Make the Bag Disappear: Carrying Status-invariant Gait-based Human Age Estimation using Parallel Generative Adversarial Networks

Xiang Li[1,2] Yasushi Makihara[2] Chi Xu[1,2] Yasushi Yagi[2] Mingwu Ren[1]
[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
[2] The Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan
{lixiangmzlx, xuchisherry}@gmail.com {makihara, yagi}@am.sanken.osaka-u.ac.jp
renmingwu@mail.njust.edu.cn

Abstract

*Existing approaches to gait-based human age estimation seldom consider variations such as carried objects, which greatly alter appearance of gait features (*e.g.*, gait energy images) and result in poor age estimation results. Therefore, we propose a method of gait-based human age estimation robust against carrying status using generative adversarial networks. Specifically, we consider a generative network that outputs a gait feature without carried objects (*i.e.*, it makes the carried objects disappear) given an input gait feature with or without carried objects, and then learns the parallel generative network with shared weights from a pair of gait features of the same subject with and without carried objects. Thereafter, the generated gait features are separately fed into the same subsequent age regression network for age estimation, which is trained in an end-to-end manner in conjunction with the parallel generative network. Because of this design, the proposed method can generate gait features without carried objects that are suitable for gait-based age estimation regardless of the carrying status of the input gait features, and therefore improve the age estimation accuracy under carrying status variations. Experimental results on a large gait dataset with carried objects show the state-of-the-art performance of the proposed method.*

1. Introduction

Gait is a unique biometric that can be captured at a distance from cameras without any cooperation from subjects. Although gait-based studies mainly focus on person authentication or recognition [23, 26], gait-based person attribute estimation (*e.g.*, age [13, 15, 17, 19, 31], gender [34], and ethnicity [35]) is another meaningful aspect. Of the person attribute estimation methods, gait-based age estimation is of great importance. Compared with face-based age estimation [3, 5], it has its own advantages especially in surveillance scenarios where captured facial images may be of low-resolution or even covered by a mask, while gait can still be well captured under these conditions. Therefore, gait-based age estimation could provide many potential applications in surveillance such as finding wandering children or elderly through surveillance camera footage and the automatic control of public entrances to certain areas with restricted age ranges.

The human body's biological and kinematic features, such as stride length, stride frequency, and head-to-body ratio, have clear changes during growth. Some approaches [4, 10] have used these gait features as clues to predict age information. For example, Davis [4] demonstrated the differences in stride frequency of children and adults. Ince *et al*. [10] used the head-to-body ratio to classify children and adults. On the other hand, appearance-based gait features, such as gait energy images (GEIs) [8] (also known as average silhouettes [14]), are widely used for real surveillance scenes where human bodies are usually captured with low resolutions. They are therefore considered to be effective gait representations in conjunction with some manifold learning techniques [16, 17] to learn a low-dimensional discriminant subspace, classification-based approaches [15], regression-based approaches [13, 17, 19], or the recent deep learning-based approaches [20] for gait-based human age estimation.

Although the GEI provides convincing clues for gait-based human age estimation in [13, 19, 31] (*i.e.*, there are obvious change in head-to-body ratio during a human's growth and clear appearance of middle-aged spread and stoop after people get older), it is very easily affected by many covariates such as carried objects (COs) that largely

978-1-7281-1522-1/19/$31.00

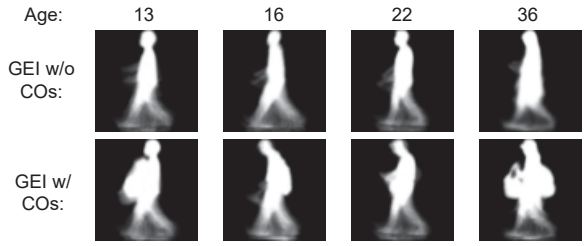

Figure 1. Examples of gait energy images with some typical types of carried objects. Each column indicates the same subject.

alter the aforementioned characteristics and degrade the age estimation accuracy. Figure 1 shows some typical types of COs that affect the appearance of the GEI. As we can see, younger people may have a rounder body shape that is similar to the middle age spread when they carry COs at the front, or they may have stoop similar to that of older people when they carry COs on their back. Alternatively, they could even have almost all their body shape covered by multiple COs. All these cases significantly increase the difficulty of gait-based human age estimation.

Inspired by the great success of the recently generative adversarial networks (GANs) for generating realistic images [9, 32, 33], the generation of GEI without any COs from source GEI with COs is a straightforward yet reasonable approach to deal with CO variations for gait-based human age estimation. Therefore, we propose a method based on GANs for gait-based human age estimation robust against CO variations. Although some GAN-based works like GaitGAN [33] and MGANs [9] have been proposed for cross-view gait recognition, they cannot be directly applied to our task of gait-based age estimation. The main contributions of this paper are summarized as follows:

1) A unified parallel GAN-based method for carrying status-invariant gait-based human age estimation.

Unlike GaitGAN [33] and MGANs [9], which learn view-invariant or view-specific features using only one input feature, the proposed method uses a parallel GAN-based network to learn carrying status-invariant features from a pair of gait features of the same subject with and without COs. Moreover, we incorporate an age regression network into the generative network for age estimation and train the whole network in an end-to-end manner.

