International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 14, No.1 (Sep-2023)

http://dx.doi.org/10.12785/ijcds/140157

# Front and Back Views Gait Recognitions Using EfficientNets and EfficientNetV2 Models Based on Gait Energy Image

**Tengku Mohd Afendi Zulcaffle[1], Fatih Kurugollu[2], Kuryati Kipli[1,], Annie Joseph[1,] and David B. L. Bong[1,]**

[1]*Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia*
[2]*College of Computing and Informatics, University of Sharjah, Sharjah, United Arab Emirates*

*Received 24 Sep. 2022, Revised 29 Jul. 2023, Accepted 10 Aug. 2023, Published 01 Sep. 2023*

**Abstract:** Front and back views gait recognitions are important, especially for narrow corridor applications. Hence, it is important to experiment with new algorithms on the front and back views gait recognitions. In this paper, we present the experiments on gait recognition using the pretrained EfficientNets and EfficientNetV2 models and Gait Energy Image. These models are chosen because they are among the best deep learning models in computer vision. The pretrained models were used in this experiment because it can produce faster and better accuracies compared to training the models from scratch. In addition to the pretrained models, we also propose ensemble models so that they can produce better accuracies. The result shows that the EfficientNetB7-Augm+ EfficientNetB6-Augm is the best overall accuracy (79.59%). However, combining the models slow down the inference speed. So, for recognition speed, EfficientNetB6 and EfficientNetB6-Augm are the best with 87.01ms speed per input image. The results produced are very good considering no cross-view algorithms applied to the Gait Energy Image. Future works will include the cross-view algorithms to further improve the accuracies of the proposed method.

**Keywords:** Gait Recognition, Deep Learning, EfficientNets, EfficientNetV2

## 1. INTRODUCTION

Gait is a walking process that involves the combination of posture and bodily motion [1]. These produce features that are useful for biometric applications [2]. It is applied in biometrics due to the uniqueness of the gait patterns of an individual [2], [3]. Unlike other biometrics, gait features can be identified using a camera, so it is contactless, without any specific spot, and can be acquired without the awareness and cooperation of the individuals under observation. Also, gait is difficult to be concealed and disguised. Gait recognition can be employed using low-resolution cameras and the features are identifiable at a long distance [2].

The gait features can be captured and applied at any view angle. The front and back views are the most suitable angles to capture gait features in narrow corridor situations [4]. Recently, many studies focused on cross-view gait recognition. Cross-view involves a gallery (the data/images made known to the algorithm) of one view and a probe (the data/images tested on the algorithm) in different views. The experiments on cross views of front and back are included in the studies. Most of these research employ the gait image representation known as Gait Energy Image (GEI) invented by Han and Bhanu [5]. The GEI is widely used in gait recognition because it is robust against noise.

Among the earliest that conduct cross-view gait recognition is Yu et al. [6]. They employ the raw GEI feature and nearest neighbour as the classifier. Same as in [6], the CMCC [7] also applies the GEI in their method. However, in [7], the GEI is further processed by creating features to overcome the view difference problems. The features are produced by using the Canonical Correlation Analysis (CCA) on the same segment of different views. Like [7], the method in [8] employs the same method on overcoming the cross-view problems. In [7] and [8], sum of cosine similarity and nearest neighbour are employed as the classifiers respectively.

The methods in [9], [10], [11] employ a method known as View Transformation Model (VTM). In [9], Support Vector Regression (SVR) is employed to create the VTM. In [10], the VTM is generated using SVD from GEI and the quality measures are incorporated to further improve the VTM. In [11], Optimized Gait Energy Image and Truncated Singular Value Decomposition (TSVD) are utilized to build the VTM. For classification, the similarity of features measurement based on Euclidean distance is carried out in [9] and [11]to recognize the individuals. On the other hand, in [10], the dissimilarity score is utilized as the classifier to measure the difference between the probe and gallery.

*E-mail address: ztmafendi@unimas.my*

TABLE I. Summary of the related works

| Paper | Features | Additional Processing and Classifier |
|---|---|---|
| Yu et al. [6] | Raw GEI | Nearest Neighbour |
| Kusakunniran et al. [7] | Canonical Correlation Analysis based on GEI | Sum of Cosine Similarity |
| Xing et al. [8] | Complete Canonical Correlation Analysis based on GEI | Nearest Neighbour |
| Kusakunniran et al. [9] | VTM generated using SVR from GEI | Similarity measurement based on Euclidean Distance |
| Muramatsu et al. [10] | VTM generated using SVD from GEI and the VTM enhanced using the quality measures | Dissimilarity score |
| Kusakunniran et al. [11] | VTM based on Optimized Gait Energy Image and TSVD | Similarity measurement based on Euclidean Distance |
| Wu et al. [12] | subGEI generated from GEI using spatial pyramid matching. | New Convolutional Neural Network |
| Ben et al. [13] | CPA generated from GEI | Nearest Neighbour |
| Ben et al. [14] | CBDP generated from GEI | Improved metric learning approach |
| Isik and Ekenel [15] | Retrained the VGG16 feature layers using binary silhouette and RGB images to produce new features. | Maximum similarity using cosine classifier |
| Zhang et al. [16] | Sequence of Binary silhouettes | Fusion of convolutional variational autoencoder and deep Koopman embedding |

