IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 1, 2021, accepted February 21, 2021, date of publication February 24, 2021, date of current version March 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3061684

Learn to Walk Across Ages: Motion Augmented Multi-Age Group Gait Video Translation

YIYI ZHANG[1,2], YASUSHI MAKIHARA[2], DAIGO MURAMATSU[2], (Member, IEEE), JIANFU ZHANG[1], LI NIU[1], LIQING ZHANG[1], (Member, IEEE), AND YASUSHI YAGI[2], (Member, IEEE)

[1]MOE Key Laboratory of Artificial Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Department of Intelligent Media, The Institute of Scientific and Industrial Research, Osaka University, Osaka 565-0871, Japan

Corresponding author: Yiyi Zhang (zhang@am.sanken.osaka-u.ac.jp)

This work was supported by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (JSPS KAKENHI) under Grant JP18H04115, Grant JP19H05692, and Grant JP20H00607.

ABSTRACT We propose a framework for multi-age group gait video translation in which, for the first time, individuality-preserving aging patterns in walking style are learnt. More specifically, we build our framework on an existing multi-domain image translation model. Because the existing multi-domain image translation model was originally designed for a still image, we extend it to gait video by introducing a motion-augmented network architecture with three streams, where gait period, period-normalized phase-synchronized gait video, and its frame difference sequence are each input to one stream. We then train the network to ensure three aspects: aging effect (using an age group classification loss), individuality preservation (using a reconstruction loss), and gait realism (using an adversarial loss). Our framework quantitatively and qualitatively outperforms state-of-the-art age progression/regression methods on the largest gait database, OULP-Age, with respect to both age group classification and identity recognition.

INDEX TERMS Gait aging, gait video generation.

I. INTRODUCTION

Gait refers to a person's walking style and is considered as one of behavioral biometrics, which contains a variety of attributes of the person such as age, gender, health status, and identity. Among the attributes, since age changes in gait is inevitable during a person's life span, it is considered as one of covariates for gait-based person identification a.k.a. gait recognition. Viewed from another perspective, the gait provides a cue to estimate an age or age group and hence gait-based age estimation also enjoy a rich body of literature [1]–[6].

Taking the above-mentioned fact into consideration, if we realize a function of age progression/regression on gait, i.e., translation of the gait from one age to another by preserving the identity (see Fig. 1), we may be able to employ it for many applications. For example, assuming that a criminal investigator tries to find a perpetrator by gait recognition after a long time (e.g., 10 years), he/she can mitigate intra-subject variations between a matching pair of an enrollment and a query by progressing the age of the enrolled gait by 10 years with the function. As another application in the healthcare field, a person may pay more efforts to keep youthfulness in gait by watching his/her age-progressed gait video with the function.

Although age progression/regression has been extensively studied in the face analysis community [7], [8], the study [9] is the only one on gait age progression/regression, to the best of our knowledge. In their work, the authors progress a gait energy image (GEI) [10] (also called an averaged silhouette [11]), which is the most widely used gait template in the gait recognition community, using a subject-independent geometric warping field. Taking into account a variety of potential applications, it is naturally preferable to translate not a static gait template such as a GEI but a gait video (e.g., recent gait recognition no longer relies on a static gait template but employs a gait video as input instead). In addition, since one process of gait age progression/regression may differ from the process in another individual, it is preferable to translate a gait video in a subject-dependent way, unlike [9], which adopts a subject-independent method of translation. Moreover, most gait analysis studies overlook an important component in gait, i.e., the gait period (or gait cycle), or they regard it just as a normalization factor. In fact, static gait

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Sharif.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

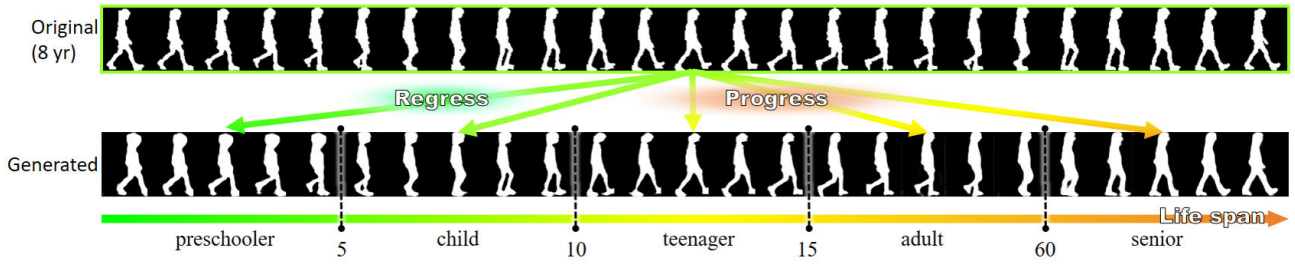


FIGURE 1. Given one's current gait video, we attempt to translate into his/her gait video in different age groups of life.

templates such as GEI [10], frequency-domain feature [12], chrono-gait image [13], are all period-normalized templates (e.g., in case of GEI, gait silhouettes are aggregated over a gait period and then divided by the gait period to obtain an average), and hence the gait period is ''washed out'' in those gait templates. In addition, the gait period has not yet been explicitly exploited, even in recent deep learning-based gait analysis [14]–[16]. This is partly because the gait period is unstable due to intra-subject variations (e.g., changes in walking speed) in the context of gait-based person identification. The gait period is, however, naturally quite important information in the context of age. For example, a child's gait period is usually shorter than that of an adult or elderly person, in other words, children's gaits have higher cadence (or frequency). The gait period is therefore considered to be one of key components for representing the characteristics of each age group.

In this work, we therefore propose a method of gait video translation among multiple age groups. For better representation of the age groups' characteristics, we convert a gait video into the gait period and a period-normalized (or rate-normalized) gait video with a fixed number of frames, and then translate both of them in an original age group to those in a target age group. As such, we can reflect aging effect not only in gait sequence of one gait period but also in the gait period (see the supplementary videos to recognize the difference in gait periods among age groups). In addition, we make the gait video translation subject-dependent by introducing StarGAN, a kind of conditional GAN framework, which was originally designed for generic static image translation. The main contributions of this work include the followings.

A. GAIT VIDEO TRANSLATION AMONG MULTIPLE AGE GROUPS

We propose a method of gait video translation from one age group to another for the first time, unlike the existing work [9] translates just a static gait template. This is beneficial for potential applications such as cross-age gait recognition using not a static gait template but a gait video as an input.

