 2015.

[65] S. Yu, D. Tan and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *18th Int. Conf. on Pattern Recognition (ICPR'06)*, Hong Kong, China, pp. 441–444, 2006.

[66] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv*:1409.1556, 2014.

[67] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang *et al.,* "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv*:1704.04861, 2017.

[68] P. P. Min, S. Sayeed and T. S. Ong, "Gait recognition using deep convolutional features," in *7th Int. Conf. on Information and Communication Technology (ICoICT)*, Kuala Lumpur, Malaysia, pp. 1–5, 2019.

[69] H. M. L. Aung and C. Pluempitiwiriyawej, "Gait biometric-based human recognition system using deep convolutional neural network in surveillance system," in *Asia Conf. on Computers and Communications (ACCC)*, Singapore, pp. 47–51, 2020.