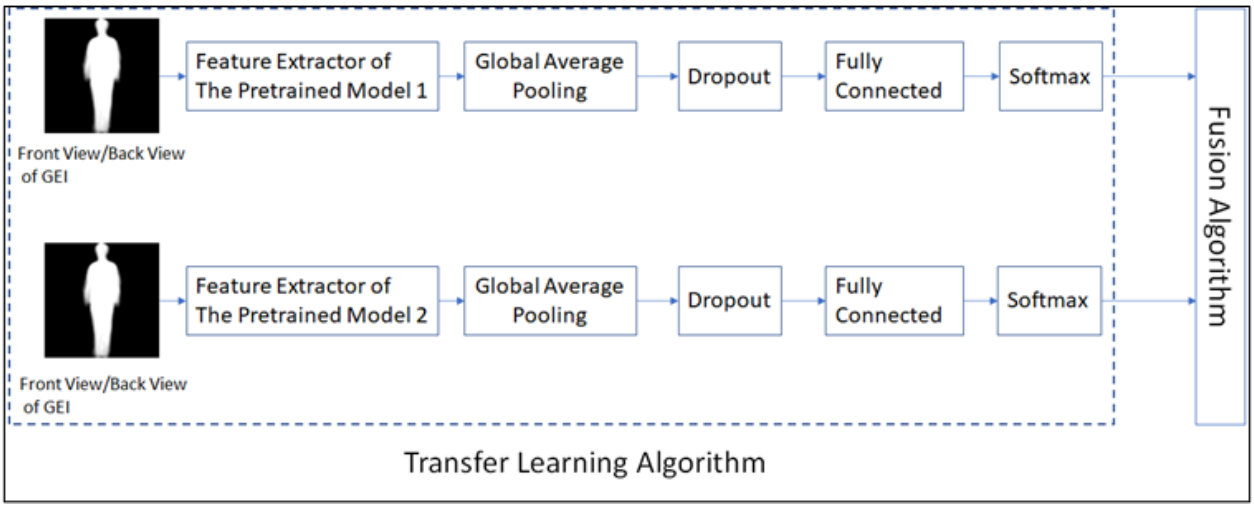


Figure 1. The Proposed Algorithm

To overcome the cross-view limitation, Wu et al. [12] used the spatial pyramid matching in [17] to produce a gait image representation called subGEI from the GEI. Later they created a new Convolutional Neural Network (CNN) model to classify the individuals. The CPA [13] and CBDP [14] apply many parameters to extract features from GEI for cross-view gait recognition. In [13] and [14], they use the nearest neighbour and improved metric learning approach as classifiers respectively. In [15], Isik and Ekenel used the gait sequence in the forms of silhouette and RGB images as the features and employs VGG16 to extract the features. After that, the maximum similarity based on the cosine technique is employed as the classifier.

Zhang et al. [16] used the silhouette from the gait sequence and employ the fusion of convolutional variational autoencoder and deep Koopman embedding to recognize individuals based on their gait patterns.. The multiview gait recognition studies are not discussed here because they used multiple views as a gallery and any one view as the probe such as the recent studies in [18], [19], [20].

The summary of the related works is presented in Table I. Based on the related works carried out, none of the methods employed pretrained CNNs. Using the pretrained model can speed up the development process. Hence, it is important to study the effectiveness of the pretrained CNNs, especially the state-of-the-art models such as the EfficientNets [21] and EfficienNetV2s [22].

The EfficientNets [21] and EfficienNetV2s [22] are two groups of models that show great performance compared to many models when tested using publicly available datasets. Hence, it is important to understand the impact of these models on gait recognition. Therefore, our contributions are:

- Rigorous experimental studies of the best EfficientNets [21] and EfficienNetV2s [22] models for front and back views gait recognition using raw GEI.

- A new ensemble model based on the best models of EfficientNets [21] and EfficienNetV2s [22].

The organization of this paper is as follows: Section 2 explains the proposed methodology for the front and back views gait recognition algorithms. The experimental results of this research are presented in Section 3. Section 4 analyzes and discusses the results. Section 5 concludes the paper.

## 2. METHODOLOGY

In this experiment, several pretrained models based on the EfficientNets [21] and EfficientNetV2 [22] are employed. In addition to that, we proposed ensembled models based on the models in [21] and [22] to improve the accuracies.

The EfficientNets is a family of models created for better performance by scaling up MobileNets and ResNets in terms of the depth and width of the models and input image resolutions. Eight models in the EfficientNets were created and the transfer learning experiments based on multiple datasets shows that in general, EfficientNetB7 was the best algorithm and other EfficientNets' models performed better than most of the previous models [21].

In addition to EfficientNets, Mingxing Tan, and Quoc Le created another family of models called EfficientNetV2. The motivation of their research was to have faster training speed and parameter efficiency. This was carried out by including adaptive regularization techniques to produce the expected results. Based on the transfer learning results, EfficientNetV2-L and EfficientNetV2-M were the best models [22].