B. GAIT-ORIENTED MOTION-AUGMENTED VIDEO TRANSLATION

In order to better translate all contents in a cyclic gait video, we design a video translation framework with three streams: a period-normalized gait silhouette sequence, a period-normalized frame difference sequence, and a gait period. While the first one mainly encodes body shape aspects in a gait video, the latter two encode motion aspects in it. This technically differentiates the proposed method from other work on still image-based face age regression/progression and general image/video translation.

C. STATE-OF-THE-ART PERFORMANCE

Our method outperforms modified versions of state-of-the-art of facial age progression/regression qualitatively and quantitatively in both person identification and age group classification tasks with the OULP-Age, the largest publicly available gait database with age information.

II. RELATED WORK

A. FACE AGE PROGRESSION/REGRESSION

Extensive studies on face aging can be mainly divided into three groups: physical model-based approaches, prototype-based approaches, and deep learning-based approaches [17].

The physical model-based approaches correlate biological and physical mechanism (i.e., craniofacial growth, skin, and winkles) with human age using models such as an and-or graph, a concatenational graph evolution aging, a craniofacial growth. Although those models are carefully designed, they rely heavily on the imperfect human knowledge [8] and require sufficient training samples with aging sequence over a long age span for each individual, which are almost impossible to be collected for gait.

In the prototype-based approaches, averaged faces are created as the prototype for each group, the difference between each prototype is regarded as the transition pattern. To offset the missing personal characteristic in the transition pattern due to the averaged faces, Shu *et al.* proposed novel aging dictionary learning methods to better preserve personality [18], [19]. However, it still relies on dataset (i.e. CACD [20]) with short period paired data, which is unlikely to be acquired for gait. Meanwhile, to obtain progressed and regressed facial image, the prototype-based approaches need to be trained twice [21].

Recently, the deep learning-based approaches to facial age progression/regression have achieved the state-of-the-art performance in age group classification and identity preservation with no paired data. Particularly, Zhang *et al.* [21] proposed a

conditional adversarial autoencoder where each facial image corresponds to a point on the manifold. Translated aging facial images are obtained through stepping along the aging axis on the manifold. Other GAN-based approaches [7], [8], [22] adopt a pretrained classifier and a conditional GAN architecture to generate faces conditioned on age during progression and regression to preserve the identity. Although deep learning-based approaches to face aging can generate good simulation results, a gait is not a static image but a video (image sequence), and hence we consider to directly handle the gait video as an input/output and to better handle motion and appearance aspect.

B. GAIT-BASED AGE ANALYSIS

Prior studies in gait-based age group classification have demonstrated the fact that gait contains discriminative aging patterns in a long-elapsed time period. For example, Mannami have adopted a frequency domain feature [23] to classify four age groups: children (under 15 years old), adult males, adult females, and the elderly (over 65 years old). Chuen *et al.* have investigated correlation of image-based gait features (i.e. stride length, body length, head-to-body ratio) to distinguish between children and adults [24]. These internally contained discriminative aging pattern in gait video make the task of multi-age group gait translation practical. Actually, Xu *et al.* have conducted the first gait age translation among multi-age groups [9]. However, similar to most of the current analysis in gait that utilize image-based features such as frequency domain feature [12], chrono-gait image [13], gait flow image [25], Gabor GEIs [26], their work rely on the GEI, which also falls into the category of image-based gait features.

Although the static image-based gait feature including the GEI had been considered simple yet effective representations for gait analysis, they have several problems such as highly compressed motion information and entanglement of appearance (body shape) and motion information. They have been therefore gradually replaced with spatio-temporal or more disentangled representations. For example, Chao *et al.* have introduced GaitSet [14], a framework that make use of gait sequences to achieve the state-of-the-art performance in gait recognition, outperforming methods that utilize image-based features by a large margin. In this paper, we also leverage gait videos to obtain gait aging pattern from both appearance and motion.

C. VIDEO-TO-VIDEO TRANSLATION

Video-to-video translation addressed in [27]–[30] has aroused attention from researchers recently. Given a video to drive a motion, these methods transfer the driving motion to an input static image with a different content, and generate a video with the different content and the same driving motion. Unlike the previous work tries transferring the same motion, our task aim to learn the aging pattern automatically across ages without the driving motion.

As an extension of image-to-image translation, prior works in video-to-video translation [27], [28] require paired data. Bansal *et al.* proposed RecycleGAN [29] to first facilitate unpaired video-to-video translation. Meanwhile, these approaches have limitations: they rely on the variants of CycleGAN [31], and hence they are only capable of learning the relations between two different domains at a time, which lack scalability in handling multiple domains.

III. THE PROPOSED METHOD

A. OVERVIEW

Our goal is to translate a gait video among multiple age groups. In order to efficiently handle multi-age group translation, we build our method upon StarGAN [32] because it can translate images among multiple domains just by a single unified model (i.e., with less network parameters compared with pairwise translation models). The StarGAN is a kind of a conditional GAN, and hence it has a generator and a discriminator as shown in the overview in Fig. 2. In addition, for better treatment of all the components appeared in the gait video, we take a triplet of period-normalized gait silhouette sequence, frame difference sequence, and a gait period as an input for the generator/discriminator. We will describe preprocessing, the generator/discriminator, and loss functions, in the following subsections.

B. PREPROCESSING

We briefly describe preprocessing to prepare input data, i.e., a triplet of period-normalized gait silhouette sequence, frame difference sequence, and a gait period, for our generator/discriminator.

Given a gait video, gait silhouettes were first extracted by graph-cut segmentation supported by background subtraction [33], since color and texture information are relevant with neither gait nor body shape. We then obtained size-normalized and registered silhouette sequences in size 88 by 128 pixels. One gait period (or cycle) was detected by maximizing auto-correlation along the temporal axis [12]. We morphed the sequence to produce a period-normalized phase-synchronized gait silhouette sequence per subject with N_{img} frames [9], where N_{img} is experimentally to 25. Examples of the resultant sequence is shown in Fig. 3. We also extract frame difference sequence, which is also used in a motion-oriented gait recognition method [34].

The image size of each frame in the gait silhouette sequence and the frame difference sequence was finally converted to 128 by 128 pixels by adding zero-padded columns at left and right sides, and then frames in each sequence are stacked over channel dimensions, for the convenience of treatment in a network architecture.

