

Gait Energy Image-Based Human Attribute Recognition Using Two-Branch Deep Convolutional Neural Network

Shaoxiong Zhang[✉], Yunhong Wang[✉], *Fellow, IEEE*, and Annan Li[✉], *Member, IEEE*

Abstract—Gait is an attractive biometric identifier, playing an essential role in addressing the issue of identity and attribute recognition in surveillance for its non-invasive and non-cooperative features. In this study, we propose a two-branch deep convolutional neural network for gait-based attribute recognition, including age estimation and gender recognition. We improve the estimation module by predicting a joint distribution instead of two independent distributions. In addition, several improvements are also proposed for improving the final performance of human attribute recognition, including data augmentation methods and loss functions. We implement several gait-based attribute recognition experiments on the OULP-Age and OUMVLP datasets. Experimental results show that the proposed method outperforms existing approaches. Finally, we elicit different body regions' contributions on attribute recognition tasks. Our conclusions can help improve the robustness of gait-based human attribute recognition systems in future.

Index Terms—Gait recognition, gait energy image, human attribute recognition, age estimation, convolutional neural network.

I. INTRODUCTION

GAIT is an essential biometric identifier that recognizes pedestrians by their body shape and the way they walk. In contrast to other biometrics such as fingerprint or iris, gait image can be captured at a distance using off-the-shelf sensors, which means low cost and low user cooperation. For these reasons, gait recognition has a wide range of applications in video surveillance. For example, prior arts demonstrated the feasibility of gait recognition in criminal investigation [1] and in court [2].

Besides identification, gait can also be used to recognize human attributes such as age and gender [3]. A gait-based attribute recognition method can predict human age (a precise age or an age-group) and gender from gait images without registration. Therefore, it enables a clue-based suspect retrieval in criminal investigation, where the identity of this suspect is still unknown.

Manuscript received 27 January 2022; revised 19 April 2022, 16 June 2022, and 4 August 2022; accepted 7 August 2022. Date of publication 31 August 2022; date of current version 22 December 2022. This work was supported by the Key Program of National Natural Science Foundation of China under Grant U20B2069. This article was recommended for publication by Associate Editor M. Nixon upon evaluation of the reviewers' comments. (Corresponding author: Annan Li.)

The authors are with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China (e-mail: zhangsx@buaa.edu.cn; yhwang@buaa.edu.cn; liannan@buaa.edu.cn).

Digital Object Identifier 10.1109/TBIOM.2022.3203149

Gait energy image (GEI) [4] and gait silhouette sequence are common input of a gait recognition system. A GEI is obtained by averaging silhouette sequence along the temporal dimension with at least one complete gait cycle and is proved effective on human identification [5], age estimation [6], and gender recognition [7]. Although silhouette sequence-based gait recognition models have achieved higher identification performances than GEI-based methods [8], GEI representation saves both storage space and computation time. Considering related studies on comparison between GEI and silhouette sequence on human attribute recognition are limited, we believe that GEI is still worth studying. Therefore, in this paper, we focus on GEI-based human attribute recognition.

The performance of gait-based attribute recognition is dramatically improved with the help of convolutional neural network (CNN). Our previous work [9] has shown that the error of gait-based age estimation can be reduced to 5.58 years using a CNN model on OULP-Age dataset [3], and this performance is much better than the methods using hand-crafted features.

Based on the deep neural networks, multi-task learning can be also utilized to improve the performance of attribute recognition, including both parallel and sequential multi-task frameworks [9], [10], [11]. In these works, a shared block is used for feature extraction, then task-specific layers are built for several attribute recognition tasks, i.e., gender recognition, age-group classification, and age estimation respectively. The effectiveness of multi-task learning frameworks indicates that the relationship between gait-based human attribute recognition tasks is not completely independent.

To further explore the relationship between gender recognition task and age estimation task, in this paper, we combine the above two recognition tasks by designing an estimation module, where a joint distribution is estimated instead of two independent distributions (see Figure 1). In addition, we construct another branch with an independent backbone network and classification module for age-group recognition. We believe that age-group labels can further help improving the performance of attribute estimation.

This work is an extension of our previous study [9], in which we first introduce deep neural network to gait-based attribute recognition. More specifically, we make the following contributions in this paper.

- We propose a deep learning method, which is referred as *TBResNet*, for gait-based attribute recognition. Compared

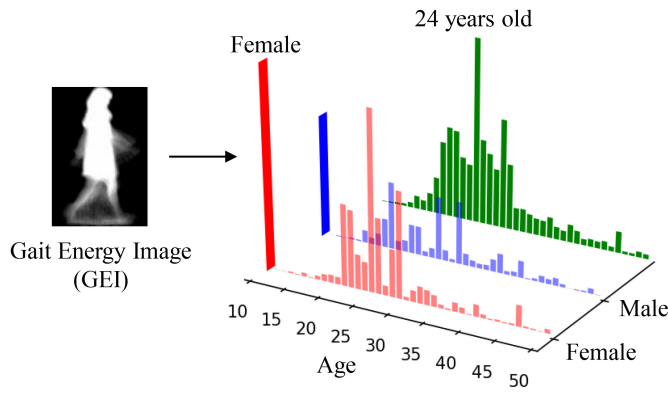


Fig. 1. GEI-based human attribute recognition. We combine the gender recognition task and the age estimation task with a joint distribution instead of two independent distributions.

to our previous framework [9], both network architecture and recognition module are improved.

- We investigate the performance of the proposed method on OULP-Age and OU-MVLP dataset. Results show that we achieve the state-of-the-art on both age estimation and gender recognition tasks.
- Extensive experiments are also conducted to elicit the contribution of different body regions on attribute recognition tasks. The conclusions can help improve the robustness of gait-based human attribute recognition systems in future.

The remainder of this paper is organized as follows. Related works are discussed in Section II. In Section III, we describe the details of proposed model. Experimental setting and results are shown in Section IV, while some discussions on ablation experiments and contributions of body regions are given in Section V. Finally, conclusions of this study are summarized in Section VI.

II. RELATED WORK

A. Gait Feature Extraction

Convolutional Neural Network (CNN) [12] has been widely used on computer vision tasks, and it also dramatically improve the performance of gait recognition. In general, CNN-based gait recognition methods can be grouped into model-based approaches and appearance-based approaches.

Model-based approaches usually use original RGB images to reconstruct the human body structure by using pose estimation [13], [14], [15] or parametric three-dimensional human models [16]. Model-based approaches heavily rely on the accuracies of pose estimation, and are generally robust to appearance variations, generally introduced by viewpoint or clothing changes [17].

Appearance-based approaches directly extract gait embeddings with CNN from GEI [5], [18], [19] and gait silhouettes sequences [8], [20], [21]. Compared to model-based approaches, appearance-based methods are more flexible and effective for their relatively low requirement on alignment. However, gait silhouettes are more sensitive to

appearance changes, including different clothing and carrying conditions [17].

In addition, the importance of local feature on gait recognition has attracted attention. Some works focus on part-based gait feature extraction [8], [20], [22], [23], in which gait images or embeddings are horizontally cropped for discriminative feature extraction. For example, Rida *et al.* [24] divides the horizontal motion of the human body into four parts. Huang *et al.* [25] defined dynamic region of gait energy images by calculating the heights of pelvis. Lishani *et al.* [26] defined two regions of interest representing the dynamic areas in gait energy image Zhang *et al.* [20] divided silhouettes into four horizontal parts by a learned partition and fed each horizontal part into an individual CNN. The above works inspire us that horizontal strips on the human body contribute differently to gait recognition. Although we do not apply the part-based method for learning embeddings in this work since gait-based attribute recognition is a more straightforward task than identification, we conduct extensive experiments to elicit the contributions of different body regions on attribute recognition tasks. The conclusions of extensive experiments may help improve the robustness of gait-based human attribute recognition systems in the future.