2) State-of-the-art performance on a large gait dataset with COs.

The proposed method achieves the best age estimation results when compared with other state-of-the-art approaches through experiments on a very large (more than 54,000 subjects) gait dataset with COs with subject ages ranging from 2 to 88 years old.

2. Related work

2.1. Traditional gait-based human age estimation approaches

Early studies on gait-based human age analysis began with gait-based age group classification [1, 4] on relatively small datasets with limited age ranges and coarse age groups. With the recent release of large datasets with wide age ranges, more efforts on gait-based human age estimation have been made. Approaches to traditional gait-based human age estimation are divided into three main categories: classification-based approaches [15], regression-based approaches [17, 19], and hybrids of classification- and regression-based approaches [13].

The classification-based approaches consider the age estimation problem as a classification problem in which age labels are viewed as individual classes. For example, Lu and Tan [15] first encoded each age label into a binary sequence with multiple labels, then applied a multilabel k-nearest neighbor classification algorithm on gait features (*i.e*., GEIs and their Gabor representations) in conjunction with a multilabel-guided subspace learning scheme. Lastly, they decoded the label vector into the final age information.

The regression-based approaches regard the age estimation problem as a regression problem from gait features to continuous age values. For example, Makihara *et al*. [19] directly applied Gaussian process regression (GPR) [2] on GEIs for gait-based human age estimation. Lu and Tan [17] extracted a low-dimensional discriminative representation from original GEI feature space using ordinary preserving manifold analysis methods (*i.e*., ordinary preserving linear discriminant analysis (OPLDA) and ordinary preserving margin Fisher analysis (OPMFA)), and then used a quadratic regression model to estimation the final age labels from these low-dimensional features.

Unlike the first two categories, which employ a common age group-independent estimation model for all age ranges, Li *et al*. [13] proposed an age group-dependent framework for gait-based human age estimation, which is a hybrid of classification and regression-based approaches considering the difference of gait aging process among different age groups (*e.g*., children, adults, and the elderly). Specifically, they first separated the whole age range into five age groups for age group classification, and then regressed the classified samples to their estimated ages by age group-dependent regression models for different age groups.

All these traditional approaches do not, however, consider covariates for gait-based human age estimation, despite the fact that the covariates such as COs may easily affect gait features and degrade age estimation accuracy.

2.2. Deep learning-based human age estimation approaches

With the great success of deep learning-based frameworks compared with traditional ones in many research areas, deep learning-based frameworks have also been introduced into human age estimation tasks, especially for face-based human age estimation [12, 22, 24]. However, there are only a few studies focus on gait-based age estimation [20]. In [20], a deep multi-task learning was proposed for multiple gait attributes (*i.e.*, identity, gender, and age), which included shared feature learning by convolutional neural networks (CNNs) and attribute-specific feature learning for each attribute. As for the estimation of the age attribute, a single unit of a fully connected layer was regressed by the CNN as the final estimated age label. This was then followed by a Tukey's bi-weight loss as the regression loss function.

However, similar to the traditional approaches, current deep learning-based approaches cannot handle CO variations well. They simply learn discriminative features for the COs on the training data, and thus may face large generalization errors on the test data for unseen COs.

2.3. Deep learning-based approaches for gait recognition

Deep learning-based gait recognition approaches mainly include CNN-based approaches [27, 30] and GAN-based approaches [9, 33]. The CNN-based approaches usually take one input to learn invariant gait features in a latent space [27] or take a pair input to perform similarity learning through a Siamese network [30]. In contrast, the GAN-based approaches aim to generate new gait images. GaitGAN [33] was the first work introducing GAN to the gait recognition, and it obtained encouraging results. The GAN model was used as a regressor to generate invariant gait images that are in a canonical view (*i.e.*, side view) from input gait images at arbitrary view angels. Another approach, MGANs [9], was proposed to learn view-specific features by using the assumption of a view angle manifold and transform gait images from the probe view to the gallery view.

Because carrying statuses contain various appearances that may not have an intrinsic manifold, we propose a method that learns carrying status-invariant features in a similar way as GaitGAN. However, there are two main differences between GaitGAN and ours. First, we use a parallel GAN framework as well as paired input GEIs without and with COs of the same subject, while GaitGAN uses one single input. Second, we incorporate an age regression network into the generative network for our final purpose of age estimation and train all the networks in an end-to-end manner, whereas GaitGAN only generates invariant gait features and ignores further process of the invariant features for the recognition task.

3. Proposed method

3.1. Overview

In this subsection, the test and training processes of the proposed method are presented. In the test process, given an input GEI, regardless of whether it has a CO or not, a generative network outputs a GEI without COs while keeping the same identity as the input. Then, the GEI is fed into an age regression network to output an estimated age.

To train the networks, we consider a parallel GAN framework as well as paired input GEIs without and with COs of the same subject. The structure is shown in Fig. 2 (a). Given a pair of GEIs with and without COs of the same subject, we first generate features that have no COs to eliminate the effect of carrying status variations using a parallel GAN-based generative network with shared weights. Then, the generated features are separately fed into the same regression network for age estimation. Finally, the whole network is trained in an end-to-end manner by combining three losses: adversarial loss for GAN-based generative network, pixel-wise loss for preserving the identity information, and age loss to measure a difference between estimated and ground truth ages. Note that the two branches of the parallel GAN share weights with each other so that we can freely handle with any single input GEI regardless of its carrying status in the test process.