As for the gait period [frame], we represent it an indicator vector rather than an integer scalar for more flexible non-linear translation. We therefore first experimentally set minimum and maximum gait periods as $P_{min} = 20$ and $P_{max} = 41$, respectively, and then set $N_{period}(= P_{max} - P_{min}+1)$-dimensional one-hot vector $\vec{x}_{period} \in \mathbb{R}^{N_{period}}$ whose i-th component represent the gait period $(P_{min} + i - 1)$.

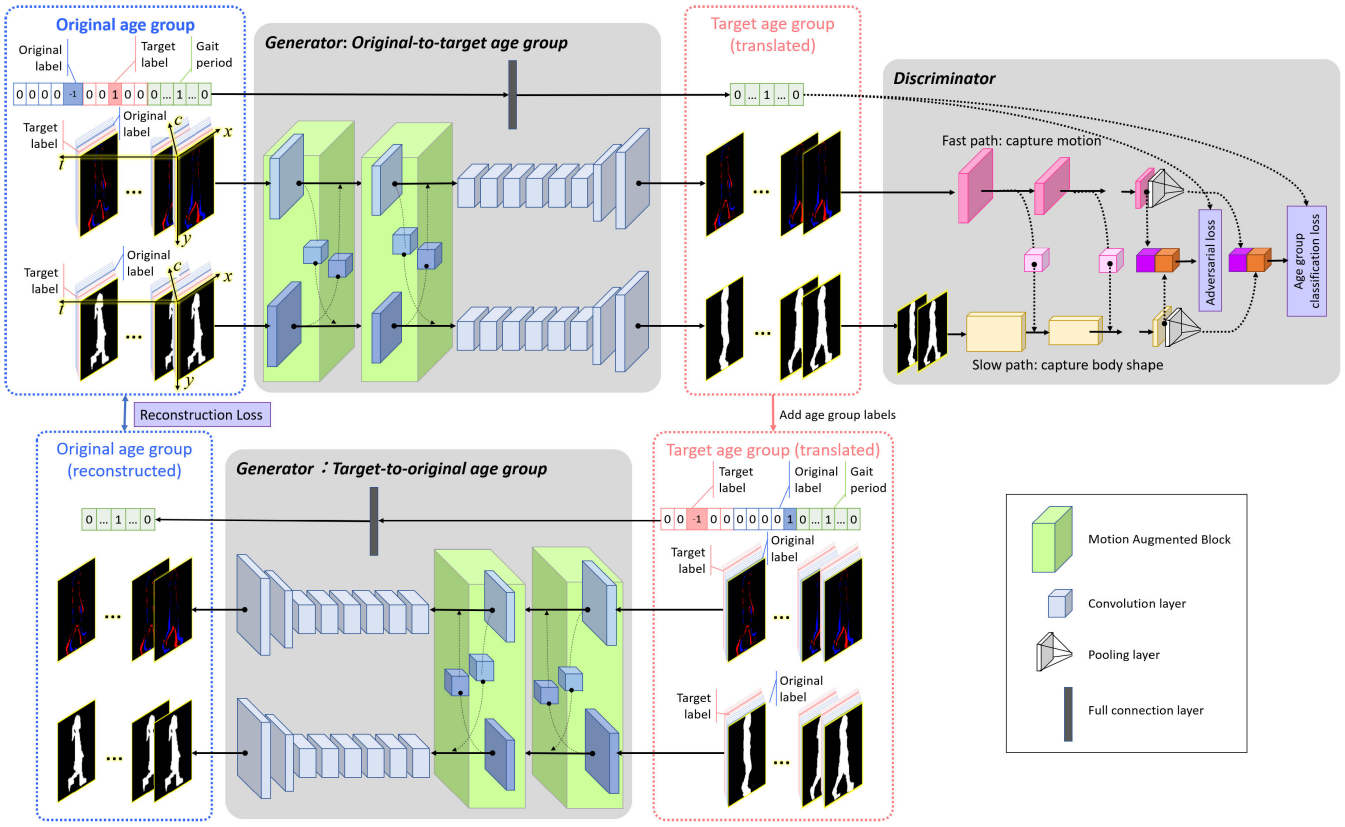


FIGURE 2. Overview of the proposed framework, consisting of a three-stream generator/discriminator. We assign positive and negative values in frame difference to red and blue just for visualization purpose. Given an original triplet of period-normalized gait silhouette sequence, frame difference sequence, and a gait period, as well as an original and target age groups' labels, the generator translates the triplet to the target age group. The translated triplet of the target age group is further translated back to that of the original age group to preserve individuality by a reconstruction loss, while the triplet is also fed into the discriminator to ensure reality and characteristics of the target age group by adversarial and age group classification losses, respectively.



FIGURE 3. Period-normalized phase-synchronized silhouette sequences (top: an adult, bottom: a child). Thirteen frames from a half gait period are shown due to space limitation.

Specifically, given an input gait period P, the $(P-P_{min}+1)$-th component of the vector $\vec{x}_{period}$ is 1 and the others are zero-padded.

C. GENERATOR WITH MOTION AUGMENTED BLOCK

We feed the generator the triplet described in Section III-A to augment the generator with more motion information.

At the beginning, since we build our framework upon StarGAN [32], i.e., image translation framework among multiple domains, we also prepare indicators for multiple domains, (i.e., age groups).

As for the gait period, we prepare N_{age}-dimensional one-hot vector, where N_{age} is the number of age groups. In this vector, the element for the target age group is set to one and the others are set to zero. In addition, we prepare an indicator vector for an original (or source) age group with the same dimension. We then concatenate the two indicator vectors to the gait period indicator vector $\vec{x}_{period}$ in channel dimension, which results in $(N_{period}+2N_{age})$ channels in total (see Fig. 2).

As for the gait silhouette sequence, we prepare an indicator set of N_{age} matrices with 128 × 128 pixels, where a matrix for a target age is set to all ones and the other matrices are all zero-padded. We also prepare the indicator set of matrices for an original age group. We then concatenate both of them for each frame of the gait silhouette sequence in channel dimension, which results in $(2N_{age}+1)$ channels in total. The input structure of the silhouette sequence is consequently a 4th-order tensor of $128\times128\times(2N_{age}+1)\times N_{img}$ (see Fig. 2). Similarly, we concatenate the indicator set of matrices for each frame of the frame difference sequence with $(N_{img}-1)$ frames, which results in a $128\times128\times(2N_{age}+1)\times(N_{img}-1)$ structure.