Also based on Keras Applications [23], the best four Rank-1 accuracy are EfficientNetV2-L, EfficientNetV2-M, EfficienNetB7, and EfficentNetB6. These models outperformed many previous models on the ImageNet dataset [24]. Hence, in this research, transfer learning and ensemble learning experiments are conducted based on these four models. The general framework of the proposed algorithm is shown in Fig. 1. In this experiment, the GEI with the size of $240 \times 240$ is used as the input images to the deep learning models. The GEI is the mostly used gait image representation which was created by Han and Bhanu [5]. Before producing the GEI, the human silhouette in each frame is extracted and binarized. The binarization process eliminates the impact of colour and illumination on gait recognition accuracy. All binary silhouettes in one gait cycle are averaged to produce the GEI. The averaging process reduces errors caused by the silhouette extraction algorithm. One gait cycle involves two strides, left and right strides.

The first part of the proposed algorithm is the transfer learning process as indicated by the dotted box in Fig. 1. As mentioned in the previous section, the EfficientNetV2-L, EfficientNetV2-M, EfficienNetB7, EfficentNetB6 are experimented. The models were pretrained using ImageNet which contains 1000 classes of objects with RGB colour images. The feature extractors of these models as in [23] are utilized, and this is carried out by removing the classifiers or the top parts of the models. The Global Average Pooling averaging reduces the dimension of the last layer of the feature extractor. Next, the dropout layer is included as the regularization technique to reduce overfitting with a dropout rate is 0.5. After that, the fully connected network is applied to classify the individuals based on their gait patterns. Finally, the Softmax activation function is employed to produce the probabilistic output of the classification.

For the training of the transfer learning model, the Adaptive Moment Estimation (Adam) optimizer is employed with learning rate, $\eta = 0.001$, exponential decay rate parameters, $\beta 1 = 0.9$, $\beta 2 = 0.9999$, and $\epsilon = 10^{-8}$.

These parameters are chosen because they are good parameters in general as recommended in the Adam paper [25]. The categorical cross entropy is used as the loss function. All the models are trained up to 500 epochs with a batch size is 32. In avoiding overfitting, the early stopping method is introduced. The training will stop if the validation loss does not change for 50 epochs. The fusion algorithm or ensembled technique is implemented by using the following formula:

$$T = argmax \sum_{i=1}^{j} y(x)_i \quad (1)$$

where T is the identity of a subject being tested, the y are the predicted probabilities of the $x$ subjects in the gallery by model $i$ in the ensembled models. This equation sums the predictions of all the models and finds the index of the summation results with the maximum value as the individual identity.

## 3. EXPERIMENTAL RESULTS

In this experiment, besides the test on accuracies, the inference time was also computed. Both are carried out using Python 3.10.4 programming language with deep learning libraries such as Tensorflow 2.8.0 and Keras 2.8.0. The programs are run on a computer with the following specifications:

- Computer System: HP Pavilion Gaming Laptop 15-dk0xxx
- Microprocessor: Intel® CoreTM i5-9300H CPU @ 2.40GHz
- Installed RAM: 8.00 GB
- Operating System: Windows 10 Home Edition (64-bit)
- Graphic Processor Unit (GPU): NVIDIA GeForce GTX 1650, 1560MHz, 4096 MB GGDR5.

In this experiment, the CASIA B dataset [6], [26] is used to examine the performance of the proposed algorithms. Only the normal walks gait sequences are used with NM01-NM04 used as the gallery and NM05-NM06 are utilized for the probe. The GEI images were used are as provided by [26]. In this experiment, only front (0°) and back (180°) views are considered. The results of the accuracies are presented in Table II. Hence, we have four categories of experiments: i) gallery 0° and probe 0° ii) gallery 0° and probe 180° iii) gallery 180° and probe 0° iv) gallery 180° and probe 180°. In this experiment, first, we conduct training on the first four models in Table I, the EfficientNetV2-M, EfficientNetV2-L, EfficentNetB6, and EfficienNetB7. Due to the better performance of EfficentNetB6 and EfficienNetB7 than EfficientNetV2-M and EfficientNetV2-L, the experiments on the performance of EfficentNetB6, and EfficienNetB7 on augmented images are also carried out. Furthermore, experiments on the fusion of models based on EfficentNetB6 and EfficienNetB7 are also conducted with the expectation to produce better results.

In addition to accuracy experiments, the speeds of the models are also compared. Fig. 2 shows the overall accuracy versus the inference time plot for all the algorithms in this experiment. The inference time is the average time for each of the models that run 10 times and identification is using a single input image. The inference process was run on the aforementioned GPU. For the models that have image augmentation, such as the combinations of EfficientNetB6 and EfficientNetB6-Augm, EfficientNetB7 and EfficientNetB7-Augm, we use the same result because the augmentation only affects the values of the trainable parameters and does not change other factors that affect the inference speed. Similarly, for the EfficientNetB6+EfficentNetB7, the result is using the same inference time results as the following combined proposed models: EfficientNetB7-Augm+EfficientNetB6-Augm, EfficentNetB6-Augm+EfficientNetB7, and also the EfficientNetB7-Augm+EfficientNetB6.

