Investigating the Impact of Occlusion on Gait-Based Age and Gender Estimation

Ryan Chin Zheng Qin1, a), Tee Connie 1, 2 b) and Michael Kah Ong Goh1, 2, c)

1Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia

2Centre for Image and Vision Computing, CoE for Artificial Intelligence, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia

Corresponding author: b) tee.connie@mmu.edu.my

a)1211307645@student.mmu.edu.my

c)michael.goh@mmu.edu.my

**Abstract.** Gait refers to an individual’s walking pattern. Everyone has a unique walking behaviour that is hard to disguise. Although gait recognition has seen many advances in recent years, occlusion is still a major problem affecting the performance. Occlusion is very common in real-life situations such as carried objects, clothing, obstacles, etc., and it is hard to avoid occlusion while capturing gait videos. This leads to lack of body information such as shape, pose and motion information. This paper aims to study the impact of different occlusions parts in gait recognition performance. Specifically, two different occlusions effects, namely block-wise and component-specific occlusion, are investigated. A self-collected dataset consisting of 715 side-walking GEIs from 120 subjects were synthetically occluded to simulate real-life scenarios. For block-wise occlusion, both vertical and horizontal block-wise occlusions are applied across the GEI silhouette. On the other hand, component-specific occlusion focuses on more detailed human body components such as head, shoulder, thigh, calf, etc. Experimental results show that the performances of non-occluded images have higher accuracy than those occluded images, confirming the detrimental effects of occlusion on gait performance. Besides, the size of occlusion also affects the gait performance. Within block-wise occlusion, age estimation accuracy is most severely affected (lowest accuracy of 29%) by occlusion of top area, which is around the shoulder part. For component-specific occlusion, gender estimation accuracy is mostly affected by head occlusion, while age estimation shows lower accuracy across all occluded body components. The study shows that, surprisingly, how we move our upper body regions, particularly the shoulders and head, is an important clue for estimating age and gender from walking, not just the leg movements.

# introduction

Gait generally refers to a person’s walking pattern. Gait features can be used in many applications such as age and gender estimation [1]. Gait has the advantage of being recognized from far away as compared to other biometric features. Besides, it can be authenticated from low-resolution images captured without subject cooperation. It is a widely used biometric technology that is hard for someone to disguise. Gait recognition has gained popularity in various applications such as surveillance, criminal investigation, and forensics [2],[3]. However, when it comes to practical applications, gait recognition faces several challenges. Covariates include carried objects, clothing, shoes, etc. and occlusion is the most common covariates involved when capturing subjects’ gait videos. Occlusion can be caused by obstacles such as trees and buildings or limited view when the subject moves out of view [2]. The presence of occlusion will significantly impact the performance of gait recognition since it will result in a lack of body information which increases the difficulty in extracting appropriate features.

In this study, the different impacts of occlusions are systematically analyzed. We investigate the effects of different types and locations of occlusion on the accuracy of age and gender estimation from gait patterns. To mimic real-life situations, a dataset consisting of 715 walking sequences from 120 subjects are created. Occlusions effects are syntactically applied on the gait energy images (GEIs) constructed from the video sequences. Two main types of occlusions are evaluated, namely block-wise and component-specific occlusion. In block-wise occlusion, the vertical or horizontal segments of the gait images are blocked out. On the other hand, the specific body parts like head, shoulder and calf are explicitly covered. This study provides insights into the robustness and limitations of current gait analysis techniques under realistic conditions.