In the method proposed in this paper, we choose the most widely used GEI as our gait representation because its simple yet effective properties contain both static body shape features and dynamic leg and arm movement features.

3.2. GAN

The original GAN work [6] introduced a framework for generative models via an adversarial process. There are two models: a generative model G that learns to generate new samples from the data distribution $p_{\boldsymbol{z}}(\boldsymbol{z})$ of a random vector $\boldsymbol{z}$, and a discriminative model D that judges whether its input samples come from the real training data $\boldsymbol{x}$ with distribution $p_{\mathrm{data}}(\boldsymbol{x})$ or the fake generated data from generative model. Models G and D compete with each other through a two-player minmax game with the following adversarial loss function:

$$\begin{aligned}\min_G \max_D \quad & \mathbb{E}_{\boldsymbol{x}\sim p_{\mathrm{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})]+ \\ & \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1-D(G(\boldsymbol{z})))].\end{aligned} \tag{1}$$

Moreover, many variants on GANs (*e.g.*, DCGAN [25], WGAN [7], CGAN [21]) have been proposed to improve the original GAN from many aspects, *e.g.*, stabilizing the training process or making the generated images more realistic. However, the generated images usually cannot be controlled if GANs generate images from random noise. Therefore, other GAN approaches [32, 36] use images as

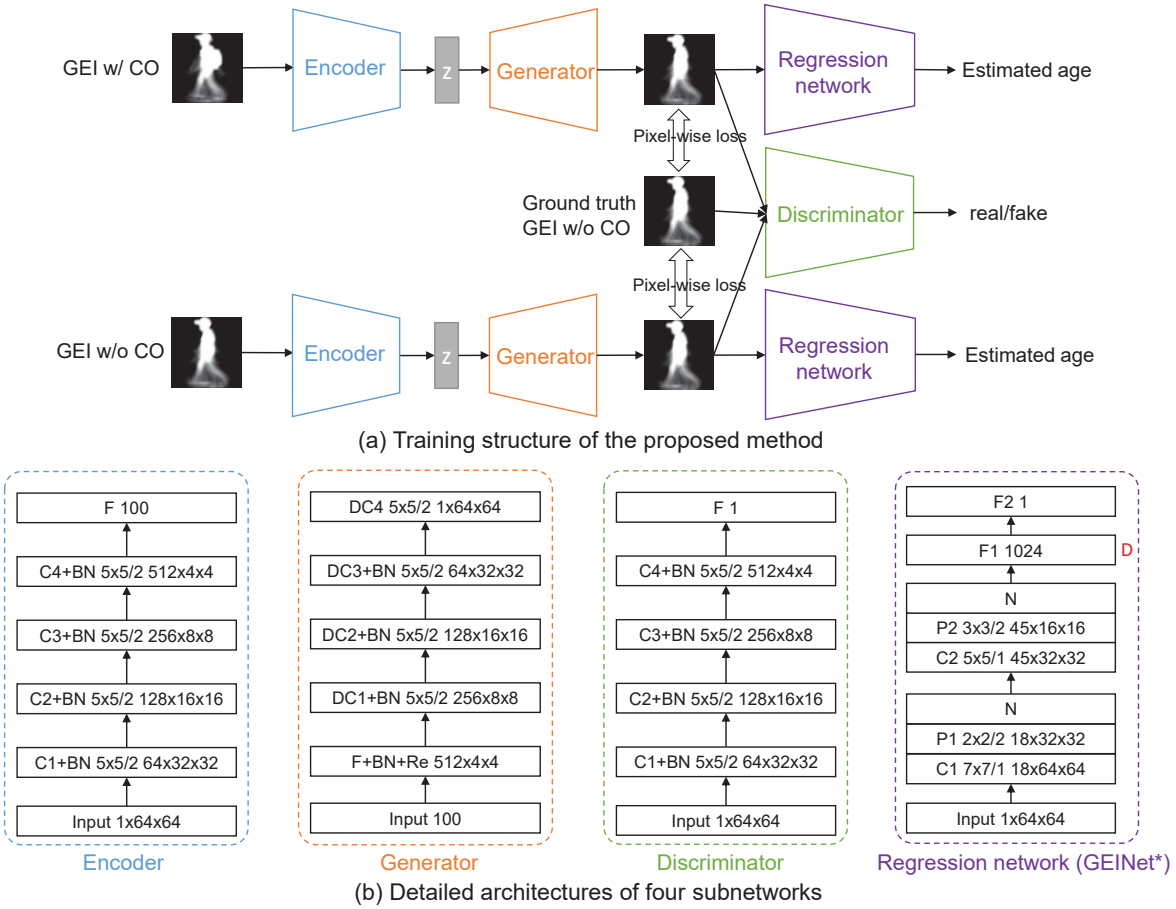

Figure 2. Training structure of the proposed method and detailed network architectures. (a) Training structure of the proposed method; (b) detailed architectures of the four subnetworks (*i.e.*, encoder, generator, discriminator, and regression network). Here, GEINet* is chosen as the regression network for age estimation. Rectangles with letters and numbers represent different kinds of layers and their parameters. Input: input data layer; C: convolutional layer; BN: batch-normalization layer; F: fully connected layer; Re: data reshape layer; DC: deconvolutional layer; P: max-pooling layer; N: local response normalization layer; and D: dropout layer. The numbers after the input layer indicate the size (channels×image height×image width) of the input image or the dimension of a single input vector; those after the (de)convolutional and pooling layers denote (kernel height×kernel width/stride) and (channels×feature height×feature width), separately; those after fully connected layers denote the number of neurons.