## 4. DISCUSSION

Based on the results in Table II, the EfficientNetB7 and EfficientNetB6 perform better than the EfficientNetV2-M, and EfficientNetV2-L. Although in [22], the EfficientNetV2 performed better than the EfficientNet, the experiment on the Flowers dataset [27], shows that the models in EfficientNets can possibly produce better results than the EfficientNetV2 models. In the experiment, the EfficientNetB7, EfficientNetV2-L, and EfficientNetV2-M produced, 98.8%, 98.8 ± 0.05%, and 98.5% ± 0.08% respectively. On the other hand, the EfficientNetV2 models performed much better than the EfficientNets models on the ImageNet dataset [24]. This might be caused by the diversified classes of the ImageNet such as CD player, canoe, baseball, land animals, sea animals, etc. Furthermore, the flower dataset is limited to flower types of classes only. Hence, the features of the Flower dataset are narrower (quite similar to each other) than the ImageNet dataset. Similarly, GEI has narrower features compared to Flower and ImageNet datasets. In terms of shape features, it is more difficult to distinguish individuals based on GEI compared to flowers in the Flower dataset. Moreover, the GEI input features are grayscale which has one channel only while flowers in the Flower dataset are in RGB format having three channels. This makes the flowers easier to be distinguished than human individuals using GEI features. This can be the reason why the EfficientNets perform better than the EfficientNetV2 in this experiment which is using GEI.

As presented in Table II, for the gallery 0° and probe 0°, as expected EfficientNetB7 performs better than EfficientNetB6. However, for the gallery 180° and probe 180°, they produce the same accuracy. For the cross-view experiments (gallery 0° and probe 180°, and gallery 0° and probe 180°), the EfficientNetB6 performs better than the EfficientNetB7. This is similar to the result in [21], based on Oxford-IIIT Pets dataset [28]. The dataset in [28] contains, cats and dogs in multiple views with narrow features (dogs and cats only). Based on this dataset, EfficientNetB6 is the best compared to all the EfficientNets models including the EfficientNetB7.

In this experiment, we perform the augmentation on

TABLE II. Top-1 accuracies of all the models

| Algorithms | Gallery View | Probe 0° | Probe 180° | Each Gallery View | Overall Accuracy |
|---|---|---|---|---|---|
| EfficientNetV2-M [22] | 0 | 88.33 | 28.33 | 58.33 | |
| | 180 | 35.00 | 89.58 | 62.29 | 60.31 |
| EfficientNetV2-L [22] | 0 | 84.58 | 20.42 | 52.50 | |
| | 180 | 28.33 | 81.67 | 55.00 | 53.75 |
| EfficientNetB6 [21] | 0 | 96.67 | 50.42 | 73.55 | |
| | 180 | 59.58 | 97.08 | 78.33 | 75.94 |
| EfficientNetB7 [21] | 0 | 98.33 | 46.25 | 72.29 | |
| | 180 | 56.25 | 97.08 | 76.67 | 74.48 |
| EfficientNetB6-Augm[a][21] | 0 | 95.42 | 55.00 | 75.21 | |
| | 180 | 56.67 | 95.00 | 75.84 | 75.53 |
| EfficientNetB7-Augm[a][21] | 0 | 98.33 | 48.75 | 73.54 | |
| | 180 | 55.83 | 97.92 | 76.88 | 75.21 |
| EfficientNetB7-Augm.+EfficientNetB6-Augm[a] | 0 | 98.33 | **56.67** | **77.50** | |
| | 180 | 65 | 98.33 | 81.67 | **79.59** |
| EfficientNetB7-Augm.+EfficientNetB6[a] | 0 | **98.75** | 52.97 | 75.84 | |
| | 180 | 62.08 | 97.92 | 80.00 | 77.92 |
| EfficientNetB7-Augm.+EfficientNetB7[a] | 0 | 98.33 | 49.17 | 73.75 | |
| | 180 | 58.75 | 97.5 | 78.13 | 75.94 |
| EfficientNetB6-Augm.+EfficientNetB7[a] | 0 | 98.33 | 55 | 76.67 | |
| | 180 | **65.42** | **98.75** | **82.09** | 79.38 |
| EfficientNetB6-Augm.+EfficientNetB6[a] | 0 | 96.67 | 52.92 | 74.80 | |
| | 180 | 59.17 | 96.25 | 77.71 | 76.26 |
| EfficientNetB7+EfficientNetB6 | 0 | **98.75** | 52.92 | 75.84 | |
| | 180 | 64.17 | 98.33 | 81.25 | 78.55 |

a. Augm stands for Augmented

the training set by performing horizontal flipping. For the EfficientB6, the performance decreases for all cases except for gallery 0° and probe 180° where it increases from 50.42% to 55.00%. For the EfficientNetB7, for both 180° probe experiments using the same view (gallery 180°) and cross-view (gallery 0°), the performances have improved from 97.08% to 97.92% and 46.25% to 48.75%. For gallery 0° and probe 0°, the result remains the same but for gallery 180° and probe 0°, it is slightly reduced from 56.25% to 55.83%. In general, if we compute the average accuracy for the augmentation experiments (EfficientNetB6-Augm and EfficientNetB7-Augm), it is 75.37% which is slightly better than the experiments without augmentation, EfficientNetB6 and EfficientNetB7 that only produce 75.21% average accuracy. Hence, we can infer that augmentation helps the models to improve by adding additional features for the models to learn.