B. Gait-Based Attribute Recognition

1) *Gender Recognition*: Several gait-based gender recognition works without deep learning methods have been proposed in the past decade. Huang and Wang [27] extracted gait features by computing the ellipse parameters and applied Support Vector Machines to classify the gender. Yu *et al.* [28] combined human knowledge given by psychological experiments with image gait features to recognize the gender. They also found that hair, back, chest, and thigh components are more discriminative than others. Since gait is a dynamic biometric relevant to body shape, Hu *et al.* [29] tackled gait-based gender classification by integrating shape features and temporal Markov modeling. Lu *et al.* [30] discussed a sparse reconstruction-based metric learning method, which minimizes the intra-class reconstruction errors and maximizes the inter-class differences simultaneously. Some other works applied typical machine learning methods on gait template images. For example, Choudhary *et al.* [7] calculated five spatio-temporal parameters and concatenated them with dimensional reduced GEI for gender recognition. Liu *et al.* [31] regarded HOG characteristics of D-GEI as gait feature representation.

2) *Age Estimation*: Prior arts can be summarized into three categories: classification approaches, regression approaches, and deep learning approaches.

At early stages, human ages are usually divided into two or three age-groups, such as children, adults, and the elderly [32], [33], [34], where precise ages are not likely to be estimated from gait images. For example, Nabila *et al.* [35] proposed a silhouette projection model for age classification. Their experiments proved that arm swing, head pitch, stride width, and body size are discriminable parameters between young and senior adults.

Some other studies estimate precise age values from gait images. Lu and Tan [36] converted human age values into binary sequences and applied conventional multi-label learning methods for age estimation. Makihara *et al.* [37] constructed a large gait database including 1,728 subjects with ages ranging from 2 to 94 years and provided the Gaussian process regression algorithm for age estimation. Li *et al.* [6] combined both classification and regression approaches and proposed an age-group-dependent gait-based human age estimation method.

With the help of deep learning frameworks, recent studies greatly improved the performance on gait-based age estimation tasks. Ordinary convolutional neural networks [11], [19], [38], [39], deep residual network [9], generative adversarial network [40] and dense convolutional network [41] are proposed to estimate human age. These works apply end-to-end deep learning methods for gait feature extraction and attribute recognition and perform better than non-deep learning methods. Recently, Xu *et al.* [42] proposed a method to estimate age from a single frame, which is more suitable for an application with real-world scenarios.

3) *Multi-Task Learning*: Multi-task learning has proven effective in many computer vision tasks, such as facial detection [43]. Therefore, multi-task learning is also utilized to improve attribute recognition performance based on deep learning frameworks. Marin-Jimenez *et al.* [10] concluded that both the accuracy of identification task and the convergence speed of training are increased by combining gait-based tasks. Our previous work [9] presented that the error can be reduced by 0.11 years (from 5.58 to 5.47 years) when we select gender recognition as an auxiliary task on the OULP-Age dataset. Zhu *et al.* [38] also found that the gender information indeed improves the performance of gait-based age estimation. Instead of conventional parallel multi-task learning framework, Sakata *et al.* [11] proposed a sequential multi-stage deep network for three tasks: i.e., gender estimation, age-group estimation, and age estimation respectively, which inspires us that the age-group estimation task may contribute to the precise age estimation task.

III. METHOD

The framework of our proposed method is shown in Figure 3, which is referred as *TBResNet*. TBResNet consists of an age-group classification (AGC) branch and an attribute estimation branch. Compared to our previous method [9], in which a single branch of the backbone network was proposed, we add another branch for age-group classification. We hold the idea that as an auxiliary task with an independent backbone network, age-group classification task can improve the final performance of the age estimation task. In addition, some minor improvements are also utilized. The details are presented below.

A. Gait Energy Image

Gait energy image (GEI) [4] is an averaged image of gait silhouettes with a duration of at least one gait cycle, and it is also well known for its robustness and simplification. Prior art has confirmed that GEI can be used for both identification [5], [18]

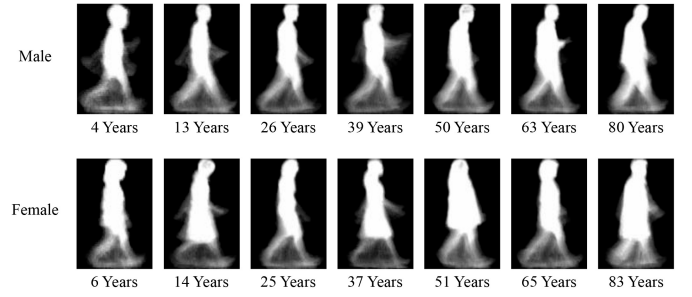


Fig. 2. Exemplar GEIs of male and female subjects on the OULP-Age dataset. It is hard for a human to estimate their exact age from a GEI.

and age estimation [9], [38]. Although Xu *et al.* [39] demonstrated that silhouette-based backbone model outperforms their GEI-based model on identification task, we still select GEI as input data on attribute recognition task for the following three reasons: (i) GEI representation saves both storage space and computation time; (ii) GEI brings enough appearance information for attribute recognition task, and (iii) only size-normalized GEI is provided on the OULP-Age dataset [3]. GEI can be calculated by

$$GEI(x, y) = \frac{1}{N} \sum_{t=1}^N B_t(x, y), \quad (1)$$

where $B_t(x, y)$ is a binary gait silhouette image at time t , and N is the number of frames in the complete cycle of a gait silhouette sequence. Figure 2 shows exemplar GEIs with different ages and genders. The size of original GEI provided by the OULP-Age dataset is $128 \times 88 \times 1$. To facilitate pre-trained backbones, we duplicate single color channel into three and resize them by a bicubic interpolation over 4×4 pixel neighborhood in OpenCV. Consequently, the resulting size of input GEI is $224 \times 224 \times 3$.

B. Data Augmentation

Data augmentation is an important technique used to increase the amount of data by slightly modifying existing data on deep learning-based models because fewer data and complex models may lead to overfitting, limiting the performance of deep models. In this work, we adopt two common data augmentation methods, random rotation and random erasing [44], to enlarge the original gait dataset.

The random rotation technique randomly rotates an input image by some number of degrees. In general, rotating the pedestrian image does not affect human discrimination on its attribute recognition. Therefore, the random rotation technique can increase the variation.