Once the generator takes an input triplet of period-normalized gait silhouette sequence, frame difference sequence, and a gait period at the original age group, it translates them to a triplet at the specified target age group. As for

the gait period, we simply employ 3 full connection layers to output the N_{period}-dimensional gait period indicator vector.

As for the period-normalized silhouette sequence and frame difference sequence, we extend network structure used in CycleGAN [31] by using 3D convolution to handle not a still image but a video (i.e., sequence).

Moreover, we introduce a motion augmented block (MAB) so that the model can share and exchange information about the body shape-dominant silhouette sequence and motion-dominant frame difference sequence in the intermediate layers. This is because the body shape and motion have some correlation, for example, a certain body shape (e.g., a fat body shape) may limit the gait motion (e.g., range of joint motion).

More specifically, let the feature maps of the gait sequence and the frame difference sequence at the i-th stage be C_i and M_i, respectively. We first apply three-dimensional convolution filters f^c_{conv} (*resp.*, f^m_{conv}) for C_i (*resp.*, M_i) to obtain intermediate feature maps C'_i (*resp.*, intermediate difference map M'_i) as

$$C'_i = f^c_{conv}(C_i) \tag{1}$$
$$M'_i = f^m_{conv}(M_i). \tag{2}$$

We further apply three-dimensional convolution filters $f^{c\to m}_{conv}$ (*resp.*, $f^{m\to c}_{conv}$) to the intermediate feature maps C'_i (*resp.*, intermediate difference map M'_i), and obtain the feature maps at the second stage by exchanging the appearance and motion information as

$$C_{i+1} = C'_i + f^{m\to c}_{conv}(M'_i) \tag{3}$$
$$M_{i+1} = M'_i + f^{c\to m}_{conv}(f_{diff}(C'_i)). \tag{4}$$

Note that the three-dimensional convolution filters have the same structure (i.e., kernel size 1, stride 1, dilation 1) yet have different weights. After passing two MABs, we apply some more convolution and deconvolution layers to generate a silhouette sequence and a frame difference sequence of the target age group.

D. DISCRIMINATOR WITH SLOWFAST PATH

We adopted a discriminator with a SlowFast path inspired by [35] as a primitive analogy to human visual system to better discriminate the generated progressed/regressed sequences. While sparsely sampled translated silhouette sequence (i.e., every five frames) are fed into the slow path to capture more body shape-relevant features with high channel capacity, the translated frame difference sequence with full frames are fed into the fast path to capture more motion-relevant features with low channel capacity.

Moreover, we fuse features from the fast path to the slow path by using lateral connection, which has demonstrated its effectiveness in optical flow-based two-stream network [36], [37]. Finally, output from the both paths as well as the translated gait period indicator vector are fed into an adversarial loss and age group classification loss through several layers, as described in the following subsection.

E. LOSS FUNCTIONS

As for the generator, we compute reconstruction losses with L1-norm between original and reconstructed silhouette sequences/frame differences and with a cross entropy between original and reconstructed gait period to preserve individuality as demonstrated in [7], [8]. Specifically, similar to [31], [32], [38], we apply the generator twice: original age group to target one; target age group to original one, to get reconstruction. Let $\vec{x}_{sil}$, $\vec{x}_{diff}$, and $\vec{x}_{period}$ are the silhouette sequence, the frame difference sequence, and the gait period of the original age group c_{org}, respectively, and $G_s(\vec{x}_s; c_{trg})$ is a generator to the target age group c_{trg} for a component $s \in \{sil, diff, period\}$. The reconstruction loss is then computed as

$$L_{rec} = \sum_{s\in\{sil,diff,period\}} \lambda_s \mathbb{E}_{\vec{x}_s,c_{org},c_{trg}}[d_s(\vec{x}_s, c_{org}, c_{trg})] \tag{5}$$

$$d_s(\vec{x}_s, c_{org}, c_{trg}) = \begin{cases} ||\vec{x}_s - G_s(G_s(\vec{x}_s; c_{trg}); c_{org})||_1, & \text{if } s \in \{sil, diff\} \\ f_{ce}(\vec{x}_s, G_s(G_s(\vec{x}_s; c_{trg}); c_{org}), & \text{if } s \in \{period\}, \end{cases} \tag{6}$$

where f_{ce} stands for cross entropy loss, weights λ_{sil}, λ_{diff}, and λ_{period} are set to 1, 0.01, and 1, respectively.

As for the discriminator, we apply two loss functions: an adversarial loss and a age group classification loss.

The adversarial loss is introduced to make translated gait videos realistic. More specifically, we choose Wasserstein GAN [39] to make generated sequences indistinguishable from real sequences and stabilize training as

$$L_{adv} = \lambda_{adv}\mathbb{E}_{\vec{x}}[D(\vec{x})] - \lambda_{adv}\mathbb{E}_{\vec{x},c}[D(G(\vec{x}, c))] - \lambda_{gp}\mathbb{E}_{\tilde{\vec{x}}}[(||\nabla_{\tilde{\vec{x}}} D(\tilde{\vec{x}})||_2 - 1)^2], \tag{7}$$

where $\vec{x}$ denotes a concatenated vector of $\vec{x}_{sil}$, $\vec{x}_{diff}$, and $\vec{x}_{period}$, $\tilde{\vec{x}}$ denotes a uniformly sample straight line of the concatenated vector between real and fake ones, and c is an age group label.

The age group classification loss is introduced to preserve a age group-specific property in gait. Unlike previous studies [7], [8] use a pre-trained age group classifier, we train the age group classifier through optimization of both generator and discriminator in an end-to-end manner similarly to [32]. Specifically, the age group classification loss poses constraints on real videos to optimize discriminator, and pose constraints on fake videos to optimize generator as follows

$$L^{real}_{cls} = \mathbb{E}_{\vec{x},c_{org}}[-\log D(c_{org}|\vec{x})]$$
$$L^{fake}_{cls} = \mathbb{E}_{\vec{x},c_{trg}}[-\log D(c_{trg}|G(\vec{x}, c_{trg}))]. \tag{8}$$

Finally, the full loss function composed of the losses introduced above are defined for each of the discriminator and the generator as

$$L_D = -L_{adv} + \lambda_{cls}L_{real}$$
$$L_G = L_{adv} + \lambda_{cls}L_{fake} + \lambda_{rec}L_{rec}. \tag{9}$$

TABLE 1. Statistics of subset of OULP-age.