To enhance the accuracy of the EfficientNets models, we propose ensemble learning based on (1). Since both EfficientNetB6, EfficientNetB7, EfficientNetB6-Augm, and EfficientNetB7-Augm have their pros and cons in terms of the accuracies, we decided to combine any two of the algorithms, so that they may produce better results. For gallery 0° and probe 0° result, the two combined algorithms, EfficientNetB7-Augm+EfficientNetB6 and EfficientNetB7+EfficientNetB6 produce the highest accuracy. EfficientNetB7-Augm+EfficientNetB6-Augm has the highest accuracy score for gallery 0° and probe 180°. For gallery 180° and for both probes 0° and 180°, the EfficientNetB6-Augm+EfficientNetB7 is the best for both experiments as stated in Table II. Based on this experiment, the ensemble

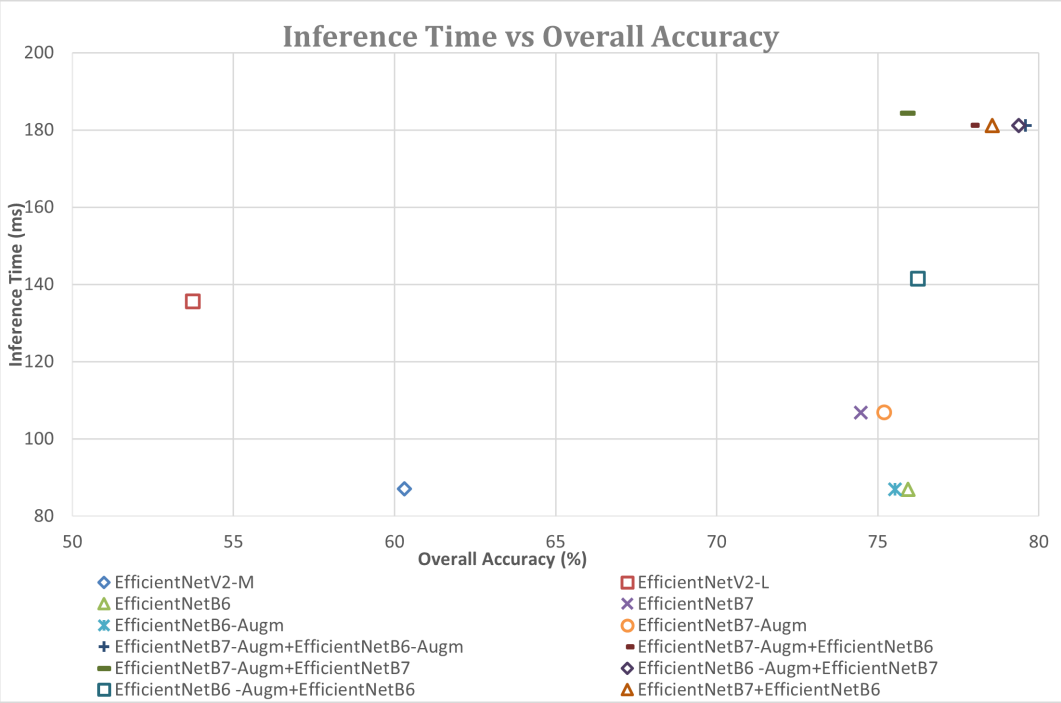


Figure 2. Inference Time (ms) vs Overall Accuracy (%)

of any of the two models produces better results than its constituent model performance.

Based on the evaluation of each gallery view, as presented in Table II, the EfficientNetB7-Augm+EfficientNetB6-Augm is the best for using 0° gallery view and for the one using 180° view, the combination of EfficientNetB6-Augm+EfficientNetB7 is the best. For the overall best performance by using both views, the EfficientNetB7-Augm+EfficientNetB6-Augm (79.59%) is the best and followed by EfficientNetB6-Augm+EfficientNetB7 (79.38%). So, for the effective ensemble in frontal view gait recognition, the EfficientNetB6 must be augmented and combined with EfficientNetB7 with or without augmentation. The reason for the better performance of the fusion algorithms is that, different deep learning models learn dissimilar complex nonlinear relationships between the inputs and outputs. The models' architecture response differently to the inputs, initial random weights, noise and hyperparameters in the training dataset. Hence, one model can produce good result on one type of inputs but failed on other types. By having the ensemble learning, the output is capable to produce the best results of both models.

Referring to the plot in Fig. 2, we want the point to be in the bottom right-hand corner of the plot. In this area, the accuracy is very high and the inference time is very low. The lowest inference time is EfficientNetB6 and EfficientNetB6-Augm with 87.01ms. The worst performance is EfficientNetV2-L (135.51ms). The EfficientNetB7 and EfficientNetB7-Augm have lower overall accuracy and slower recognition speed compared to both EfficientNetB6 and EfficientNetB6-Augm. Also, the ensemble of EfficientNetB6-Augm+ EfficientNetB6 produces better overall accuracy and recognition speed compared to EfficientNetB7-Augm+ EfficientNetB7.

All four proposed ensemble models of EfficientNetB6 combine with EfficientNetB7 produce very good accuracies but have the lowest inference speed. Hence, we can conclude that the EfficientNetB6 model is important in terms of producing high speed recognition and good accuracy. The inference time of a deep learning model is affected by the size of the models and the number of parameters used. Based on Table III, for single models, we can infer that the number of parameters is directly proportional to the inference time. This is also applicable if comparing is made within the fusion models. However, comparing the EfficientNetV2L, and other combined models except EfficientNetB7-Augm+EfficientNetB7, EfficientNetV2L with 119 million parameters has a lower inference time compared to those combined models (86.6 million and 111.0 million parameters). This is caused by the fusion algorithm computation.