In traditional methods, Awai et al. [5] developed a robust gait recognition method to improve the search for missing elderly people using the security camera to manage image occlusions. The work used pose estimation data and a random forest algorithm; it achieved 72.5% accuracy on the CASIA-B dataset. Meanwhile, Paul et al. [1] proposed a neural-network-based model to reconstruct occluded frame sequences from binary silhouettes that are extracted from RGB frames to improve the gait recognition accuracy. This work used dual LSTM models to reconstruct occluded gait frames from binary silhouettes, resulting in enhanced recognition accuracy with CASIA-B and OU-ISIRLP datasets, achieving high dice scores of 0.866 and 0.898 and Rank-1 accuracy at around 77%. Not only that, but deep learning methods are also widely used in gait recognition under occlusion. Gupta et al. [6] proposed a method to model intrinsic occlusion type awareness into any state-of-art of occluded gait recognition and design an auxiliary detection module to produce occlusion encodings that contains useful information for occluded gait recognition. The work used CNN-based models (GaitGL, GaitPart, GaitBase) and evaluated on GREW and BRIAR datasets. Similarly, Tsuji et al. [7] proposed a method to register silhouettes based on occlusion estimation related to the occluding elements to solve the occlusion problem where recognition accuracy drops without the silhouette registration as the full-body bounding boxes are often unavailable in real-world situations, so by using a CNN-based occlusion ratio estimator and a spatial transformer network for differentiable registration followed by a pairwise mask to handle residual variations, the method achieved a Rank-1 accuracy of 73.6% and EER of 1.45% on the OU-MVLP dataset.

# Proposed Method

The overall framework of this research starts by acquiring the gait videos, which are the raw videos of participants walking. After that, gait energy images (GEIs) are generated based on the gait video. Next, the GEIs are synthetically occluded to simulate real-life occlusion. The non-occluded and occluded GEIs are be used in feature extraction and classification during model training. The output is labelled as Male (0) and Female (1) for the gender model, and Child (0), Adult (1), and Senior (2) for the age model. Figure 1 shows the overall framework of this work.

A diagram of a person's body

AI-generated content may be incorrect.

**FIGURE 1*.*** Overall framework

## Data Collection

In total, there are 120 videos of people of different genders and ages walking 3 times each in 4 directions, which are front, back, left, and right. The gender includes male and female while the age is categorized as child (0-14 years old), adult (15-64 years old), and senior (>65 years old). The videos are segmented according to the walking directions, and only the side-walking part (left and right directions) are considered in this study. During video acquisition, the camera is placed at a position that can capture the subject’s whole body starting from the head to the toes. Next, the participants are required to walk in 4 directions which are front, back, left, and right, three times in each direction. Participants’ consents are obtained prior to the recording.

## Gait Energy Image (GEI)

After the gait videos are fully collected, each of the side-walking videos will be processed into a gait energy image. Gait refers to an individual’s walking manner or pattern and a GEI means the average of all silhouettes in one gait cycle. The GEI is generated by implementing a python code that takes the raw left-right walking edited videos as a batch of inputs, then it extracts individual frames using OpenCV. After that, each frame is segmented by using a pre-trained DeepLabv3 model from PyTorch where it detects pixels that represent ‘person’ and label it as ‘True’, else the background is label as ‘False’. A binary mask is generated to set the target ‘person’ to 255 and the background is 0, resulting in a series of binary frames. Next, the contours are extracted from the binary mask using OpenCV and to find the gait cycle, it will search for the initial frame to the ending frames that conclude a complete gait cycle. Equation (1) is the formula for generating a GE where *F* represents the total number of frames in a gait cycle and *If* represents the gait silhouette at the frame.

*f*(1)

To identify a complete gait cycle, the individual needs to walk two steps from the starting stance and must end at the same stance. These GEIs will be used in the model training and testing later.

## Occlusion Simulation

The GEI is synthetically occluded to simulate real-life occlusion scenarios and train the gait recognition model under such conditions. In this work, the GEIs are occluded in two broad ways, which are block-wise occlusion and component-specific occlusion. The objective is to observe the effects of occlusion to the gait performance and make discussions from the results.