The *Random Erasing* method randomly selects a rectangle region in an image with probability p , then erases pixels in it by random values. To simplify the implementation, we use zero in this work. We believe that a generalized model should be able to recognize human attributes when input gait images have been rotated, as well as been partly masked, due to redundant information on human gait images. For example, gender can be recognized from both hairstyle and body movement. Without random erasing, the model tends to only focus on

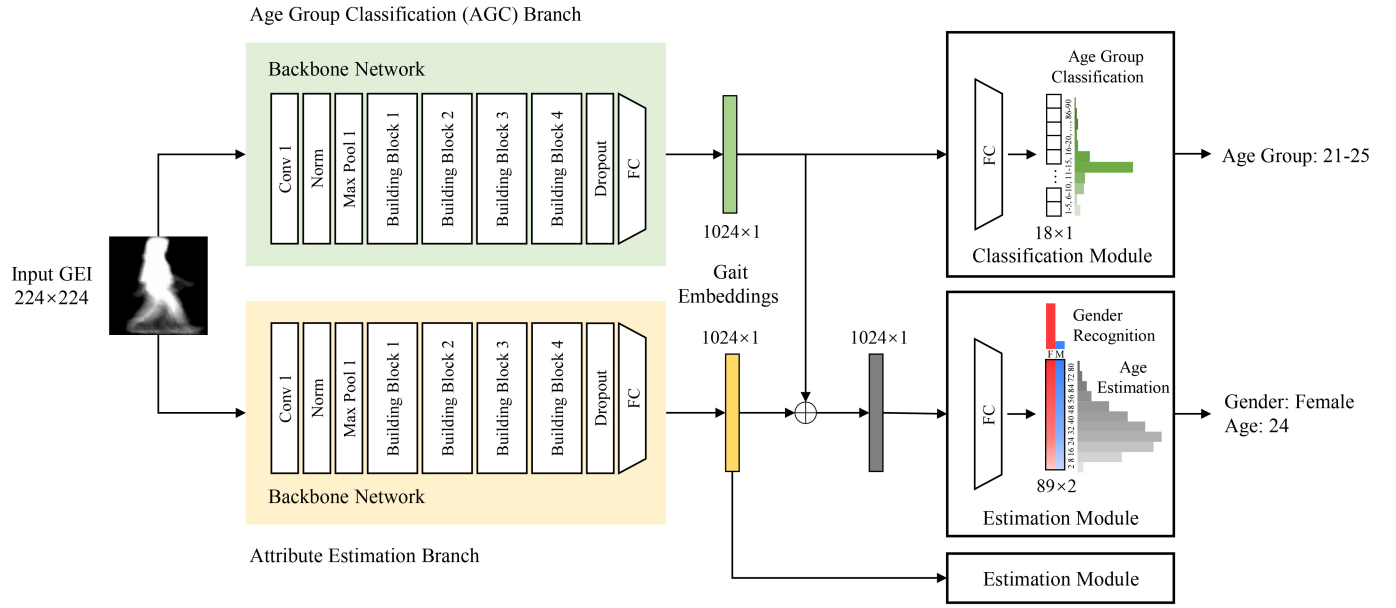


Fig. 3. Framework of proposed method. The entire model can be divided into two branches. The AGC branch is used for age-group classification, while the other branch is for attribute estimation. The input GEI is sent into two individual branches, and gait embeddings are summed together for final attribute estimation.

hairstyle rather than movement, which increases the risk of overfitting.

C. Backbone Network

ResNet is a popular backbone due to its astounding performance on computer vision tasks. Considering that we have achieved great performance by extracting gait features with ResNet in our previous work [9], we continue to use pre-trained ResNet-18 [45] as a backbone network for gait feature extraction. Since our network has two branches, we construct two independent backbone networks. Each branch extracts gait embeddings for different tasks: i.e., age-group classification and attribute estimation, respectively.

D. Age-Group Classification Module

Sakata *et al.* [11] proposed a sequential multi-stage deep network. Three tasks are considered in their network, i.e., gender estimation, age-group estimation, and age estimation respectively. They proved that the proposed multi-stage network outperforms the benchmarks on age estimation. Inspired by their work, we also apply age-group classification as an auxiliary task to improve the performance of precise age estimation. However, instead of a multi-stage network, we design a multi-branch network.

As described in [11], the disadvantage of a sequential multi-stage network is the errors introduced by previous estimators, which have a great impact on the next-step recognition tasks. Considering this weakness, we decide to use a parallel network. On the other hand, we do not follow the design of a conventional multi-task network, in which different tasks share the same backbone network, because the age-group classification task and precise age estimation task are not independent. Age-group labels can be directly calculated from the integer

age labels. If we simply add an age-group classifier paralleling the precise age estimator, the features learned by the backbone network may not be improved and the performance on age estimation also cannot be improved either.

In this work, we add an independent branch for the age-group classification, named AGC branch. AGC branch consists of a backbone network and a classification module, where the backbone network does not share weights with the other backbone network in the main branch (Attribute Estimation Branch in Figure 3). The two gait embeddings after these two independent backbone networks are sent into two recognition modules with two different tasks. For the Attribute Estimation Branch, the task is precise age estimation and it is more difficult, thus the backbone network learns more discriminative embeddings, while ACG Branch learns simpler embeddings. These two complementary embeddings are summed together to make the final age estimation better.

As shown in Figure 3, the learned gait embedding is sent into a simple fully connected network. Following Lu and Tan [36], we divide age into 18 groups by the five-year-intervals, i.e., 1-5, 6-10, ..., 86-90. The fully connected network with output size (18×1) represents the result of age-group classification. The cross-entropy loss $\mathcal{L}_{AgeGroup}$ is selected as the optimization criterion for age-group classification.

E. Recognition Module

In our previous work [9], we directly estimate age value and gender distribution independently from gait embeddings. The results encouraged our opinion that these two tasks may be closely relevant. In this work, we combine these two estimation tasks and estimate their joint probability mass function instead of two independent distributions. Since there is a criterion applied to joint distribution estimation, as a result, all

parameters contribute to both tasks, and two tasks are combined into a single task. This combination reduces the conflict between two separated tasks and therefore can further improve performance of the attribute recognition. The cost of task combination is dimension increasing. Since gender recognition is a two-category classification task, combining age and gender brings a small dimension increase, which is still acceptable for performance increasing.

Formally, for a subject i and x is its gait embeddings extracted from the backbone network. Our estimation module generates a joint probability mass function $\mathbf{p}_{A,G}$ for both age and gender, and then estimates the age value $\hat{age}(x)$ and gender value $\hat{gender}(x)$ from marginal probability mass function \mathbf{p}_A and \mathbf{p}_G by

$$\begin{aligned}\hat{age}(x) &= \sum_{a \in \mathbb{A}} a \cdot \mathbf{p}_A(A = a) \\ &= \sum_{a \in \mathbb{A}} a \cdot \sum_{g \in \mathbb{G}} \mathbf{p}_{A,G}(A = a, G = g)\end{aligned}\quad (2)$$

and

$$\begin{aligned}\hat{gender}(x) &= \sum_{g \in \mathbb{G}} g \cdot \mathbf{p}_G(G = g) \\ &= \sum_{g \in \mathbb{G}} g \cdot \sum_{a \in \mathbb{A}} \mathbf{p}_{A,G}(A = a, G = g),\end{aligned}\quad (3)$$

where \mathbb{A} is defined as age value set $\{2, 3, \dots, 90\}$ and \mathbb{G} is gender value set $\{0, 1\}$. $g = 0$ and $g = 1$ indicate female and male, respectively.

The loss function for recognition module consists of three components, i.e., the joint distribution loss \mathcal{L}_{Joint} , the gender recognition loss \mathcal{L}_{Gender} , and the age estimation loss \mathcal{L}_{Age} respectively. \mathcal{L}_{Joint} is used to constrain the joint probability mass function of age and gender so that it will converge to the true distribution. In this work, a simple ℓ_1 loss is used to measure the mean absolute error between the estimated and the ground truth,

$$\mathcal{L}_{Joint} = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{p}}_{A,G} - \mathbf{p}_{A,G}\|, \quad (4)$$

where N is the batch size, and $\|\cdot\|$ is ℓ_1 -norm.

Furthermore, we also apply this criterion for both \mathbf{p}_A and \mathbf{p}_G . Considering that we achieved superior performance on the gender recognition task in our previous work [9], we still choose the cross-entropy loss \mathcal{L}_{Gender} as the optimization criterion for gender recognition.