Age group	[0, 5]	[6, 10]	[11, 15]	[16, 60]	Over 60	Total
#Training	639	4,148	2,961	15,108	687	23,543
#Test	71	462	329	1,678	76	2,616

IV. EXPERIMENTS

A. DATASETS

We trained and evaluated our model on OULP-Age dataset [40], which is the largest gait database with age information in the world. We used a subset of 26,159 subjects with ages ranging from 2 to 90 years old. We randomly chose 2,616 subjects (roughly 10% of the entire dataset for quantitative and qualitative evaluation, and used the remaining for training.

We then set the age groups. Gait reflects human's physical growing process, where the difference between neighboring age group is not necessarily the same [41], [42]. Generally, young children and teenagers (i.e., under 20 years old) grow more rapidly than adults, since the body of adults have grown into a mature physical state [43]. Therefore, instead of dividing ages with a uniform interval, we defined five age groups: [0, 5], [6, 10], [11, 15], [16, 60], and over 60 similarly to [44]. Examples of the dataset are shown in Fig. 3, and its statistics are shown in Table 1.

B. EXPERIMENTAL SETUP

We use Adam optimizer [45] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ for optimization. The learning rate starts from 0.0001 and fix for 100 epochs, then decays by 0.1 every 100 epochs, and stays at 0.000001 finally. The batch size is set to 8. Similar to [39], we perform one generator update after five discriminator updates. We compared the proposed method with the state-of-the-arts on deep generative adversarial facial aging including CAAE [21], IPCGAN [7], and S2GAN [8] qualitatively and quantitatively. For qualitative evaluation, we visualize the generated gait sequence in subsection IV-C. For quantitative evaluation, we conducted experiment on age group classification and cross-age gait recognition (i.e., identity recognition) in subsections IV-D and IV-E, respectively.

C. QUALITATIVE VISUALIZATION

We visualize the generated gait videos of different methods given target age group for comparison in Fig. 4. CAAE [21] projects the encoded vector to a latent manifold which is constrained by a simple uniform distribution, and does not well preserve individuality (see supplementary material also). While IPCGAN [7] adopts a pre-trained classifier to better preserve individuality, it produces artifacts in arm in some cases. S2GAN [8] is a state-of-the-art facial aging method and introduced a well-designed S2 module which well captures age group-specific characteristics (e.g., senior people tend to be fatter). However, it fails to preserve individuality (e.g., the woman in Fig. 4 is slimmer than general one, but the generated video by S2GAN ignores this characteristic).

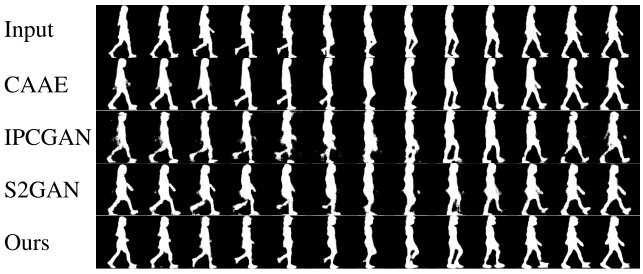


FIGURE 4. Input: Original gait video (1st row, age group [16-60]). Output: translated ones to age group [6-10] by CAAE, IPCGAN, S2GAN, and our method (from the 2nd to 5th row).

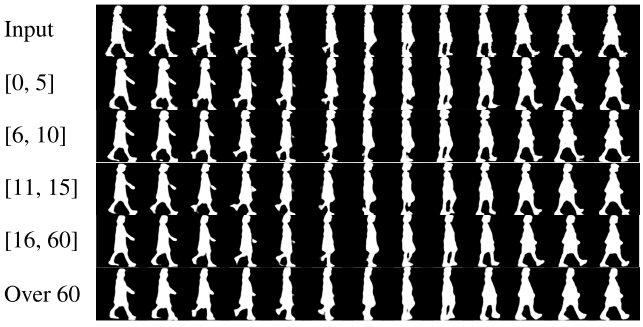


FIGURE 5. Input: Original gait video (1st row, age group [16-60]). Output: translated ones to age groups [0, 5], [6, 10], [11, 15], [16, 60], over 60 by our method (from the 2nd to 6th row).

On the other hand, our method well captures both age-specific characteristics and individuality, yielding the best qualitative result.

We also visualize the age progressed and regressed gait video for multiple age groups with our method in Fig. 5. We can see that our method successfully realizes realistic body shapes such as body-length, head-to-body ratio, leg length for kid (0–5 years old), child (6–10 years old), and teenager (11–15 years old), respectively. Meanwhile, our method generates smaller stride length in the gait video of senior age group (over 60 years old), which corresponds to the physical phenomenon that people tend to walk slower when aged.

D. AGE GROUP CLASSIFICATION

We conducted age group classification experiment in order to evaluate the generation quality of aging patterns, that is, to check whether the age progressed/regressed image truly presented the characteristics of the intended age group.

Specifically, we first designed an age group classifier using a modified ResNet-18 architecture [46] by following [32], and then trained it with real gait silhouette sequence of OULP-Age (the same training and test set split). The input dimension of first convolution layer of ResNet-18 is modified to $N_{img} = 25$ in order to handle not a single still image but a gait silhouette sequence. By classifying the generated gait silhouette sequences from different methods using the same pre-trained ResNet-18 classifier for an unseen test set, we can check to which extent each method generates age

TABLE 2. Age group classification accuracy for benchmarks. E2E indicates end-to-end training. Bold font indicates the best accuracy.

Method	Age group (Ground Truth)					Avg	E2E	#Param
	[0, 5]	[6, 10]	[11, 15]	[16, 60]	Over 60			
CAAE	10.53	58.75	35.09	92.69	16.73	42.76	√	44.4M
IPCGAN	12.16	65.63	42.24	81.50	16.17	43.54		36.4M
S2GAN	11.72	62.52	46.73	93.86	19.85	46.93		36.7M
Ours	**22.74**	**72.90**	**84.17**	**98.01**	**49.54**	**65.47**	√	39.3M

progressed/regressed video that can present the characteristics of the intended age group.

Experimental results on age group classification accuracies among benchmarks are shown in Table 2. We also report two more properties: whether the multi-age group gait sequence generation model is trainable in an end-to-end manner, and the number of network parameters.