their inputs and incorporate an encoder to avoid random sampling of $\boldsymbol{z}$ so as to keep the identity of the input images. Inspired by this family of GANs, we also incorporate an encoder with a generator to generate GEIs without COs that have the same individual-preserving characteristics as the input GEIs.

3.3. Proposed network

The proposed network consists of four subnetworks: an encoder, a generator, a discriminator, and a regression network. Figure 2 (b) shows their respective configurations. We introduce each subnetwork in detail as follows.

Encoder. The encoder is a CNN that takes a GEI as an input and outputs a low-dimensional vector $\boldsymbol{z}$ that preserves the individual gait feature of the input GEI while removing the feature of the COs. The input GEI is of size $1 \times 64 \times 64$. There are four convolutional layers with a stride of two. Like [25, 36], we do not use a pooling strategy (*e.g.*, max-pooling) because stride convolution can also learn its differentiable spatial downsampling. Each convolutional layer is followed by a batch-normalization layer and a rectified linear unit (ReLU) activation function. At last, a fully connected layer is employed to embed the low-dimensional vector.

Generator. Low-dimensional vector $\boldsymbol{z}$, which is an output of the encoder, is fed into the generator so as to output a GEI without COs. Therefore, we achieve the purpose of eliminating the effect of CO variations. The architecture of the generator is just the reverse of that of the encoder. The first fully connected layer incorporates a reshape oper-

ation to transfer the low-dimensional input vector into the image domain with multiple channels. Then, four deconvolutional layers with a stride of two are used for spatial upsampling. Each deconvolutional layer is followed by a batch-normalization layer and a ReLU activation function, except for the last layer, which uses a Tanh activation function to output the generated GEI.

Discriminator. Similar to the traditional GAN, the purpose of the discriminator is to determine whether the input images are real or fake in order to help the whole network to generate realistic images. The inputs are from the fake GEIs generated by the encoder and generator, the real GEIs without COs from the training data, and their corresponding fake and real labels. After going through a CNN similar to the encoder, the discriminator outputs a scalar value to represent the probability that the input is fake or real. Following the DCGAN [25], we use a LeakyReLU activation function for all convolutional layers in the discriminator.

Regression network. The purpose of the regression network is to predict the age of the input GEI generated from the encoder and generator. In this paper, to show the effectiveness of the proposed GAN-based generative network, we simply use the GEINet [27], which is a standard CNN for cross-view gait recognition with a single input, as the regression network. Note that it could be easily replaced by other CNNs for better performance in future works. To meet our requirements for age estimation, we make a slight modification in that we set the dimension of the last fully connected layer to one so as to regress the estimated age, while the number of the last layer's nodes for the original GEINet is the number of subjects to be classified. We refer to the modified network as GEINet* in this paper.

3.4. Objective function

To train the proposed network, we designed an objective function by fusing three kinds of losses: the adversarial loss for generating GEIs without COs, the pixel-wise loss for preserving the identity of the generated GEIs, and the age loss to measure a difference between estimated and ground truth ages.

Adversarial loss. Suppose that x_{CO}, x is the input pair of GEIs with and without COs from the same subject, and that x is also regarded as the real ground truth GEI without COs. The encoder E, generator G, and discriminator D can be trained using the following adversarial loss:

$$\begin{aligned} L_{\mathrm{Adv}}(E,G,D) =& \mathbb{E}[\log D(x)] + \mathbb{E}[\log(1 - D(G(E(x))))] \\ &+ \mathbb{E}[\log(1 - D(G(E(x_{\mathrm{CO}}))))]. \end{aligned} \tag{2}$$

Pixel-wise loss. To preserve the identity of generated GEIs, the pixel-wise loss is introduced to minimize the pixel-wise difference between the ground truth GEI x (real) and the generated GEIs (fake). This loss is written as

$$L_{\mathrm{Pix}}(E,G) = \|G(E(x_{\mathrm{CO}})) - x\| + \|G(E(x)) - x\|, \tag{3}$$

where $\|\cdot\|$ is the L1 norm.

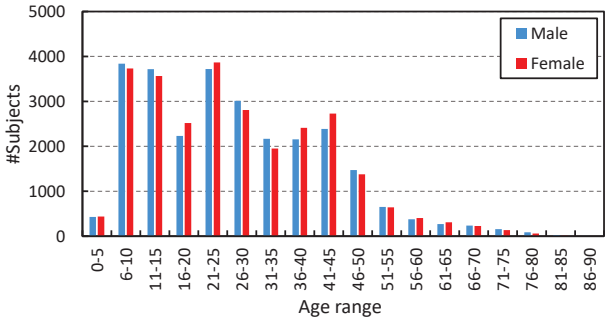

Figure 3. Statistics of subjects' age and gender in the dataset.