As for gender estimation from \mathbf{p}_G , we choose *Earth Mover's Distance (EMD)* as the metric between two distributions following [38], where the cross-entropy loss \mathcal{L}_{CE} is also added. The EMD loss is proposed for single-label classification [46], which uses the predicted probabilities of all classes and penalizes the miss-predictions according to a ground distance matrix that quantifies the dissimilarities between classes. The EMD loss \mathcal{L}_{EMD} can be calculated by

$$\mathcal{L}_{EMD} = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K (CDF_k(\mathbf{p}_A) - CDF_k(\hat{\mathbf{p}}_A))^2, \quad (5)$$

where $CDF(\cdot)$ is a function that calculates the cumulative density function of an input, $CDF_k(\cdot)$ is the k -th element of $CDF(\cdot)$, and N is the batch size.

Therefore, the criterion for age estimation \mathcal{L}_{age} can be written as

$$\mathcal{L}_{age} = \alpha \mathcal{L}_{EMD} + \mathcal{L}_{CE}, \quad (6)$$

where α is a weight to balance these two functions. In this work, we set $\alpha = 1$. Finally, the total loss for the whole model is written as

$$\mathcal{L} = \mathcal{L}_{AgeGroup} + \mathcal{L}_{Joint} + \mathcal{L}_{Gender} + \mathcal{L}_{Age}. \quad (7)$$

F. Age Label Smoothing

Considering the uniqueness of gait-based age recognition task, local smoothing technique is applied in this work. There are two reasons for this improvement. Since body shape and movement habits may stay stable for several years, it is hard to predict his exact age. Furthermore, the ground truth age label is not accurate. When answering our age, we might remove the decimals of our age value. For example, a man who will be 26 years old in the next month may answer 25 for his age in a questionnaire.

We apply an approximate local Gaussian window $w = [0.05, 0.25, 0.4, 0.25, 0.05]$ to smooth the original one-hot age label \mathbf{p}_A by

$$\tilde{\mathbf{p}}_A = \mathbf{p}_A \circledast w, \quad (8)$$

where \circledast indicates 1-D convolution with zero padding, and $\tilde{\mathbf{p}}_A$ is a smoothed age label.

IV. EXPERIMENTS

A. Dataset

The experiments are conducted on two popular gait datasets with attribute labels, OULP-Age dataset [3] and OU-MVLP dataset [47].

OULP-Age dataset is a large dataset for human gait analysis with age and gender labels. There are 63,846 subjects from 2 to 90 years old. For each subject, a GEI of size 88×128 from a side view is provided shown in Figure 2. Since the training and test set division is already given, we follow their experimental protocol [3]. Furthermore, we selected the last 3,000 subjects from the training set as a validation set for better evaluation. Therefore, we have 28,923 subjects for training, 3,000 subjects for validation, and 31,923 subjects for final testing.

OU-MVLP dataset is the largest cross-view gait dataset, which contains 10,307 subjects (5,114 males and 5,193 females) with 14 views ($0^\circ, 15^\circ, \dots, 90^\circ, 180^\circ, 195^\circ, \dots, 270^\circ$) per subject and two sequences per view with ages ranging from two to 87 years old. We follow the experimental protocol proposed in [42], in which 5,153 subjects are selected for training and the other 5,154 subjects for test. Furthermore, we select the last 300 subjects from the training set as a validation set similar to the OULP-Age dataset.

B. Implementation Details

For data augmentation parameters, we set $p = 0.9$ in the random erasing algorithm, which means the probability of being erased for each input GEI is 90%. We also set the area ratio range of erasing region $s_l = 0.02$ and $s_h = 0.4$; min aspect ratio $r_1 = 0.3$. All training images are randomly rotated with a degree less than 15° , either clockwise or counterclockwise.

We initialize the weights of backbone network by a pre-trained ResNet-18 on the ImageNet Dataset and initialize the weights and bias of fully connected layers by normal initialization with a standard deviation equaling 0.01. In optimization, the Adam algorithm [48] was implemented with a start learning rate 10^{-4} .

For the OULP-Age dataset, we trained each model for 200 epochs with batch size equaling 128 while calculating the performance on the validation set for each epoch. After 50 epochs, the learning rate is reduced to 1/10 of the original ones. For the OU-MVLP dataset, the learning rate is reduced after 5 epochs. Finally, the model with the best validation performance during training is selected as the final model for evaluation on the test set.

C. Evaluation

Mean absolute error (MAE) and cumulative score (CS) are evaluation metrics for age estimation [39]. MAE is defined as the average absolute error between the estimated ages and ground truth ages. MAE can be calculated as

$$MAE = \frac{1}{N} \sum_{i=1}^N |a\hat{g}e_i - age_i|, \quad (9)$$

where N is the total number of test subjects, age_i is the ground truth age, and $a\hat{g}e_i$ is the estimated age for test sample i . CS is defined as the rate of test subjects whose absolute error is less than or equal to a given number. Given N test subjects, n_k is the number of test subjects for which $|a\hat{g}e_i - age_i| \leq k$, then $CS(k)$ can be calculated as

$$CS(k) = \frac{n_k}{N}. \quad (10)$$

For gender recognition, a simple correct classification rate (CCR) is calculated for evaluation.

D. Comparison

We compare our proposed method with the state-of-the-art methods on the OULP-Age dataset in Table I, including conventional methods and deep learning methods. The MAEs of all conventional methods are larger than 6.5 years, while deep learning methods reduce MAE to less than 6 years. Our proposed method outperforms all previous deep learning methods on age estimation and gender recognition tasks. Compared with our previous work, ResNet with multi-task framework [9], our new method can reduce MAE by 0.61 years (from 5.47 to 4.86 years) and raise CCR of gender recognition by 0.54% (from 98.10% to 98.64%), which proves the effectiveness of our improvement.

It should be noted that the uncertainty-aware model [39] is also evaluated GaitSet backbone [8], in which silhouette

TABLE I
COMPARISON ON THE OULP-AGE DATASET. GENDER RECOGNITION PERFORMANCE IS ALSO LISTED FOR WORKS WITH MULTI-TASK FRAMEWORK METHODS. AGE ESTIMATION RESULTS OF NON-DEEP LEARNING METHODS ARE CITED FROM [39]

Method	Age				Gender CCR
	MAE	CS(1)	CS(5)	CS(10)	
ITML [30], [49]	-	-	-	-	94.39%
MLG [36]	10.98	16.7	43.4	60.8	-
GPR (k = 10) [37]	8.83	9.1	38.5	64.7	-
GPR (k = 100) [37]	7.94	10.5	43.3	70.2	-
GPR (k = 1000) [37]	7.3	10.7	46.3	74.2	-
SVR (Linear) [50]	8.73	7.9	38.2	67.6	-
SVR (Gaussian) [50]	7.66	9.4	44.2	73.4	-
OPLDA [51]	8.45	7.7	37.9	67.6	-
OPMFA [51]	9.08	7.0	34.9	64.1	-
ADGMLR [6]	6.78	18.4	54.0	76.2	-
ResNet [9]	5.47	-	-	-	98.10%
DenseNet [41]	5.79	22.5	55.9	80.4	-
Multi-stage [11]	5.84	25.3	62.6	82	-
Joint CNN [19]	6.27	-	-	-	-
ODR-GLCNN [38]	5.06	-	-	-	97.80%
Uncertainty-aware [39]	5.43 (5.41 ^b)	23.5	61.7	82.5	-
+ GaitSet ^a [39]	5.01 (4.91 ^b)	25.9	65.1	84.4	-
TBResNet (Ours)	4.86	35.1	68.3	86.2	98.64%

^a Gait silhouettes are used, which provide additional information than GEI.
^b Performance with the augmented training set.