As a result, we can see that the proposed method outperforms the other benchmarks and that it yielded the best accuracies for all of the age groups. Besides, unlike IPCGAN [7] and S2GAN [8] requires a pre-trained age group classifier to train the multi-age group translation model, the proposed method does not require it, i.e., it can train the model in an end-to-end manner, which can save training time and efforts for the pre-training. Moreover, we notice that the number of network parameters for the proposed method is comparable to the benchmarks, which shows a good scalability of the proposed method.

E. CROSS-AGE GAIT RECOGNITION

We conducted cross-age gait recognition experiments to evaluate the preservation of individuality using the age progressed/regressed generated gait sequence in addition to real ones. The dataset for the cross-age gait recognition experiment composed of three subsets: a training set, a gallery set, and a probe set. The training set contains $23,543$ subjects, while the gallery and probe sets form a set composed of the other $2,616$ subjects that are disjoint from the training set. Since each subject in OULP-Age has a single gait sequence, we generated five gait sequences per subject, which correspond to the five age groups, i.e., each subject has 6 gait sequences in total (one real and five generated).

We employed GaitSet [14], which is a state-of-the-art network structure in gait recognition for gait silhouette sequences (as opposed to static gait templates). In both training and testing phases, the input sequences are prepossessed into size of 64×64 for GaitSet requirement. In the training phase, both the generated and the original sequences of a given subject are regarded as having the same identity. In the test phase, the real sequences were assigned to the gallery, while the generated sequences were assigned to the probe.

In an identification scenario, we matched a probe to all the subjects in gallery and evaluated rank-1 identification rate based on dissimilarities (i.e., L2 norm between the final representations in the trained GaitSet network). We computed the standard deviation (uncertainty) sFRR of false rejection rate (FRR) pFRR in case of a single attempt per subject

TABLE 3. Rank-1 identification rates [%] and EER [%] (±standard deviation [%]) for each age group in probe. Bold font indicates the best performances.

Measure	Method	Age group					Avg.
		[0, 5]	[6, 10]	[11, 15]	[16, 60]	Over 60	
Rank-1	CAAE	84.9 (±0.70)	84.7 (±0.70)	84.2 (±0.71)	84.2 (±0.71)	85.4 (±0.69)	84.68
	IPCGAN	99.3 (±0.16)	99.3 (±0.16)	98.4 (±0.25)	99.1 (±0.18)	99.4 (±0.15)	99.10
	S2GAN	98.5 (±0.24)	97.3 (±0.32)	97.4 (±0.31)	98.8 (±0.21)	98.8 (±0.09)	98.16
	Ours	**99.7 (±0.11)**	**99.8 (±0.09)**	**99.8 (±0.09)**	**99.9 (±0.06)**	**99.6 (±0.12)**	**99.76**
EER	CAAE	0.93 (±0.19)	0.93 (±0.19)	0.96 (±0.19)	0.99 (±0.19)	0.94 (±0.19)	0.95
	IPCGAN	0.72 (±0.17)	0.73 (±0.17)	0.98 (±0.19)	0.91 (±0.19)	0.84 (±0.18)	0.83
	S2GAN	0.73 (±0.17)	0.88 (±0.18)	0.99 (±0.19)	0.66 (±0.16)	0.65 (±0.16)	0.78
	Ours	**0.29 (±0.11)**	**0.19 (±0.09)**	**0.29 (±0.11)**	**0.34 (±0.11)**	**0.30 (±0.11)**	**0.28**

according to [47], [48], which is represented as:

$$\sigma_{FRR} = \sqrt{\frac{p_{FRR}(1 - p_{FRR})}{n - 1}} \tag{10}$$

where n is the number of subjects and it is $2,616$ in our case. The standard deviation of true acceptance rate and rank-1 identification rate can be computed in the same manner. Take results for the age group $[0, 5]$ as an example, in the form of rank-1 identification (±the standard deviation), the following results are obtained. CAAE: 84.9%(±0.70%), IPCGAN: 99.3%(±0.16%), S2GAN: 98.5%(±0.24%), Proposed: 99.7%(±0.11%). We therefore confirmed that there is still statistical significant difference between the proposed method and the benchmark method even though the absolute difference of the rank-1 identification rate is less than 1% (between the proposed method and IPCGAN).

In a verification scenario, an input pair of a probe and a gallery is accepted as the same subject (i.e., positive sample pair) if the dissimilarity measure between them is below an acceptance threshold, and is rejected otherwise (i.e., negative sample pair or different subject pair). We computed the equal error rate (EER) of the false acceptance rate and false rejection rate as a typical performance measure. As reference, we also confirmed the statistical significant difference in terms of EER.

Experimental results of the cross-age gait recognition are summarized in Table 3. As a result, we can see that the proposed method yielded the best accuracy for all age groups, which indicates the superiority of the proposed method in terms of individuality preservation.

F. ABLATION STUDY

We made ablation studies on individual modules and report the accuracies of age group classification in Table 4.

First, in order to validate the effectiveness of the gait period, we removed the period stream in both generator and discriminator. As a result, it turns out that the accuracy decreases by approx. 15% without the gait period, which indicates that the gait period is essential for age analysis in gait.

TABLE 4. Ablation study of different modules in our network on average age-group classification accuracy [%]. We show the results for the final proposed method (top row) and compare them with the results obtained when individual modules are removed to validate their effectiveness (second to bottom rows). #Params: number of network parameters in millions.

Modules			Accuracy	#Params in million	
MAB	SlowFast	Period		Generator	Discriminator
√	√	√	**65.47**	20.01	19.35
√	√		50.34	20.00	19.30
√		√	63.21	20.01	44.83
	√	√	58.05	16.60	19.35

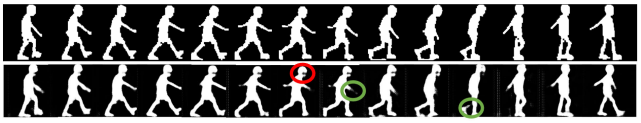


FIGURE 6. False example. Input: 1st row, original gait video of age group [0, 5] years old. Output: 2nd row, translated gait video of age group [16, 60] years old.