Age loss. The age loss is an essential part of the final objective function that not only includes the mean absolute difference of the estimated age and ground truth age, but also includes the difference of the ages estimated from the input pair of GEIs. We define it as follows:

$$\begin{aligned} L_{\mathrm{Age}}(E,G,R) =& \|R(G(E(x_{\mathrm{CO}}))) - a\| + \\ & \|R(G(E(x))) - a\| + \\ & \|R(G(E(x_{\mathrm{CO}}))) - R(G(E(x)))\|, \end{aligned} \tag{4}$$

where a is the ground truth age. R denotes the regression network for age estimation, and the pair $R(G(E(x_{\mathrm{CO}}))), R(G(E(x)))$ consists of the estimated ages of the input pair x_{CO}, x. The last term is introduced to reduce the gap between the estimated ages of input pairs with and without COs.

Then, we define the objective function as a weighted sum of the above three losses as

$$\begin{aligned} L(E,G,D,R) =& \lambda_{\mathrm{Adv}} L_{\mathrm{Adv}}(E,G,D) + \lambda_{\mathrm{Pix}} L_{\mathrm{Pix}}(E,G) + \\ & \lambda_{\mathrm{Age}} L_{\mathrm{Age}}(E,G,R), \end{aligned} \tag{5}$$

where λ_{Adv}, λ_{Pix}, and λ_{Age} are three hyper-parameters that balance the effect of different losses. Finally, we aim to optimize

$$\min_{E,G,R} \max_{D} \quad L(E,G,D,R). \tag{6}$$

4. Experiments

4.1. Dataset

We conducted the experiments on a large gait dataset with COs. The dataset is an intersection of the currently largest publicly available datasets with age information and CO variations, *i.e.*, OULP-Age [31] and OULP-Bag [28]. Specifically, we keep the subjects in OULP-Bag that have

age labels in OULP-Age because these two datasets were both collected through a demonstration of video-based gait analysis in a science museum [18]. In the final dataset, there remain 54,117 subjects (26,936 males and 27,181 females) with ages ranging from 2 to 88 years old. The statistics distribution of the dataset is shown in Fig. 3. Each subject has two sequences: one is without any COs and another one is with their own COs. These COs are carried in various locations and include almost all common COs in real life. For example, some subjects have their backpacks carried on their back or in front, some subjects have their handbags carried at the body side region, and some subjects even have multiple COs carried at multiple regions. Some typical GEI samples can be found in Fig. 1. In the experiment, we randomly shuffled the whole dataset into two subsets (one has 27,059 subjects, another has 27,058 subjects) and used two-fold cross-validation for the evaluation (the protocal is available at `http://www.am.sanken.osaka-u.ac.jp/BiometricDB/index.html`.)

4.2. Implementation details

We initialized the weights of each layer using a Gaussian distribution with a mean of zero and a standard deviation of 0.02. All the bias terms were initialized with a constant of zero. For the LeakyReLU, the slope of the leak was set to 0.2. In the training stage, we first selected 20% of the training set as a validation set while the remaining 80% was used for training, and then chose the hyper-parameters of the objection function in Eq. (5) based on the best performance on the validation set. As a result, set $\lambda_{\mathrm{Adv}} = 0.0001$, $\lambda_{\mathrm{Pix}} = 1$, and $\lambda_{\mathrm{Age}} = 10$. The dimension of the low-dimensional embedded feature z produced by the encoder was set to 100. The whole network parameters were updated by the Adam optimizer [11] with a mini-batch size of 100. Similar to [25, 36], the learning rate was set to 0.0002 and the momentum term β_1 was set to 0.5. We stopped the training process after 600 epochs.

4.3. Evaluation metrics

We employed two widely used measures, *i.e.*, the mean absolute error (MAE) and cumulative score (CS), to evaluate the accuracy of age estimation. The MAE is the average of the absolute errors between the estimated ages and ground truth ages, which is defined as $\mathrm{MAE} = \frac{1}{M}\sum_{i=1}^{M}|\hat{a}_i - a_i|$, where a_i is the ground truth age for the i-th test sample, $\hat{a}_i$ is its estimated age, and M is the total number of test samples. The CS for the j-years absolute error tolerance $\mathrm{CS}(j)$ is defined as $\mathrm{CS}(j) = \frac{M_{e\leq j}}{M}$, where $M_{e\leq j}$ is the number of test samples whose absolute errors are less than j years.

Table 1. Total MAEs (years) (*i.e.*, mean±standard deviations) for all methods. Bold and italic bold fonts indicate the best and second-best results, respectively. This convention is consistent throughout this paper.