TABLE II
COMPARISON ON THE OU-MVLP DATASET. THE THREE METHODS LISTED BELOW TAKE DIFFERENT IMAGES AS INPUT DATA; THUS, IT IS HARD TO COMPARE THESE MODELS DIRECTLY

Method	Input	Age				Gender CCR
		MAE	CS(1)	CS(5)	CS(10)	
Single Image [42]	Single frame	8.39	15.8	48.0	68.4	94.27%
+ GT cycle [42]	Silhouettes	6.63	-	-	-	96.04%
TBResNet (Ours)	GEI	6.70	20.57	55.87	76.59	96.71%

images were used as input data for age estimation. In this experimental setting, it achieved performances of 4.98 years error (without augmented data) and 4.91 years error (with augmented data) on age estimation (second row from the bottom in Table I). Since silhouette images provide additional temporal information and they have been proved more effective than GEI [8], it is hard to compare these two performances with others directly. Nonetheless, our proposed method using GEI achieves a better performance on the age estimation task.

We also compare our proposed method with the Single Image model [42] on the OU-MVLP dataset in Table I. In general, the MAE of proposed method is 6.70 years for age estimation, and CCR is 96.71% for gender recognition, which is worse than the performance on OULP-Age dataset because the OU-MVLP dataset is more challenging since the GEIs are sampled from 14 different views rather than a single side-view. Compared to the Single Image model [42], our method outperforms its single frame version since GEI which generated from a sequence contains more information. Our method also achieves a similar age estimation performance and a better gender recognition performance than the GT

TABLE III
ABLATION EXPERIMENTS ON THE OULP-AGE DATASET. MAE AND CCR INDICATE THE PERFORMANCE OF AGE ESTIMATION AND GENDER RECOGNITION, RESPECTIVELY. RR: RANDOM ROTATION. RE: RANDOM ERASING. AGC: AGE-GROUP CLASSIFICATION

Data Augmentation	AGC Branch	Backbone Architecture	Age-group Classification	Age and Gender	Age Label Smoothing	Performance	
						MAE	CCR
–	✓	ResNet-18	✓	Joint	✓	5.37	96.39%
RE	✓	ResNet-18	✓	Joint	✓	5.35	97.92%
RR	✓	ResNet-18	✓	Joint	✓	5.31	97.75%
RR + RE	–	ResNet-18	–	Joint	✓	4.94	98.53%
RR + RE	–	ResNet-18	✓ ^a	Joint	✓	4.94	98.47%
RR + RE	✓	ResNet-18	–	Joint	✓	4.94	98.58%
RR + RE	–	ResNet-34	–	Joint	✓	4.96	98.58%
RR + RE	✓	ResNet-18	✓	Only Age Separated ^b	✓	5.00	–
RR + RE	✓	ResNet-18	✓		✓	4.95	98.59%
RR + RE	✓	ResNet-18	✓	Joint	–	4.89	98.58%
RR + RE	✓	ResNet-18	✓	Joint	✓	4.86	98.64%

^a Age-group classification loss is combined into the attribute estimation branch.

^b Age estimation task and gender recognition task are regarded as two separated tasks, rather than joint distribution estimation.

cycle version [42]. It should be pointed out that our model uses GEI, which is actually a single image and just a simple average of a sequence. Even though, a competitive age estimation performance and a better gender recognition performance compared to existing silhouette sequence-based model can be achieved. It supports our opinion that GEI-based method for gait recognition is still worth studying. In general, we can conclude that the proposed method is the state-of-the-art on gait-based human attribute recognition.

E. Ablation Studies

Results of ablation experiments are listed in Table III. We consider data augmentation, branch and network architecture, and multi-task loss function of our proposed methods.

This work applies two data augmentation methods: random rotation (RR) and random erasing (RE). When both data augmentation methods are applied, the MAE reduces from 5.37 years to 4.86 years, showing the most improvement in age estimation. This result indicates that while a deep CNN model works well on a gait dataset, data augmentation can further avoid over-fitting and improve the performance. In addition, the validation set is crucial in the gait-based age estimation task because early stopping can be applied according to the performance on the validation set, which also helps prevent over-fitting.

To evaluate the impact of proposed AGC branch, we conduct ablation experiments. If we remove the AGC branch, or remove age-group classification task, the age estimation performance will drop. In addition, we replaced the backbone network by ResNet-34 without AGC branch and achieved similar performance. Since the parameter number of ResNet-34 is almost twice as ResNet-18, this counterevidence shows that the performance improvement is not simply derived from the increase of network parameters. We can see that supervised by age-group label, the AGC module can predict the age-group information and help the whole system achieving better age estimation performance.

We also evaluated the performance of without gender recognition and with separated tasks. If we only consider the age

estimation task without gender information, the performance is 5.00 years error. If we use a conventional multi-task framework and regarded gender recognition as an auxiliary task, the performance improves 0.05 years. This result shows that gender information improves the performance of age estimation, which is consistent with the previous findings [9], [38]. After we improved the separated two tasks, age estimation and gender classification, to a joint estimation, the age estimation performance improves by 0.09 years (from 4.95 to 4.86). In summary, the ablation results show that all components of our proposed method contribute to the final gait-based age estimation performance.

V. DISCUSSION

A. Distribution of Estimated Attribute

Figure 4 shows a scatter plot between the ground truth age and the estimated ages for proposed TBResNet, where the results on gender recognition are also marked with different colors. We observe that the deviation of age estimation is slight for children under ten years old, but the accuracy of gender recognition is lower than adults. This observation is because there is a strong correlation between a child's age and body height, but the gender of a child is not evident in appearance. As age increases, the variance for estimated age increases. It is difficult for the elderly over 60 years old to estimate their ages. The estimation results are usually less than the ground truth ages. In addition, older women have a greater probability of being predicted to be the wrong gender. We believe this is due to the fact that the number of elderly samples is small. Thus, it is hard to learn the gait features of the elderly from a small amount of GEI in the OULP-AGE dataset.

Figure 5 shows the distribution of age estimation errors and the gender recognition accuracy among different age-groups. The age estimation error on the 0-5 years age-group is stable because it is easy to classify little children correctly. In the age-group 6-10 and 11-20 years age-group, the deviation of age estimation is still slight (see the light blue violin plot in Figure 5), but the biggest errors of age estimation reach about

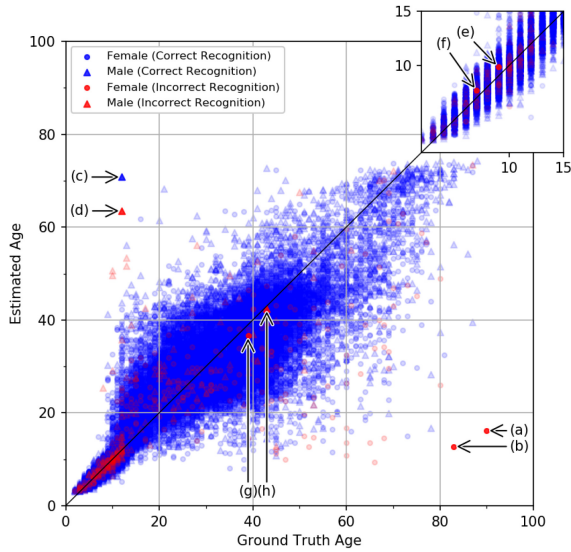


Fig. 4. Scatter plots between the ground-truth age and estimated age for proposed TBResNet. Each point is a subject from the test set on the OULP-Age dataset. Shape indicates to ground truth gender, and color indicates whether the gender recognition is correct. The details of subject (a) to (h) are listed in Table V.