The front and back views gait recognitions that are employed in this research are suitable for narrow corridor applications where cameras cannot be mounted from different view angles to capture the gait features. Furthermore, the advantage of the setting suggested in this research, only one camera is needed to capture the training data

TABLE III. Number of parameters and inference time

| Model | Parameters (Million) | Inference Time (ms) |
|---|---|---|
| EfficientNetB6 | 43.3 | 87.0 |
| EfficientNetB6-Augm | 43.3 | 87.0 |
| EfficientNetV2M | 54.4 | 87.1 |
| EfficientNetB7 | 66.7 | 106.8 |
| EfficientNetB7-Augm | 66.7 | 106.8 |
| EfficientNetV2L | 119 | 135.5 |
| EfficientNetB6-Augm+ EfficientNetB6 | 86.6 | 141.4 |
| EfficientNetB7 + EfficientNetB6 | 111.0 | 181.2 |
| EfficientNetB7-Augm + EfficientNetB6 | 111.0 | 181.2 |
| EfficientNetB6-Augm+ EfficientNetB7 | 111.0 | 181.2 |
| EfficientNetB7-Augm+ EfficientNetB7 | 133.4 | 184.3 |

either from the front or back view. The probe for the inference can also be captured from the front or back view. Figures 3 and 4 illustrate the setting in which the front or back view trained models can be used for the inference of front and back views. Therefore, the best algorithm for each category can be used together to be implemented in the settings. The EfficientNetB7+EfficientNetB6 and EfficientNetB7-Augm+EfficientNetB6 can be applied for front view training data with front view and back view inference. EfficientNetB7+EfficientNetB6 is chosen instead of EfficientNetB7-Augm+EfficientNetB6 because the augmentation requires more time to produce augmented images and also increases training time due to more images have to be trained. Since the EfficientNetB6 -Augm+EfficientNetB7 produces the best results, hence the algorithm is the most appropriate for the front and back views inference using the back view training data. Table IV shows the comparison of the proposed method with other methods found in the literature. Since the algorithms can be applied separately depending on their situations, hence we combine the results and tabulate them in Table IV. Comparing the results of the same views, the proposed methods have almost similar results to the other methods. On the other hand, for different view angles results, the proposed methods were outperformed by Işık and Ekenel [15] and Ben et al. [14]. For the Ben et al. [14] method, the same view results are not available. Işık and Ekenel [15] proposed a method that was implemented by using all the binary silhouettes in one gait sequence which is more than one gait cycle. Unlike the GEI which is based on the average of binary silhouettes in one gait cycle. The main problem with this method is if the

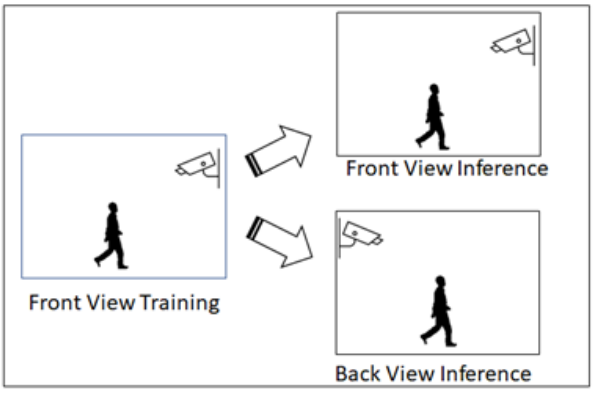


Figure 3. Applications of the proposed method: front view training for front and back views inference

architecture is trained with sequences of binary silhouettes and later when a camera captures shorter or longer gait sequences, it may not be able to generate similar gait features, hence producing inaccurate recognition. This kind of setting were not tested in their experiments. The method in [14] produces the best results but uses a smaller number of subjects for the probe compared to the experiment conducted in this paper. This is because the method uses different subjects for training and testing. The problem with this method is that it requires more than one view to produce the projection of other views. Hence, the method is not possible to be implemented, if only one view gait sequences are available for the training. This can happen especially in the narrow corridor situation. Hence, the method is not

TABLE IV. Comparison of the proposed methods with other state of the art methods in the literature

| Algorithms | View | 0° | 180° |
|---|---|---|---|
| Yu et al. [6] | 0 | 99.2 | 37.9 |
| | 180 | 41.1 | 99.6 |
| Proposed Methods | 0 | 98.8 | 56.7 |
| | 180 | 65.4 | 98.8 |
| Işık and Ekenel [16] | 0 | 100 | 90 |
| | 180 | 94 | 100 |
| Ben et al. [15] | 0 | - | 96.1 |
| | 180 | 96.4 | - |

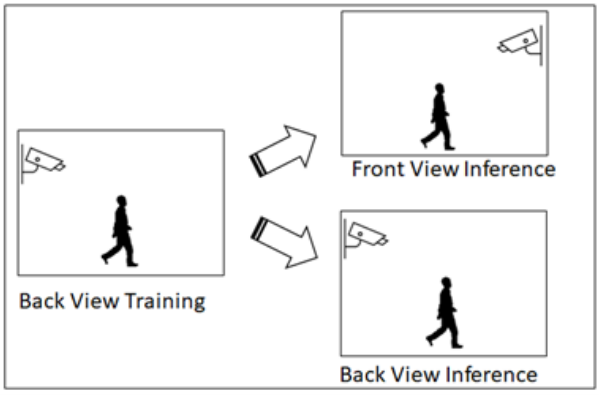


Figure 4. Applications of the proposed method: back view training for front and back views inference

fully suitable for narrow corridor applications. Furthermore, it requires several manually tuned parameters to perform the view projection.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, we present the front and back views for gait recognition algorithms using the best EfficientNets and EfficientNetV2 models. New ensemble learning algorithms based on the best results of the aforementioned models were created. The input images used by the models are raw GEI without any enhanced feature extraction.