### Block-Wise Occlusion

The block-wise occlusion is implemented to the GEIs, it is categorized as vertical and horizontal occlusions two part and is achieved by using cv2 functions. At first, load the GEIs from a specified folder using cv2.imread(), then return the coordinates of the silhouette region (non-zero pixel) using cv2.findNonZero(), pass the coordinates to the cv2.boundingRect() function to enclose the silhouette region with a rectangle. After that, the occlusion can be done by changing the pixels value of the specified range of height or width of the bounding rectangle to 0. For vertical occlusion, the width is divided into 4 parts equally and the occlusions are named as leftA, leftB, rightA and rightB. The leftA and rightA occlusions occluded ¼ of the width starting from the left and right while the leftB and rightB occlusions occluded ½ of the width from the left and right. The same applies to horizontal occlusion, where the height is divided into 5 equal parts, and each part is occluded individually. The occlusions are named topA, topB, topC, topD, and topE, where each occludes 1/5 of the height in a different position. In total, there are 9 types of occlusions occurring in different positions in the GEIs. Figure 2 first five images show the vertical occlusions, and the rest shows the horizontal occlusions.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A white light in the dark  AI-generated content may be incorrect. | A white light in the dark  AI-generated content may be incorrect. | A white light coming out of a door  AI-generated content may be incorrect. | A person standing in a dark room  AI-generated content may be incorrect. | A white light coming out of a hole  AI-generated content may be incorrect. | A white light in the dark  AI-generated content may be incorrect. | A white light in the dark  AI-generated content may be incorrect. | A white light in the dark  AI-generated content may be incorrect. | A white light in the dark  AI-generated content may be incorrect. |
| leftA | leftB | rightA | rightB | topA | topB | topC | topD | topE |

**FIGURE 2.** Block-Wise Occlusions

### Component-Specific Occlusion

The second type of occlusion applied to the GEIs is the component-specific occlusion. It is different from the block-wise occlusion as the occlusion part in the GEI is more specific than the block-wise occlusion. The component-specific occlusion is manually occluded using the Windows Paint application, it is achieved by using the shapes feature in paint, removing the outline and filling it with black colour, then just putting the shape on the specific part to occlude it. There are 9 parts in component-specific occlusion which are head, shoulder (s), right hand (rh), left hand (lh), butt, right thigh (rt), right calf (rc), left thigh (lt) and left calf (rc). Figure 3 shows the component-specific occlusions.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A white light coming out of a black background  AI-generated content may be incorrect. | A white light in the dark  AI-generated content may be incorrect. | A white light in the dark  AI-generated content may be incorrect. | A white light in the dark  AI-generated content may be incorrect. | A white figure with a black background  AI-generated content may be incorrect. | A white figure with a black square  AI-generated content may be incorrect. | A white light in the dark  AI-generated content may be incorrect. | A white light in the dark  AI-generated content may be incorrect. | A white light coming out of a black box  AI-generated content may be incorrect. |
| Head | Shoulder | Right-hand(rh) | Left-hand (lh) | Butt | Right-thigh (rt) | Right-calf (rc) | Left-thigh (lt) | Left-calf (lc) |

**FIGURE 3.** Component-specific occlusions

## Feature Extraction and Classification

The convolutional neural network (CNN) deep learning algorithm is implemented during model training in this age and gender estimation project. A CNN is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers, it is commonly used for image recognition and processing tasks. This work uses the CNN algorithm as a benchmark for the gait performance, focusing mainly on observing the impacts that occlusion brought rather than the method to train the models. The data is preprocessed before creating the CNN model. The GEIs are resized to (224,224). After that, normalization is performed to the [0,1] range. Each GEI filename is named as **subject\_1\_F\_58\_22072024\_left\_1\_gei\_topA**.

The convolutional layers operate on the input GEI images by using weighted kernels to detect features such as edges or complex patterns. It can learn subtle differences in gait patterns such as stride length, posture, balance, etc for age classification and hip movement, body structure, walking dynamics, etc. for gender classification. It slides a filter across the input data to produce a feature map. There are 3 convolutional layers applied in the models. The Equation (2) shows the computation for the Convolutional layer (Conv 2D) where *I*(*i*,*j*) represents the pixel value at position *I*(*i*,*j*) in the input image, *K*(*m*,*n*)represents the weight of the kernel at position(*m*,*n*) and *S*(*i*,*j*)denotes the output feature map at position (*i*,*j*).