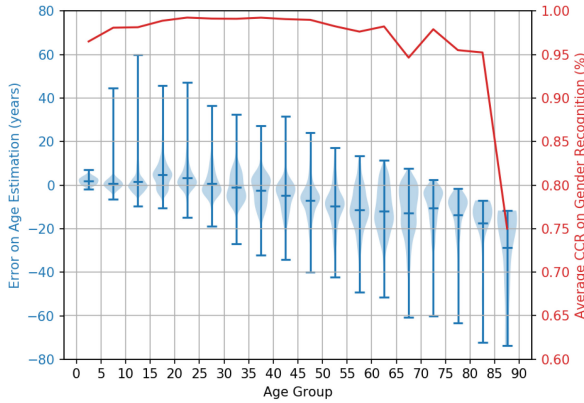












Fig. 5. Distribution of age estimation and the gender recognition accuracy among different age-groups on the OULP-Age dataset. The x-axis shows 18 age-groups with five-year intervals (i.e., 1-5, 6-10, ..., 86-90). The blue violin plot shows errors between the estimated age and the ground truth age in each age-group, including the error distribution, the extreme, and the means. The red curve indicates the average correct classification rate of gender recognition in each age-group.

40 years. The mean error becomes negative when the age is older than 30 years. The absolute error in age estimation is within 10 years for most samples until age increases to 55 years. For the elders older than 55 years, the deviation of age estimation becomes large, and the average CCR on gender recognition also drops obviously.

B. Contribution of Body Region

Previous gait-based age and gender recognition works use full human body images for recognition. Then, some gait identification works have focused on part-based methods [20], [52], [53]. For local information extraction, they split a human body image or a gait feature map into several horizontal strips. This step is effective in gait recognition,

TABLE IV
AVERAGE PERFORMANCES OF AGE ESTIMATION (MAE) AND GENDER RECOGNITION (CCR) WITH DIFFERENT HUMAN BODY REGIONS ON OULP-AGE DATASET

Two Regions		Four Regions		
Upper Body				Head Region
	7.41, 95.98%	9.18, 93.37%	5.76, 94.97%	
				
Lower Body				Pelvic Region
	6.59 , 93.39%	12.05, 82.99%	5.43, 97.89%	Limb Region
				
		7.81 , 86.27%	6.27, 97.33%	

which indicates that different body regions do not contribute equally to recognition. Inspired by the above works, we conduct experiments on the contributions of different human body regions.

We horizontally split the human body into different regions: upper body and lower body for a two-parts-division; head region, chest region, pelvic region, and limb region for a four-parts-division following [52]. Then we use masked GEI [54] generated by masking out (i.e., set to zero) the specific region to re-evaluate our model. Since we have applied random erasing data augmentation methods during training, our model can recognize human attributes from masked GEI. We list the performances in Table IV.

From Table IV, if we mask the lower body and keep the upper body for each subject in the test set, MAE is 7.41 years for age estimation, and CCR is 95.98% for gender recognition. On the other hand, if we mask the upper body, MAE is 6.59 years, and CCR is 93.39% for age and gender recognition, respectively. This result shows that we can get a better age estimation performance from the lower body than the upper body, while better gender recognition performance from the upper body.

From results of the four-parts-division, we achieve 7.81 years for age estimation with only limb region, which is the best performance in all of the four regions. If the limb region is masked, age estimation performance drops dramatically to 6.27 years (from 4.86 years of the full image), the worst age estimation performance among four masked regions. Therefore, we conclude that the lower body, especially the limb region, provides more information for age estimation than other regions.

However, things are different on gender recognition. The Head region is the most critical region to gender recognition since there is little performance degradation if we mask

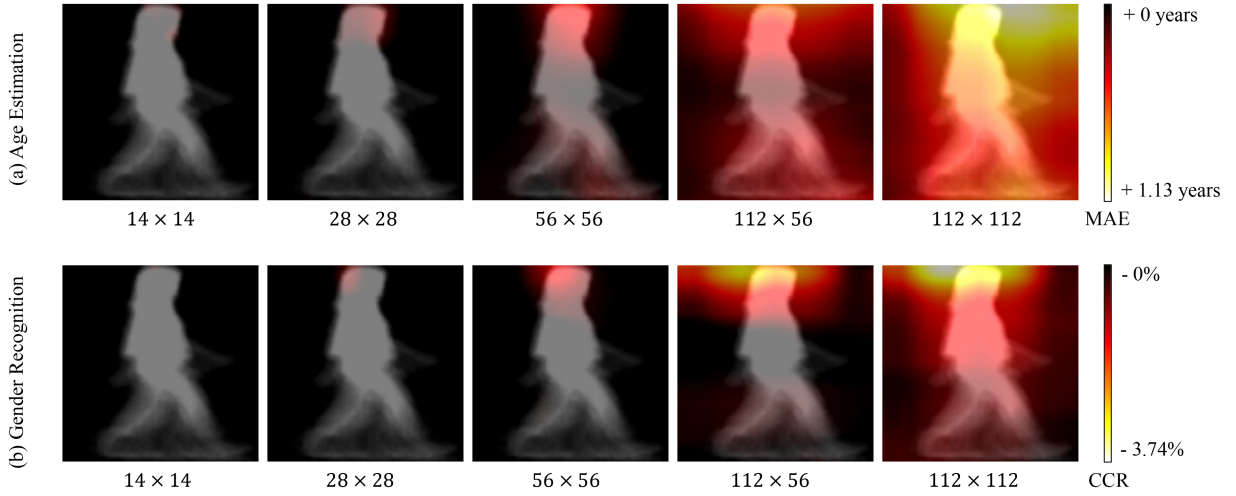


Fig. 6. Visualization of contributions for each local portion on age estimation and gender recognition tasks on the OULP-Age dataset. The first 1,000 subjects in the test set are used instead of the complete set to reduce calculation costs. The size of GEI is 224×224 , while the mask size is listed below each heatmap, and the sliding step size is 2. MAE and CCR are used to evaluate the contribution of each portion on (a) age estimation and (b) gender recognition, respectively. The color on each pixel shows the average performance degradation if this pixel is masked.

chest, pelvic, and limb regions. This conclusion is consistent with human intuition because we often recognize gender from hairstyle and head shape from a silhouette.

C. Contribution of Local Region

In this section, we conducted experiments to show the contributions of each local portion. Compared to body regions described in the last section, the local portion indicates a minor part of the human body (e.g., hair, an arm, or a foot).

To evaluate the contributions for each local portion, masked GEI is utilized again. Instead of masking an entire region, in this experiment, we use a small fixed-size mask that slides horizontally and vertically over a GEI with a specific step size. We estimate age and gender from this masked GEI with our pre-trained model at each mask position. Finally, for each pixel in a GEI, average performance is calculated from the attribute recognition performances where the mask covers this pixel. The averaged performance on each pixel shows the contribution of this pixel to the final attribute recognition task because the performance will drop if a critical portion is masked. We generate heatmaps to visualize the contribution of different local portions (see Figure 6), which help us understand which human body portion contributes to attribute recognition.

From Figure 6, the average performance drops dramatically when the mask size becomes large. On gender recognition task, our recognition system is more sensitive to the hairstyle. This observation is easy to understand since hairstyle is an important feature that recognizes the human gender. Meanwhile, leg movement contributes less to gender recognition. Average performance degradation on gender recognition is hard to be observed when a mask of size 112×56 covers the human leg. On age estimation, average performance degradation can be observed when the human leg is masked with size 56×56 and 112×56 , a significant difference from gender recognition. The fact revealed in this experiment implies that leg movement contributes to age estimation.