Second, in order to validate the effectiveness of the SlowFast path, we replaced it with two extended one-stream discriminators derived from StarGAN [32], i.e., we used the one-stream discriminator for both gait silhouette sequence stream and frame difference sequence stream. As a result, it turns out that the accuracy decreases by approx. 2% without the SlowFast path. Moreover, the SlowFast path is more efficient than the one-stream discriminator [32] w.r.t. the number of network parameters (i.e., the number of parameters with the one-stream discriminator is more than twice). This is because the SlowFast path can save the number of parameters by limiting the temporal resolution in the slow path (i.e., sampling by every five frames) as well as by limiting the channel capacity in the fast path.

Finally, we removed the MAB to validate the effectiveness of interaction between the gait silhouette sequence and the frame difference sequence. As a result, it turns out that the accuracy decreases by approx. 7% without the MAB. This indicates that interaction between motion and body shape by the MAB is essential to age group classification.

This ablation study consequently demonstrates the benefits of the MAB, the SlowFast path, and the gait period.

G. FAILURE MODE ANALYSIS

We list a typical false example in Fig. 6, where the first row is the input of age group [0, 5] years old, the second row is the output of translated ones to age group [16, 60] years old. Young children usually have larger head-to-body ratio compare to adults. When translating from age group [0,5] to [16, 60], our method can successfully generate smaller head-to-body ratio (small head for adult), but will sometimes lead to artifacts in head (red circle). Meanwhile, since our method is image-based one, generated arm and legs might not be consistent across frames (green circle). We will solve it by introducing a model-based method, which enables us to prevent from such artifacts and inconsistent body shape, in future work.

V. CONCLUSION

In this paper, we introduced an end-to-end motion augmented multi-age group gait video translation framework, which exploits both motion and body shape information from gait sequences. Specifically, we proposed three-stream generator/discriminator with the gait period, period-normalized gait silhouette sequence, and the frame difference sequences in conjunction with motion augmented block and the SlowFast path. Experiments on OULP-Age demonstrated the superiority of the proposed method quantitatively and qualitatively among other state-of-the-art methods. Future research avenues are extension to multi-view analysis and fine-grained age progression/regression (e.g., one-year interval).

REFERENCES

[1] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi, "Gait-based age estimation using a whole-generation gait database," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Washington, DC, USA, Oct. 2011, pp. 1–6.

[2] J. Lu and Y.-P. Tan, "Ordinary preserving manifold analysis for human age and head pose estimation," *IEEE Trans. Human-Machine Syst.*, vol. 43, no. 2, pp. 249–258, Mar. 2013.

[3] J. Davis, "Visual categorization of children and adult walking styles," in *Proc. Int. Conf. Audio-Video-Based Biometric Person Authentication*, Jun. 2001, pp. 295–300.

[4] R. K. Begg, M. Palaniswami, and B. Owen, "Support vector machines for automated gait classification," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 5, pp. 828–838, May 2005.

[5] S. Zhang, Y. Wang, and A. Li, "Gait-based age estimation with deep convolutional neural network," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.

[6] 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 10th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Sep. 2019, pp. 1–9.

[7] X. Tang, Z. Wang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7939–7947.

[8] Z. He, M. Kan, S. Shan, and X. Chen, "S2GAN: Share aging factors across ages and share aging trends among individuals," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9439–9448.

[9] C. Xu, Y. Makihara, Y. Yagi, and J. Lu, "Gait-based age progression/regression: A baseline and performance evaluation by age group classification and cross-age gait identification," *Mach. Vis. Appl.*, vol. 30, no. 4, pp. 629–644, Jun. 2019.

[10] 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.

[11] Z. Liu and S. Sarkar, "Simplest representation yet for gait recognition: Averaged silhouette," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2004, pp. 211–214.

[12] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, "Gait recognition using a view transformation model in the frequency domain," in *Proc. Eur. Conf. Comput. Vis.*, Graz, Austria, May 2006, pp. 151–163.

[13] C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan, "Human identification using temporal information preserving gait template," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2164–2176, Nov. 2012.

[14] H. Chao, Y. He, J. Zhang, and J. Feng, "Gaitset: Regarding gait as a set for cross-view gait recognition," in *Proc. 33rd AAAI Conf. Artif. Intell. (AAAI)*, 2019, pp. 8126–8133.

[15] Z. Zhang, L. Tran, X. Yin, Y. Atoum, X. Liu, J. Wan, and N. Wang, "Gait recognition via disentangled representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019.

[16] C. Fan, Y. Peng, C. Cao, X. Liu, S. Hou, J. Chi, Y. Huang, Q. Li, and Z. He, "GaitPart: Temporal part-based model for gait recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14225–14233.