Based on the experiment, EfficientNetB6 and EfficientNetB7 perform better than EfficientNetV2-M, and EfficientNetV2-L. The augmentation by horizontal flipping on the GEI input images has a mixture of results on the performance of both EfficientNetB6 and EfficientNetB7 but on average produces better results. In addition to that, in general, ensembled algorithms of EfficientNetB6 and EfficientNetB7 seem to have improved the recognition accuracies in all the experiments based on the matching between gallery 0° and probe 0°, gallery 0° and probe 180°, gallery 180° and probe 0°, and gallery 180° and probe 180°. If only one gallery view is used, for gallery view 0° (probe either 0° or 180°) the EfficientNetB7-Augm+EfficientNetB6-Augm is the best algorithm, but the EfficientNetB6-Augm+ EfficientNetB7performed the best for gallery view 180° (probe either 0° or 180°). The EfficientNetB7-Augm+EfficientNetB6-Augm is the best overall algorithm in terms of accuracy. However, the EfficientNetB6+EfficientNetB6-Augm and ensembled models do not perform well in terms of inference time.

Therefore, we can conclude that the EfficientNetB6-Augm is important for the front and back views gait recognitions using GEI. It can produce good accuracy and high recognition speed. The fusion of EfficientNetB6 with EfficientNetB7 either augmented or not can improve the performance with the drawback in terms of inference time.

However, the algorithms do not perform well for the cross-view gait recognition (gallery 0° and probe 180°, and gallery 180° and probe 0°), this is due to no cross-view enhancement algorithms applied to the GEI model. Also, for the cross-view performance, methods in the literature have higher accuracies than the proposed methods. However, they have several disadvantages in terms of practicality in implementing them in real applications. Therefore, future works can consider the cross-view algorithms to further enhanced the proposed methods.

The gait recognition research was carried out in this paper is based on front and back views with the normal walking pattern. Other covariates such as carrying objects, different types of clothing, shoes, and walking surfaces are not considered. Other problems such as occlusion if two people are walking together are also not experimented in this research. These covariates and problems require a new dataset and will also be considered in our future works

## ACKNOWLEDGEMENT

We would like to acknowledge Universiti Malaysia Sarawak (UNIMAS) through Research Innovation and Enterprise Centre (RIEC) for Special MyRA Assessment Funding, F02/SpMYRA/1721/2018, in funding this re-

search.

## References


[1] D. S. Matovski, M. S. Nixon, S. Mahmoodi, and J. N. Carter, "The effect of time on gait recognition performance," *IEEE transactions on information forensics and security*, vol. 7, no. 2, pp. 543–552, 2011.

[2] T. M. A. Zulcaffle, F. Kurugollu, D. Crookes, A. Bouridane, and M. Farid, "Frontal view gait recognition with fusion of depth features from a time of flight camera," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 4, pp. 1067–1082, 2018.

[3] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception & psychophysics*, vol. 14, pp. 201–211, 1973.

[4] T. Afendi, F. Kurugollu, D. Crookes, and A. Bouridane, "A frontal view gait recognition based on 3d imaging using a time of flight camera," in *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 2435–2439.

[5] 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.

[6] 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 international conference on pattern recognition (ICPR'06)*, vol. 4. IEEE, 2006, pp. 441–444.

[7] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, and L. Wang, "Recognizing gaits across views through correlated motion co-clustering," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 696–709, 2013.

[8] X. Xing, K. Wang, T. Yan, and Z. Lv, "Complete canonical correlation analysis with application to multi-view gait recognition," *Pattern Recognition*, vol. 50, pp. 107–117, 2016.

[9] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Support vector regression for multi-view gait recognition based on local motion feature selection," in *2010 IEEE Computer society conference on computer vision and pattern recognition*. IEEE, 2010, pp. 974–981.

[10] D. Muramatsu, Y. Makihara, and Y. Yagi, "View transformation model incorporating quality measures for cross-view gait recognition," *IEEE transactions on cybernetics*, vol. 46, no. 7, pp. 1602–1615, 2015.

[11] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, "Multiple views gait recognition using view transformation model based on optimized gait energy image," in *2009 IEEE 12th international conference on computer vision workshops, ICCV workshops*. IEEE, 2009, pp. 1058–1064.

[12] Z. Wu, Y. Huang, L. Wang, X. Wang, T. Tan, T. Liu, D. Tao, M. Song, S. Maybank, Y. Xiao *et al.*, "A comprehensive study on cross-view gait based human identification with deep cnns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1–1.

[13] X. Ben, C. Gong, P. Zhang, X. Jia, Q. Wu, and W. Meng, "Coupled patch alignment for matching cross-view gaits," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3142–3157, 2019.