Although our experiments are conducted on datasets in a controlled environment, we believe that our findings still contribute to robust gait recognition systems in other scenarios. For example, in a high-reliability requirement scenario, since a gait-based gender recognition system is more sensitive to the head region than other body regions, wearing a hat should not be allowed. If a pedestrian wears a hat, the accuracy of gender recognition will reduce drastically. In addition, our study also contribute to multi-modal biometric systems, which make use of multiple biometric traits for recognition [55] and improving the overall performance by adjusting the weight of single modalities. If the head region of a pedestrian is covered, reducing the weight of gait recognition may reduce the negative impact on overall performance on the multi-modal biometric systems.

D. Failure Cases

Figure 4 shows a scatter plot between the ground truth age and the estimated ages for proposed TBResNet on the OULP-Age dataset. To better demonstrate the performance of our proposed model, we select several subjects (subjects (a) to (h) in Figure 4) in the test set of OULP-Age dataset and show their GEI and information in Table V, including both age estimation and gender recognition failure cases.

In Table V, subjects (a) and (b) are two older women. Our model fails to predict their ages and genders. Subjects (c) and (d) are two children and are both 12 years old. Our model also fails to estimate their ages. One possible explanation is that a few pedestrians' appearances and walking habits are not consistent with those of the same age or gender, which leads to recognition failure. However, we believe that the original RGB videos are essential to a detailed interpretation. Subjects (e) to (h) are four cases with wrong gender recognition results. These failures are acceptable because it is also difficult for humans to recognize their gender.

TABLE V
EXAMPLES OF GAIT-BASED AGE ESTIMATION AND GENDER
RECOGNITION RESULTS ON THE OULP-AGE DATASET. THE TOP FOUR
ROWS SHOW SUBJECT IDS, GEIS, GROUND TRUTH (GT) GENDER AND
AGE LABELS. THE LAST TWO ROWS LIST THE ESTIMATED (EST.)
GENDER (WITH PROBABILITIES) AND AGE (WITH ERRORS) LABELS.
UNDERLINE INDICATES WRONG ESTIMATION RESULTS.
M: MALE; F: FEMALE

	(a) ID20481 	(b) ID64194 	(c) ID59369 	(d) ID50749 
GT	Female 90 Years	Female 83 Years	Male 12 Years	Male 12 Years
EST.	<u>M</u> (84.8%) 16.2 (-73.8)	<u>M</u> (88.9%) 12.8 (-70.2)	M (99.9%) 70.8 (+58.8)	F (97.4%) 63.5 (+51.5)
	(e) ID56071 	(f) ID24691 	(g) ID16886 	(h) ID45812 
GT	Female 9 Years	Female 7 Years	Female 39 Years	Male 43 Years
EST.	<u>M</u> (98.8%) 9.9 (+0.9)	<u>M</u> (58.9%) 7.7 (+0.7)	<u>M</u> (94.9%) 36.6 (-2.4)	F (83.2%) 42.5 (-0.5)

VI. CONCLUSION

This study proposes a deep convolutional neural network for gait-based human attribute estimation. Several improvements are proposed, including data augmentation methods and recognition modules, to improve the final performance further. The proposed method is trained and evaluated on the OULP-Age and OU-MVLP datasets. Our experiments prove that our model outperforms the previous ones on the age estimation and gender recognition tasks. The limitation of our proposed framework is that it may only work on gender and age estimation tasks because that joint distribution representation is hard to be applied to other tasks due to dimension increase and independence among other tasks.

We also conduct experiments on contributions of the human body region and local portion. The fact revealed in this paper implies that the head region is the most critical region for gender recognition. More specifically, hairstyle is essential to recognize human gender rather than leg movement. As for age estimation, leg movement provides more information than other regions. These conclusions can help improve the robustness of gait-based human attribute recognition systems in the future.

Since the OULP-Age dataset and the OU-MVLP dataset are captured using a chroma key, i.e., the green screen, models trained on such datasets are difficult to directly adopt to real-world surveillance video. In such scenarios, background clutter, occlusion, and complex weather conditions lead to low pedestrian silhouettes quality. It can be expected that the performance degradation will still exist in a real-world

scenario. Our future work will focus on tackling the real-world scenario challenge, which includes both dataset construction and model renovation.

REFERENCES

- [1] D. Muramatsu, Y. Makihara, H. Iwama, T. Tanoue, and Y. Yagi, "Gait verification system for supporting criminal investigation," in *Proc. Asian Conf. Pattern Recognit.*, 2013, pp. 747–748.
- [2] I. Macoveciuc, C. J. Rando, and H. Borrión, "Forensic gait analysis and recognition: Standards of evidence admissibility," *J. Forensic Sci.*, vol. 64, no. 5, pp. 1294–1303, 2019.
- [3] C. Xu, Y. Makihara, G. Ogi, X. Li, Y. Yagi, and J. Lu, "The OU-ISIR gait database comprising the large population dataset with age and performance evaluation of age estimation," *IPSI Trans. Comput. Vis. Appl.*, vol. 9, no. 1, pp. 1–14, 2017.
- [4] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.
- [5] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep CNNs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 209–226, Feb. 2017.
- [6] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, "Gait-based human age estimation using age group-dependent manifold learning and regression," *Multimedia Tools Appl.*, vol. 77, no. 21, pp. 28333–28354, 2018.
- [7] S. Choudhary, C. Prakash, and R. Kumar, "A hybrid approach for gait based gender classification using gei and spatio temporal parameters," in *Proc. Int. Conf. Adv. Comput., Commun. Inform.*, 2017, pp. 1767–1771.
- [8] H. Chao, Y. He, J. Zhang, and J. Feng, "Gaitset: Regarding gait as a set for cross-view gait recognition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 8126–8133.
- [9] S. Zhang, Y. Wang, and A. Li, "Gait-based age estimation with deep convolutional neural network," in *Proc. Int. Conf. Biometr.*, 2019, pp. 1–8.
- [10] M. J. Marín-Jiménez, F. M. Castro, N. Guil, F. De la Torre, and R. Medina-Carnicer, "Deep multi-task learning for gait-based biometrics," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 106–110.
- [11] A. Sakata, N. Takemura, and Y. Yagi, "Gait-based age estimation using multi-stage convolutional neural network," *IPSI Trans. Comput. Vis. Appl.*, vol. 11, no. 1, p. 4, 2019.
- [12] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2010, pp. 253–256.
- [13] W. An *et al.*, "Performance evaluation of model-based gait on multi-view very large population database with pose sequences," *IEEE Trans. Biom., Behav., Ident. Sci.*, vol. 2, no. 4, pp. 421–430, Oct. 2020.
- [14] R. Liao, S. Yu, W. An, and Y. Huang, "A model-based gait recognition method with body pose and human prior knowledge," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107069.
- [15] X. Li, Y. Makihara, C. Xu, and Y. Yagi, "End-to-end model-based gait recognition using synchronized multi-view pose constraint," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 4106–4115.
- [16] X. Li, Y. Makihara, C. Xu, Y. Yagi, S. Yu, and M. Ren, "End-to-end model-based gait recognition," in *Proc. Asian Conf. Comput. Vis.*, 2020, pp. 1–17.
- [17] A. Sepas-Moghaddam and A. Etemad, "Deep gait recognition: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Feb. 15, 2022, doi: [10.1109/TPAMI.2022.3151865](https://doi.org/10.1109/TPAMI.2022.3151865).
- [18] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Geinet: View-invariant gait recognition using a convolutional neural network," in *Proc. Int. Conf. Biometr.*, 2016, pp. 1–8.
- [19] Y. Zhang, Y. Huang, L. Wang, and S. Yu, "A comprehensive study on gait biometrics using a joint CNN-based method," *Pattern Recognit.*, vol. 93, pp. 228–236, Sep. 2019.
- [20] Y. Zhang, Y. Huang, S. Yu, and L. Wang, "Cross-view gait recognition by discriminative feature learning," *IEEE Trans. Image Process.*, vol. 29, pp. 1001–1015, 2019.
- [21] C. Xu, Y. Makihara, X. Li, Y. Yagi, and J. Lu, "Cross-view gait recognition using pairwise spatial transformer networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 260–274, Jan. 2021.
- [22] Y. Huang, Y. Liang, Z. Han, and M. Du, "Two-stream convolutional network extracting effective spatiotemporal information for gait recognition," in *Proc. Int. Conf. Security, Pattern Anal. Cybern.*, 2019, pp. 43–48.