Method	Total MAE
MLG	11.45±0.02
GPR (K=1,000)	7.96±0.04
SVR (linear)	9.63±0.27
SVR (Gaussian)	8.15±0.04
OPLDA	9.11±0.06
OPMFA	9.60±0.03
Age group-dependent	7.57±0.04
GEINet*	***7.15±0.04***
Proposed	**6.61 ±0.03**

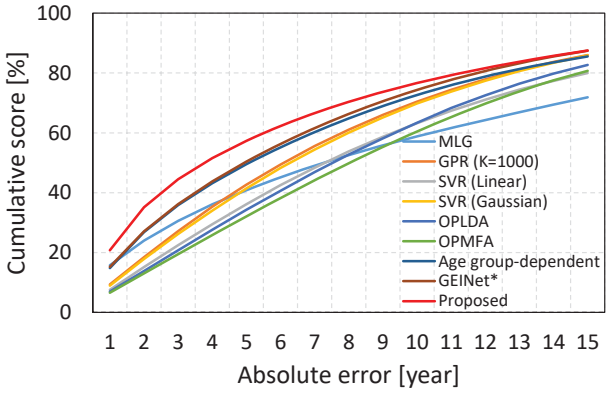

Figure 4. Mean cumulative scores for all methods with an absolute error tolerance from 1 to 15 years.

4.4. Comparison with the state-of-the-art methods

The state-of-the-art methods for comparison consist of the traditional approaches, *i.e.*, MLG [15], GPR [19], SVR [29], OPLDA [17], OPMFA [17] and an age group-dependent approach [13] (called it Age group-dependent later), as well as a CNN-based approach, *i.e.*, GEINet* [27]. Specifically, the GPR was implemented with a large K ($\mathrm{K} = 1,000$), which is a parameter that determines the number of neighboring training samples to a test sample, and according to [31], larger K results in better performance. The SVR was implemented with both linear and Gaussian kernels. The GEINet* was trained using the same settings as the proposed method.

We first predicted the ages of all subjects regardless of their carrying statuses and show the total MAEs for all methods in Table 1. The results show that the proposed method outperforms the traditional approaches by a large margin. Compared with the CNN-based approach GEINet*, it also achieves a lower MAE. Moreover, we show the mean cumulative scores for all methods with absolute

Table 2. MAEs (years) (*i.e.*, mean±standard deviations) of the subjects w/o and w/ COs for all methods.

Method	MAE w/o COs	MAE w/ COs
MLG	10.90±0.00	12.00±0.04
GPR (K=1,000)	7.59±0.02	8.33±0.05
SVR (linear)	9.29±0.31	9.98±0.22
SVR (Gaussian)	7.69±0.03	8.61±0.05
OPLDA	8.87±0.08	9.35±0.04
OPMFA	9.52±0.05	9.69±0.00
Age group-dependent	7.22±0.06	7.92±0.03
GEINet*	***6.81±0.03***	***7.49±0.05***
Proposed	**6.15±0.05**	**7.07±0.02**

error tolerance values from 1 to 15 years in Fig. 4. We can clearly see that the proposed method gets much better results than other benchmarks. These results demonstrate that the proposed method achieves the state-of-the-art performance in terms of both MAE and CS.

4.5. Effect on solving CO variations

In this subsection, we analyze the effect of carrying status on age estimation accuracy by comparing the MAEs of the subjects without and with COs. The results are shown in Table 2. When comparing the results of the two cases, we can clearly observe that the CO variations substantially degrade the accuracies of age estimation for all methods. Although this degradation of the proposed method is relatively large, this is beacause the proposed method first generates GEIs without COs from the input GEIs with and without COs and then feeds them into the regression network for age estimation which may make the age estimator overly fit to GEI without COs. However, when comparing the results of all methods, the proposed method achieves the best MAE in both cases, *i.e.*, with respect to the second-best method GEINet*, it achieves a 0.66 lower MAE for the subjects without COs and a 0.42 lower MAE for the subjects with COs.

Additionally, because we use the GEINet* as the regression network of the proposed method for predicting the age information, we present further comparisons between the proposed method and GEINet*, which can be regarded as an individual component analysis of the proposed GAN-based generative model. First, we show the mean MAEs with respect to the ground truth age at each interval of 5 years in Fig. 5. In general, for both methods, the MAEs of elderly age ranges are relatively larger than those of younger age ranges because the lack of older subjects in the dataset. When comparing with each other, it turns out that the MAEs of the proposed method are lower than those of GEINet* in almost all the ground truth age ranges, especially in the very young and elderly ranges (*e.g.*, 0–5, 6–10, 81–85, and 86–90 years). Second, we show the scatter plots for the ground truth ages versus the corresponding estimated ages of all the subjects in Fig. 6. The estimated ages of the proposed method are more compactly distributed around the diagonal line (*i.e.*, ground truth) than those of GEINet*, particularly for the younger and elderly age ranges.

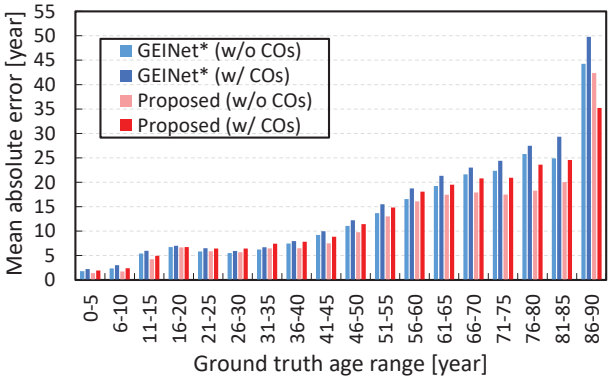

Figure 5. Mean MAEs of the subjects w/o and w/ COs for the proposed method and GEINet* in 5-year intervals.