[14] X. Ben, C. Gong, P. Zhang, R. Yan, Q. Wu, and W. Meng, "Coupled bilinear discriminant projection for cross-view gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 3, pp. 734–747, 2019.

[15] S. G. Işık and H. K. Ekenel, "Deep convolutional feature-based gait recognition using silhouettes and rgb images," in *2021 6th International Conference on Computer Science and Engineering (UBMK)*. IEEE, 2021, pp. 336–341.

[16] S. Zhang, Y. Wang, and A. Li, "Cross-view gait recognition with deep universal linear embeddings," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9095–9104.

[17] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 2169–2178.

[18] Z. He and Y. Shen, "A study of gait recognition algorithm based on multi-level pooling residual network," in *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*. IEEE, 2022, pp. 1059–1063.

[19] H. Chao, K. Wang, Y. He, J. Zhang, and J. Feng, "Gaitset: Cross-view gait recognition through utilizing gait as a deep set," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3467–3478, 2021.

[20] Z. Zhang, L. Tran, F. Liu, and X. Liu, "On learning disentangled representations for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 345–360, 2020.

[21] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[22] M. Tan and Q. V. Le, "Efficientnetv2: Smaller models and faster training," in *International conference on machine learning*. PMLR, 2021, pp. 10 096–10 106.

[23] Keras-Team, "Keras applications," 2022, https://www.keras.io/api/applications [Accessed: Jan. 1 2022].

[24] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, pp. 211–252, 2015.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[26] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, "Robust view transformation model for gait recognition," in *2011 18th IEEE international conference on image processing*. IEEE, 2011, pp. 2073–2076.

[27] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*. IEEE, 2008, pp. 722–729.

[28] O. M. Parkhi, A. Vedaldi, A. Zisserman, and C. Jawahar, "Cats and dogs," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3498–3505.

**Tengku Mohd Afendi Zulcaffle** received the B.Eng degree (Hons.) in Electronics and Computer Engineering in 1997, the M.Sc. degree in Intelligent System from Universiti Putra Malaysia in 2007, and the Ph.D. degree from Queen's University Belfast, U.K., in 2016. He is currently a Senior Lecturer with Universiti Malaysia Sarawak, Malaysia. His research interests include image and video processing, computer vision, and deep learning.

**Fatih Kurugollu** obtained BSc and MSc in Computer and Control Engineering degree from Istanbul Technical University, Turkey, in 1989 and 1994, respectively. He was awarded a PhD degree in Computer Engineering from the same university in 2000. He was employed as a research fellow by the Marmara Research Centre, which is the main governmental research unit of the Turkish Scientific Research Council (TUBITAK) in 1991. He joined the School of Electronics, Electrical Engineering and Computer Science at Queen's University, Belfast, UK, in 2000, initially as a Post-Doctoral Research Fellow. In 2003, he was appointed to a lectureship at the same department and later on was promoted to Senior Lecturer in Computer Science. He joined University of Derby, UK, as a Professor of Cyber Security in 2016. He has recently been appointed as a full Professor at University of Sharjah, UAE. His current research interests are centered around Security and Privacy in Internet-of-Things, Cloud Security, Imaging for Forensics and Security, Security related Multimedia Content Analysis, Big Data in Cyber Security, Homeland Security, Security Issues in Healthcare Systems, Biometrics, Image and Video Analysis. He has been the principal investigator and co-investigator of several projects funded by EPSRC, Royal Academy Engineering (RAEng), Leverhulme Trust, and Action Medical Research as well as the principal supervisor of several KTP projects funded by Innovate UK. He has supervised 11 PhD projects and he has authored more than 130 publications. He is a Senior Member of IEEE, a Member of the Associate College of Engineering and Physical Sciences Research Council (EPSRC), and a Fellow of the Higher Education Academy (HEA).

**Kuryati Kipli** received her B.Eng. degree in Electronic and Computer Engineering from the Universiti Malaysia Sarawak (UNIMAS), Malaysia, in 2004, the M.Sc. degree in Electronic and Computer Engineering from the University of Birmingham, U.K in 2007 and Ph.D degree Engineering at the School of Engineering, Deakin University, Australia in 2015. She is now a senior lecturer at Universiti Malaysia Sarawak. She has authored and coauthored more than 70 publications. Her current research interests include biomedical image processing and signal analysis artificial intelligence.

**Annie Joseph** received B.Eng. degree in Electronic and Electrical Engineering from College University Technology Tun Hussein Onn, Malaysia, in 2005. She received her M.S. degree in 2006 from University Science Malaysia. Then, she joined as a lecturer in Universiti Malaysia Sarawak under Department of Electronic, Faculty of Engineering in 2006. She then received Doctor of Engineering in Electrical and Electronic Engineering at Kobe University, Japan and promoted as a senior lecturer in 2014. Her research interests include online learning, concept drift, feature extraction and machine learning.

**David B.L. Bong** received his Bachelor of Electrical Engineering degree from Universiti Teknologi Malaysia, Master of Science degree in Computer Control Automation from Nanyang Technological University, Singapore, and PhD degree in Digital Signal Processing from Universiti Sains Malaysia. He is currently an Associate Professor with Universiti Malaysia Sarawak. His research interest is in digital image processing, computer vision and artificial intelligence.