(2)

Next, the LeakyRelu activation function is introduced after each convolutional layer to introduce non-linearity to the feature maps produced by the convolution, which enables the network to learn complex patterns. After that, three max pooling layers are employed to reduce the computational complexity in the network by downsampling the spatial dimension of the input. It works by selecting the maximum elements from the feature map within the kernel and outputs the most prominent features of the feature map, the output is then converted into a 1D vector by the flatten layer as the dense layer took 1D vector as input [8]. The two dense layers (connected layers) with Relu activation are then used to learn high-level features from the vectors. Dropout regularization was also performed to prevent overfitting by randomly removing some neurons during training. Lastly, the final dense layer makes predictions based on the features learned by previous layers. The activation function for the gender model is Sigmoid to perform binary classification, while the Softmax activation function is used in the age model to handle multi-class classification. The output will be either male/female for the gender model and child, adult, and senior for the age model [9]. The CNN settings for age and gender models are 3 convolutional layers, 3 pooling layers, 3 dense payers and 2 dropout layers. The activation function for the gender model is sigmoid and softmax for the age model.

# Experimental Results

First, we are going to see the results for the models trained on the non-occluded GEI. There are 715 GEIs from 120 subjects that were used, and augmentation is performed to increase the data, resulting in 4290 augmented GEIs. The train test split is set at 8:2, and the CNN settings for the gender model are 3 convolutional layers, 3 pooling layers, 3 dense layers, and 2 dropout layers. the model is trained with 15 epochs and achieved 85% test accuracy. On the other hand, the CNN settings for the age model are 3 convolutional layers, 3 pooling layers, 3 dense layers and 2 dropout layers also. The model is trained with 15epoch and achieved 80% test accuracy.

In block-wise occlusion, the models used 715 GEIs from 120 subjects in both training and testing. The training dataset consists of 4290 augmented non-occluded GEIs while the testing dataset consists of 715 block-wise occluded GEIs. There is a total of 9 types of block-wise occlusion tested on the CNN models. The CNN settings for both gender and age models are 3 convolutional layers, 3 pooling layers, 3 dense layers and 2 dropout layers. Both models are trained with 15 epoch. Table 1 shows the test accuracy for both gender and age models. The results show that the leftA and rightA occlusions yield higher accuracy compared to others because leftA and rightA occlusion size is relatively smaller compared to others which proves the effect of occlusion size on gait recognition performance. The accuracy of gender model in horizontal occlusion ranging from 57% to 65%, for some occlusion parts like topA and topB that covers important information like body posture and torso movement will make the accuracy drop. Next, the lowest accuracy is found under topB occlusion of the gender model, this implies that the upper torso and shoulder part causes the most severe impact on gender classification as older people to have rounded shoulder, this information are necessary to be included.

**TABLE 1.** Gender model test results (block-wise occlusion)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Part\_Based\_Occlusion | leftA | leftB | rightA | rightB | topA | topB | topC | topD | topE |
| Gender Model Test Accuracy (%): | 88 | 66 | 87 | 56 | 65 | 57 | 57 | 63 | 65 |
| Age Model Test Accuracy (%): | 81 | 67 | 83 | 61 | 62 | 29 | 60 | 60 | 63 |

For the component-specific occlusion, the dataset and CNN settings are the same as the block-wise occlusion experiment. There are also 9 types of component-specific occlusions implemented on each GEI, the occluded part is more specific compared to the block-wise occlusion. Table 2 shows the test accuracy for both gender and age models. The results show that the head occlusion gives the lowest accuracy for the gender model, while the calf parts produce higher accuracy, which means the calf component might be a region that provides discriminative information for gender classification. For the age model, the shoulder occlusion yields the lowest accuracy, which is similar to the block-wise occlusion findings, this indicates the shoulder is the most vital information for age estimation.