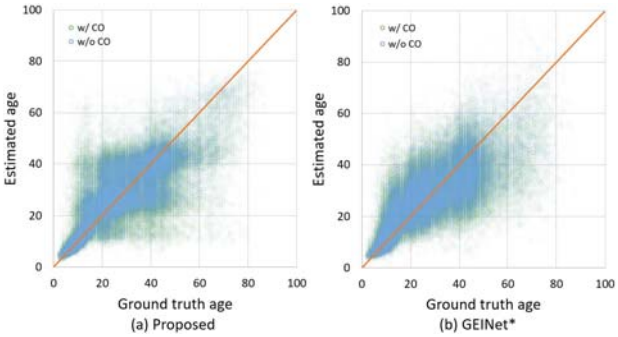

Figure 6. Scatter plots for the ground truth ages versus the corresponding estimated ages of all the subjects. The orange diagonal line represents that the estimated ages are equal to the ground truth ages.

In conclusion, all the aforementioned results show a large improvement obtained by the proposed method for the subjects both with and without COs. This indicates that the proposed method not only eliminates the effect of COs variations for the subjects with COs, but also refines the GEIs themselves even for the subjects without COs (*e.g.*, mitigating silhouette extraction errors in GEIs) through the proposed GAN-based generative model so as to make them suitable for age estimation.

4.6. Feature visualizations

To analyze the qualitative properties of the proposed method, we visualize some GEIs generated by our GAN-based generative model in Fig. 7. First, by checking the generated GEIs in the fourth column, we confirm that the

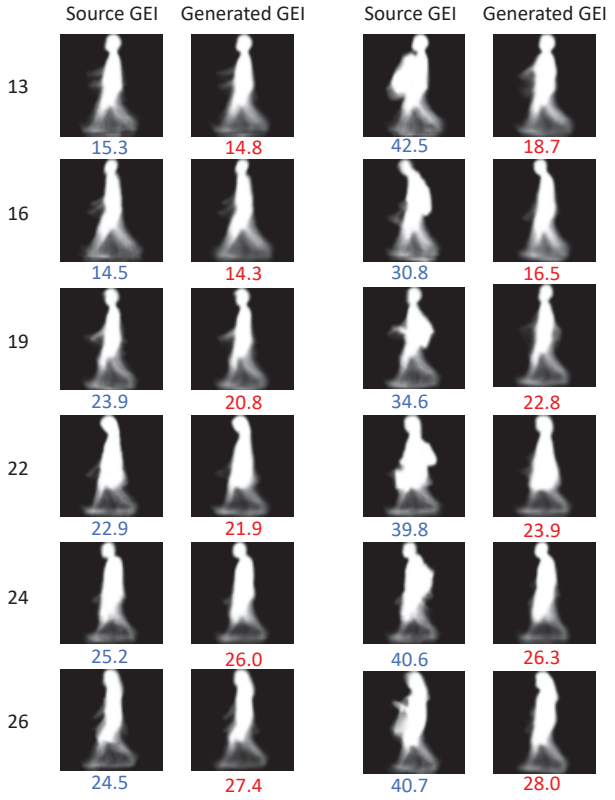

Figure 7. Six subjects' GEIs w/o and w/ COs and their corresponding generated GEIs. Each row indicates the same subject; the first and second columns show the source GEIs w/o COs and their generated GEIs, respectively; the third and fourth columns show the source GEIs w/ COs and their generated GEIs, respectively. Black digits are the ground truth ages of each subject; blue digits under the source GEIs are the estimated ages of the GEINet*; and red digits under the generated GEIs are the estimated ages of the proposed method.

proposed method can successfully eliminate the CO variations and generate well identity-preserving GEIs (c.f. the original GEIs without COs in the first column). In addition to removing the COs, the proposed method can even fix the stooping body shape to some extent caused by some heavy COs (see the second row). Second, by checking the generated GEIs in the second column, we find that the proposed method can also ensure its good performance and generate almost the same GEIs for the subjects without COs.

Moreover, Fig. 7 shows the age estimation results of the GEINet* and the proposed method for the six subjects, as well as their ground truth ages. We can see that the GEINet* can predict accurate age information for the subjects without COs, but fails for the subjects with COs. In contrast, the proposed method can achieve much better results than the GEINet* for the samples with COs and results comparable with those of the GEINet* for the samples without COs.

5. Conclusion

This paper described a parallel GAN-based method for gait-based human age estimation robust against carrying status variations. Through the parallel GAN-based model training, the proposed method can generate gait features without COs that are suitable for gait-based age estimation regardless of the carrying status of input gait features. Experimental results on a large gait dataset show the state-of-the-art performance of the proposed method when dealing with CO variations compared with the performance of benchmark methods. Because we use a relatively simple regression network (*i.e.*, GEINet*) for age estimation and current age estimation results are still not enough for real-work applications, we plan to investigate more effective networks for better age estimation results in our future work.