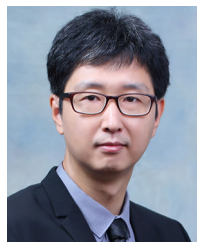
- [23] C. Fan *et al.*, “Gaitpart: Temporal part-based model for gait recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 14225–14233.
- [24] I. Rida, X. Jiang, and G. L. Marcalis, “Human body part selection by group lasso of motion for model-free gait recognition,” *Signal Process. Lett.*, vol. 23, no. 1, pp. 154–158, 2015.
- [25] J. Huang, X. Wang, and J. Wang, “Gait recognition algorithm based on feature fusion of GEI dynamic region and Gabor wavelets,” *J. Inf. Process. Syst.*, vol. 14, no. 4, pp. 892–903, 2018.
- [26] A. O. Lishani, L. Boubchir, E. Khalifa, and A. Bouridane, “Human gait recognition using GEI-based local multi-scale feature descriptors,” *Multimedia Tools Appl.*, vol. 78, no. 5, pp. 5715–5730, 2019.
- [27] G. Huang and Y. Wang, “Gender classification based on fusion of multi-view gait sequences,” in *Proc. Asian Conf. Comput. Vis.*, 2007, pp. 462–471.
- [28] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu, “A study on gait-based gender classification,” *IEEE Trans. Image Process.*, vol. 18, pp. 1905–1910, 2009.
- [29] M. Hu, Y. Wang, Z. Zhang, and D. Zhang, “Gait-based gender classification using mixed conditional random field,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 5, pp. 1429–1439, Oct. 2011.
- [30] J. Lu, G. Wang, and P. Moulin, “Human identity and gender recognition from gait sequences with arbitrary walking directions,” *IEEE Trans. Inf. Forensics Security*, vol. 9, pp. 51–61, 2014.
- [31] T. Liu, B. Sun, M. Chi, and X. Zeng, “Gender recognition using dynamic gait energy image,” in *Proc. Inf. Technol., Netw., Electron. Autom. Control Conf.*, 2017, pp. 1078–1081.
- [32] Y. Makihara, H. Mannami, and Y. Yagi, “Gait analysis of gender and age using a large-scale multi-view gait database,” in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 440–451.
- [33] B. K. Y. Chuen, T. Connie, O. T. Song, and M. Goh, “A preliminary study of gait-based age estimation techniques,” in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2015, pp. 800–806.
- [34] N. Mansouri, M. A. Issa, and Y. B. Jemaa, “Gait features fusion for efficient automatic age classification,” *IET Comput. Vis.*, vol. 12, no. 1, pp. 69–75, 2018.
- [35] M. Nabila, A. I. Mohammed, and B. J. Yousra, “Gait-based human age classification using a silhouette model,” *IET Biometr.*, vol. 7, no. 2, pp. 116–124, 2018.
- [36] J. Lu and Y.-P. Tan, “Gait-based human age estimation,” *IEEE Trans. Inf. Forensics Security*, vol. 5, pp. 761–770, 2010.
- [37] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi, “Gait-based age estimation using a whole-generation gait database,” in *Proc. Int. Joint Conf. Biometr.*, 2011, pp. 1–6.
- [38] H. Zhu, Y. Zhang, G. Li, J. Zhang, and H. Shan, “Ordinal distribution regression for gait-based age estimation,” *Sci. China Inf. Sci.*, vol. 63, no. 2, pp. 1–14, 2020.
- [39] C. Xu *et al.*, “Uncertainty-aware gait-based age estimation and its applications,” *IEEE Trans. Biom., Behav., Ident. Sci.*, vol. 3, no. 4, pp. 479–494, Oct. 2021.
- [40] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, “Make the bag disappear: Carrying status-invariant gait-based human age estimation using parallel generative adversarial networks,” in *Proc. IEEE Int. Conf. Biometr. Theory, Appl. Syst.*, 2019, pp. 1–9.
- [41] A. Sakata, Y. Makihara, N. Takemura, D. Muramatsu, and Y. Yagi, “Gait-based age estimation using a densenet,” in *Proc. Asian Conf. Comput. Vis.*, 2018, pp. 55–63.
- [42] C. Xu *et al.*, “Real-time gait-based age estimation and gender classification from a single image,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2021, pp. 3460–3470.
- [43] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, “Facial landmark detection by deep multi-task learning,” in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 94–108.
- [44] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random erasing data augmentation,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 13001–13008.
- [45] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [46] L. Hou, C.-P. Yu, and D. Samaras, “Squared earth mover’s distance-based loss for training deep neural networks,” 2016, *arXiv:1611.05916*.
- [47] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, “Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition,” *IPSN Trans. Comput. Vis. Appl.*, vol. 10, no. 1, pp. 1–14, 2018.
- [48] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, *arXiv:1412.6980*.
- [49] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, “Information-theoretic metric learning,” in *Proc. Int. Conf. Mach. Learn.*, 2007, pp. 209–216.
- [50] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.
- [51] J. Lu and Y.-P. Tan, “Ordinary preserving manifold analysis for human age and head pose estimation,” *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 2, pp. 249–258, Mar. 2013.
- [52] H. Aggarwal and D. K. Vishwakarma, “Covariate conscious approach for gait recognition based upon Zernike moment invariants,” *IEEE Trans. Cogn. Devel. Syst.*, vol. 10, no. 2, pp. 397–407, Jun. 2018.
- [53] B. Lin, S. Zhang, and X. Yu, “Gait recognition via effective global-local feature representation and local temporal aggregation,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 14648–14656.
- [54] X. Li, Y. Makihara, C. Xu, D. Muramatsu, Y. Yagi, and M. Ren, “Gait energy response functions for gait recognition against various clothing and carrying status,” *Appl. Sci.*, vol. 8, no. 8, pp. 1380–1401, 2018.
- [55] A. Jain, K. Nandakumar, and A. Ross, “Score normalization in multimodal biometric systems,” *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, 2005.



Shaoxiong Zhang received the B.S. and M.S. degrees from Beihang University, Beijing, China, in 2014 and 2017, respectively, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering. His research interests include computer vision, pattern recognition, and gait recognition.



Yunhong Wang (Fellow, IEEE) received the B.S. degree in electronic engineering from Northwestern Polytechnical University, Xi'an, China, in 1989, and the M.S. and Ph.D. degrees in electronic engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1995 and 1998, respectively. She was with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 1998 to 2004. Since 2004, she has been a Professor with the School of Computer Science and Engineering, Beihang University, Beijing, where she is also the Director of Laboratory of Intelligent Recognition and Image Processing, Beijing Key Laboratory of Digital Media. Her current research interests include biometrics, pattern recognition, computer vision, data fusion, and image processing.



Annan Li (Member, IEEE) received the B.S. and M.S. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2003 and 2006, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2011. He worked as a Scientist with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore, and a Postdoctoral Research Fellow with the National University of Singapore. He currently works with the School of Computer Science and Engineering, Beihang University. His research interests include computer vision, pattern recognition, and statistical learning.