**TABLE 2.** Gender model test results (component-specific occlusion).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Component-Specific Occlusion | head | shoulder | rh | lh | butt | lt | lc | rt | rc |
| Gender Model Test Accuracy (%): | 63 | 73 | 79 | 68 | 62 | 75 | 85 | 73 | 84 |
| Age Model Test Accuracy (%): | 61 | 48 | 62 | 64 | 66 | 71 | 67 | 61 | 70 |

Overall, models that trained on occluded GEI data have lower test accuracy compared to models that use non-occluded GEI as occlusions result in information loss.

# Conclusion

Each individual’s gait feature is unique and can’t be replicated easily by others. Hence, this unique characteristic makes the gait feature widely used in many applications, but gait recognition faces several challenges with occlusion being the most common one. This work aims to observe the impacts of occlusion on gender and age estimation performances using gait features. To achieve this, the CNN deep learning age and gender models are implemented to train synthetically occluded GEIs, the occlusions include block-wise occlusion and component-specific occlusion with varied sizes on different body parts. From the result, we can tell that the model trained on non-occluded data can capture the features from GEI effectively, allowing the model to learn necessary information from the images that aid in classifying the subject’s age and gender. However, the differences are clear when the model used occluded GEIs to train and test, the accuracy drops compared to the non-occluded results, and the accuracy is impacted more when the occluded proportion is greater. In the block-wise occlusion, there are certain parts, like the upper torso and shoulder, which greatly affect the age model's accuracy. This implies that these areas are crucial features for the model to classify age groups. The results also show that the occlusion size matters in gait performance. If the occlusion portion is bigger, the accuracy will be lower. In the component-specific occlusion, the shoulder still remains as an important part for both age and gender estimation. In conclusion, this work explored the effects of different occlusions on gait recognition. By occluding different parts of the human body or occluding them with different sizes, the results will be affected depending on the occlusion size and which part is occluded. The results show that occlusion indeed affects the gait performance by hiding some information from the GEI from the models, preventing the models from extracting relevant data for estimation.

# References

1. Paul, A., Jain, M. M., Jain, J., and Chattopadhyay, P., “Gait Cycle Reconstruction and Human Identification from Occluded Sequences,” (2022).
2. Y.F. Ti, T. Connie, and M.K.O. Goh, “GenReGait: Gender Recognition using Gait Features,” Journal of

Informatics and Web Engineering **2**(2), 129–140 (2023).

1. V.W.S. Tan, W.X. Ooi, Y.F. Chan, C. Tee, and M.K.O. Goh, “Vision-Based Gait Analysis for Neurodegenerative

Disorders Detection,” Journal of Informatics and Web Engineering **3**(1), 136–154 (2024)

1. Hi, C. L., Lew, K. L., Zheng, C., Kurniawan, T., Babale, S. A., and Lee, C. S., “A Review of Camouflage Object Detection Techniques,” International Journal of Recent Advances in Science **7**(1), 84–92 (2025).
2. Awai, S., Chikano, M., and Konno, T., “Gait Recognition Using Occlusion Classification in Security Cameras,” in *Proceedings of the 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE)*, pp. 520–521, IEEE, Nara, Japan (2023).
3. Gupta, A., and Chellappa, R., “You Can Run but not Hide: Improving Gait Recognition with Intrinsic Occlusion Type Awareness,” in *Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5881–5890, IEEE, Waikoloa, HI, USA (2024).
4. Xu, C., Tsuji, S., Makihara, Y., Li, X., and Yagi, Y., “Occluded Gait Recognition via Silhouette Registration Guided by Automated Occlusion Degree Estimation,” in *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 3191–3201, IEEE, Paris, France (2023).
5. Gan, K. B., and Teoh, C. Y. E., “An Edge Convolution Neural Network Model for Plant Health Classification Using Camera,” International Journal of Recent Advances in Science **7**(1), 1–6 (2025).
6. Sui Lyn, H., Mansor, S., AlDahoul, N., and Abdul Karim, H., “Convolutional Neural Network-based Transfer Learning and Classification of Visual Contents for Film Censorship,” Journal of Emerging Technologies and Applications in Professional Practice **2**(2), 28–35 (2020).