Acknowledgment

This work was supported by JSPS Grants-in-Aid for Scientific Research (A) JP18H04115, the National R&D Program for Major Research Instruments (Grant No. 61727802). We thank Kimberly Moravec, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

References

[1] R. Begg. Support vector machines for automated gait classification. *IEEE Trans. on Biomedical Engineering*, 52(5):828–838, May 2005.

[2] C. K. I. W. Carl Edward Rasmussen. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[3] K. Y. Chang and C. S. Chen. A learning framework for age rank estimation based on face images with scattering transform. *IEEE Transactions on Image Processing*, 24(3):785–798, 2015.

[4] J. W. Davis. *Visual Categorization of Children and Adult Walking Styles*, pages 295–300. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[5] X. Geng, Z. H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 29(12):2234–2240, Dec. 2007.

[6] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, 2014.

[7] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville. Improved training of wasserstein gans. In *NIPS*, 2017.

[8] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316– 322, 2006.

[9] Y. He, J. Zhang, H. Shan, and L. Wang. Multi-task gans for view-specific feature learning in gait recognition. *IEEE Transactions on Information Forensics and Security*, 14(1):102–113, Jan 2019.

[10] O. F. Ince, J. Park, J. Song, and B. Yoon. Child and adult classification using ratio of head and body heights in images. *International Journal of Computer and Communication Engineering*, 3(2):120, 2014.

[11] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[12] G. Levi and T. Hassner. Age and gender classification using convolutional neural networks. *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 34–42, 2015.

[13] 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 and Applications*, 77:28333–28354, 2018.

[14] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *Proc. of the 17th International Conference on Pattern Recognition*, volume 1, pages 211–214, Aug. 2004.

[15] J. Lu and Y.-P. Tan. Gait-based human age estimation. *IEEE Trans. on Information Forensics and Security*, 5(4):761–770, Dec. 2010.

[16] J. Lu and Y.-P. Tan. Ordinary preserving manifold analysis for human age estimation. In *IEEE Computer Society and IEEE Biometrics Council Workshop on Biometrics 2010*, pages 1–6, San Francisco, CA, USA, Jun. 2010.

[17] J. Lu and Y. P. Tan. Ordinary preserving manifold analysis for human age and head pose estimation. *IEEE Transactions on Human-Machine Systems*, 43(2):249–258, March 2013.

[18] Y. Makihara, T. Kimura, F. Okura, I. Mitsugami, M. Niwa, C. Aoki, A. Suzuki, D. Muramatsu, and Y. Yagi. Gait collector: An automatic gait data collection system in conjunction with an experience-based long-run exhibition. In *Proc. of the 8th IAPR Int. Conf. on Biometrics (ICB 2016)*, number O17, pages 1–8, Halmstad, Sweden, Jun. 2016.

[19] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi. Gait-based age estimation using a whole-generation gait database. In *Proc. of the Int. Joint Conf. on Biometrics (IJCB2011)*, pages 1–6, Washington D.C., USA, Oct. 2011.

[20] M. J. Marín-Jiménez, F. M. Castro, N. G. Mata, F. de la Torre, and R. Muñoz-Salinas. Deep multi-task learning for gait-based biometrics. In *IEEE International Conference on Image Processing (ICIP)*, Sept 2017.

[21] M. Mirza and S. Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014.

[22] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua. Ordinal regression with multiple output cnn for age estimation. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4920–4928, 2016.

[23] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. Int. Series on Biometrics. Springer-Verlag, Dec. 2005.

[24] H. Pan, H. Han, S. Shan, and X. Chen. Mean-variance loss for deep age estimation from a face. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[25] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, abs/1511.06434, 2015.

[26] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. G. ther, and K. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.

[27] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. Geinet: View-invariant gait recognition using a convolutional neural network. In *Proc. of the 8th IAPR Int. Conf. on Biometrics (ICB 2016)*, number O19, pages 1–8, Halmstad, Sweden, Jun. 2016.

[28] M. Z. Uddin, T. T. Ngo, Y. Makihara, N. Takemura, X. Li, D. Muramatsu, and Y. Yagi. The ou-isir large population gait database with real-life carried object and its performance evaluation. *IPSJ Transactions on Computer Vision and Applications*, 10(1):5, May 2018.

[29] V. N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.

[30] 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 Transactions on Pattern Analysis and Machine Intelligence*, 39(2):209–226, Feb 2017.

[31] 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. *IPSJ Transactions on Computer Vision and Applications*, 9(1):24, Dec 2017.

[32] D. Yoo, N. Kim, S. Park, A. S. Paek, and I.-S. Kweon. Pixel-level domain transfer. In *ECCV*, 2016.

[33] S. Yu, H. Chen, E. B. Garcia Reyes, and N. Poh. Gaitgan: Invariant gait feature extraction using generative adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[34] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu. A study on gait-based gender classification. *IEEE Trans. on Image Processing*, 18(8):1905–1910, Aug. 2009.

[35] D. Zhang, Y. Wang, and B. Bhanu. Ethnicity classification based on gait using multi-view fusion. In *IEEE Computer Society and IEEE Biometrics Council Workshop on Biometrics 2010*, pages 1–6, San Francisco, CA, USA, Jun. 2010.

[36] Z. Zhang, Y. Song, and H. Qi. Age progression/regression by conditional adversarial autoencoder. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4352–4360, 2017.