[17] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1955–1976, Nov. 2010.
[18] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan, "Personalized age progression with aging dictionary," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3970–3978.
[19] X. Shu, J. Tang, Z. Li, H. Lai, L. Zhang, and S. Yan, "Personalized age progression with bi-level aging dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 905–917, Apr. 2018, doi: 10.1109/TPAMI.2017.2705122.
[20] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 768–783.
[21] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4352–4360.
[22] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 31–39.
[23] 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.*, vol. 6493. Queenstown, New Zealand: Springer, 2010, pp. 440–451.
[24] B. K. Yeo 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. (APSIPA)*, Dec. 2015, pp. 800–806.
[25] T. H. W. Lam, K. H. Cheung, and J. N. K. Liu, "Gait flow image: A silhouette-based gait representation for human identification," *Pattern Recognit.*, vol. 44, no. 4, pp. 973–987, Apr. 2011.
[26] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[27] T. Wang, M. Liu, J. Zhu, N. Yakovenko, A. Tao, J. Kautz, and B. Catanzaro, "Video-to-video synthesis," in *Proc. Neural Inf. Process. Syst.*, 2018, pp. 1152–1164.
[28] X. Wei, J. Zhu, S. Feng, and H. Su, "Video-to-video translation with global temporal consistency," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 18–25.
[29] A. Bansal, S. Ma, D. Ramanan, and Y. Sheikh, "Recycle-GAN: Unsupervised video retargeting," in *Proc. Eur. Conf. Comput. Vis.*, vol. 11209. Munich, Germany: Springer, 2018, pp. 122–138.
[30] Y. Chen, Y. Pan, T. Yao, X. Tian, and T. Mei, "Mocycle-GAN: Unpaired video-to-video translation," in *Proc. 27th ACM Int. Conf. Multimedia*, 2019, pp. 647–655.
[31] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.
[32] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8789–8797.
[33] Y. Makihara and Y. Yagi, "Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.
[34] C. Chen, J. Liang, H. Zhao, H. Hu, and J. Tian, "Frame difference energy image for gait recognition with incomplete silhouettes," *Pattern Recognit. Lett.*, vol. 30, no. 11, pp. 977–984, Aug. 2009.
[35] C. Feichtenhofer, H. Fan, J. Malik, and K. He, "SlowFast networks for video recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6201–6210.
[36] C. Feichtenhofer, A. Pinz, and R. P. Wildes, "Spatiotemporal residual networks for video action recognition," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3468–3476.
[37] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1933–1941.
[38] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.* (Proceedings of Machine Learning Research), vol. 70. Sydney, NSW, Australia: PMLR, 2017, pp. 1857–1865.
[39] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
[40] 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 Trans. Comput. Vis. Appl.*, vol. 9, no. 1, p. 24, Dec. 2017.
[41] B. Bogin, *The Human Pattern of Growth and Development in Paleontological Perspective* (Cambridge Studies in Biological and Evolutionary Anthropology), Cambridge, U.K.: Cambridge Univ. Press, 2003, pp. 15–44.
[42] B. Bogin, J. Bragg, and C. Kuzawa, "Humans are not cooperative breeders but practice biocultural reproduction," *Ann. Human Biol.*, vol. 41, no. 4, pp. 368–380, Jul. 2014.
[43] Z. Chengju, I. Mitsugami, F. Okura, K. Aoki, and Y. Yagi, "Growth assessment of school-age children using dualtask observation," *ITE Trans. Media Technol. Appl.*, vol. 6, no. 4, pp. 286–296, Oct. 2018.
[44] 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, Nov. 2018.
[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., San Diego, CA, USA, May 2015. [Online]. Available: http://arxiv.org/abs/1412.6980
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[47] A. Jain, R. Bolle, and S. Pankanti, "Introduction to biometrics," in *Biometrics*. Boston, MA, USA: Springer, 1996, pp. 1–41.
[48] A. J. Mansfield and J. L. Wayman, "Best practices in testing and reporting performance of biometric devices," Nat. Phys. Lab. Queens Road, London, U.K., Tech. Rep., 2002. [Online]. Available: http://www.idsysgroup.com/ftp/BestPractice.pdf



YIYI ZHANG received the B.E. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2017, where she is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering. She is also pursuing the dual degree with the Graduate School of Information Science and Technology, Osaka University. Her research interests include generative models, gait recognition, and image processing.



YASUSHI MAKIHARA received the B.S., M.S., and Ph.D. degrees in engineering from Osaka University, in 2001, 2002, and 2005, respectively. He was appointed as a Specially Appointed Assistant Professor (full-time), an Assistant Professor, and an Associate Professor with The Institute of Scientific and Industrial Research, Osaka University, in 2005, 2006, and 2014, respectively, where he is currently a Professor with the Institute for Advanced Co-Creation Studies. His research interests include computer vision, pattern recognition, and image processing, including gait recognition, pedestrian detection, morphing, and temporal super resolution. He is also a member of IPSJ, IEICE, RSJ, and JSME. He has obtained several honors and awards, including the 2nd International Workshop on Biometrics and Forensics (IWBF 2014), the IAPR Best Paper Award, the 9th IAPR International Conference on Biometrics (ICB 2016), the Honorable Mention Paper Award, and the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology, and prizes for Science and Technology Research Category in 2014. He has served as an Associate Editor-in-Chief for *IEICE Transactions on Information and Systems*, an Associate Editor for *IPSJ Transactions on Computer Vision and Applications* (CVA), a Program Co-Chair for the 4th Asian Conference on Pattern Recognition (ACPR 2017), and area chairs for ICCV 2019, CVPR 2020, and ECCV 2020.



DAIGO MURAMATSU (Member, IEEE) received the B.S., M.E., and Ph.D. degrees in electrical, electronics, and computer engineering from Waseda University, Tokyo, Japan, in 1997, 1999, and 2006, respectively. He is currently an Associate Professor with the Institute of Scientific and Industrial Research, Osaka University. His current research interests include gait recognition, signature verification, and biometric fusion. He is also a member of the IEICE and the IPSJ.



JIANFU ZHANG received the B.E. degree from Shanghai Jiao Tong University, Shanghai, China, in 2015, where he is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering. His current research interests include deep learning, adversarial learning, and generative models.



LI NIU received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2011, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2017. He is currently an Associate Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. Before joining Shanghai Jiao Tong University, he was a Postdoctoral Associate with Rice University, Houston, TX, USA. His current research interests include machine learning, deep learning, and computer vision.



LIQING ZHANG (Member, IEEE) received the Ph.D. degree from Sun Yat-sen University, Guangzhou, China, in 1988. He was a Research Scientist with the RIKEN Brain Science Institute, Japan, from 1997 to 2002. He is currently a Full Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His research interests include visual cognitive computing, uncertainty reasoning, and statistical machine learning.



YASUSHI YAGI (Member, IEEE) received the Ph.D. degree from Osaka University, in 1991. In 1985, he joined the Product Development Laboratory, Mitsubishi Electric Corporation, where he worked on robotics and inspections. He was with Osaka University as a Research Associate in 1990, a Lecturer, in 1993, an Associate Professor, in 1996, and a Professor, in 2003, where he was the Director of the Institute of Scientific and Industrial Research, from 2012 to 2015, and the Executive Vice President, from 2015 to 2019. He is currently a Professor with the Institute of Scientific and Industrial Research, Osaka University. His research interests include computer vision, medical engineering, and robotics. He is a Fellow of IPSJ and a member of IEICE and RSJ. He was awarded the ACM VRST2003 Honorable Mention Award, the IEEE ROBIO2006 Finalist of T. J. Tan Best Paper in Robotics, the IEEE ICRA 2008 Finalist for Best Vision Paper, the MIRU 2008 Nagao Award, and the PSIVT 2010 Best Paper Award. He served in the various international conferences, including the Financial Chair for FG 1998 and PSVIT 2009, the Organizing Chair for OM-INVIS 2003, the Program Co-Chair for ROBIO 2006 and ACPR 2011, the Program Chair for ACCV 2007, the Technical Visit Chair for ICRA 2009, and the General Chair for ACCV 2009 and ACPR 2013. He has also served as an Editor for IEEE ICRA Conference Editorial Board from 2007 to 2011. He is also an Editorial Member of the *International Journal of Computer Vision* and the Editor-in-Chief of the *IPSJ Transactions on Computer Vision and Applications*.