Uncertainty-Aware Gait-Based Age Estimation and Its Applications

Chi Xu, Atsuya Sakata, Yasushi Makihara, Noriko Takemura, Daigo Muramatsu, *Member, IEEE*, Yasushi Yagi, *Senior Member, IEEE*, and Jianfeng Lu

***Abstract*—Gait-based age estimation is a key technique for many applications. It is well known that age estimation uncertainty is highly dependent on age (i.e., small for children and large for adults), and it is important to know the uncertainty for the above-mentioned applications. Therefore, we propose a method for uncertainty-aware gait-based age estimation by introducing a label distribution learning framework. Specifically, we design a network that takes an appearance-based gait feature as input and outputs discrete label distributions in the integer age domain. We then train the network to minimize a loss function, which is defined as the dissimilarity between the estimated age distribution and the ground-truth age distribution, in addition to the conventional mean absolute error for the estimated age. Additionally, we demonstrate that uncertainty-aware gait-based age estimation is beneficial for two applications: person search by age query and people counting by age group. Experiments on the world's largest gait database, OULP-Age, demonstrated that the proposed method can successfully represent age estimation uncertainty, and outperforms or is comparable with state-of-the-art methods in terms of age estimation accuracy. Moreover, we demonstrated the effectiveness of the uncertainty-aware framework in applications to person search and people counting through experiments on the database.**

***Index Terms*—Gait, age, label distribution learning, people counting, person search, uncertainty.**

Manuscript received December 26, 2020; revised March 31, 2021; accepted May 9, 2021. Date of publication May 14, 2021; date of current version December 1, 2021. This work was supported in part by JSPS KAKENHI under Grant JP18H04115, Grant JP19H05692, and Grant JP20H00607; in part by the Jiangsu Provincial Science and Technology Support Program under Grant BE2014714; in part by the 111 Project under Grant B13022; and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions. This article was recommended for publication by Associate Editor I. Kakadiaris upon evaluation of the reviewers' comments. *(Corresponding author: Chi Xu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the publicly available dataset OULP-Age.

Chi Xu, Atsuya Sakata, Yasushi Makihara, and Yasushi Yagi are with the Institute of Scientific and Industrial Research, Osaka University, Ibaraki 567-0047, Japan (e-mail: xu@am.sanken.osaka-u.ac.jp; sakata@am.sanken.osaka-u.ac.jp; makihara@am.sanken.osaka-u.ac.jp; yagi@am.sanken.osaka-u.ac.jp).

Noriko Takemura is with the Institute for Datability Science, Osaka University, Suita 565-0871, Japan (e-mail: takemura@am.sanken.osaka-u.ac.jp).

Daigo Muramatsu is with the Department of Computer and Information Science, Faculty of Science and Technology, Seikei University, Musashino 1808633, Japan (e-mail: muramatsu@st.seikei.ac.jp).

Jianfeng Lu is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: lujf@njust.edu.cn).

Digital Object Identifier 10.1109/TBIOM.2021.3080300

I. Introduction

FOR THE last two decades, gait has been regarded as a unique behavioral biometric that can even be used at a distance from a camera without the subjects' cooperation [1]. Because of the above-mentioned advantages, gait-based person verification/identification is expected to be applied to many applications, such as surveillance, access control, and criminal investigation using closed-circuit television (CCTV) footage. In fact, there have been several cases in which gait recognition has been applied to criminal investigation and forensics [2]. In addition to identity [3], [4], [5], [6], gait may contain a variety of cues, such as gender [7], [8], [9], age [10], [11], [12], [13], [14], [15], [16], ethnicity [17], disease [18], emotion [19], and some qualitative gait attributes [20].

Gait-based age estimation has many potential applications. For example, once a visitor's age is estimated in a shopping mall, an advertiser can change the content of digital signage to be more suitable for the estimated age. Age-based access control is also possible based on the estimated age, for example, underage people can be prevented from buying alcohol or cigarettes. Moreover, in a forensic scenario, a criminal investigator may obtain an eyewitness report regarding rough age information about a perpetrator/suspect, and thereafter may be able to automatically retrieve candidates from CCTV footage whose estimated age by gait matches the eyewitness report. It would also be possible to search CCTV footage for people such as wandering elderly and lost children with the help of a gait-based age estimator.

Technically, relevant work on gait-based age analysis mainly falls into two families: age-group classification [10], [11], [22] and age regression [12], [13], [14], [15], [23]. The above-mentioned studies use classical machine learning techniques, such as a support vector machine (SVM) [24] and support vector regression (SVR) [25] because of the limited number of training data (e.g., [4]); however, recent studies on gait-based age estimation mainly use deep learning-based approaches [16], [26], [27] because much larger-scale gait databases have been released recently that include over 60,000 subjects [28].

Regardless of the approach used for gait-based age estimation, we almost always observe that age estimation accuracies are indeed age-dependent. By closely observing the scatter plot between the ground-truth age and estimated age using one of the above-mentioned deep learning approaches [21] (see Fig. 1 (left)), we can clearly see that the uncertainty (or degree of the spread of ground-truth ages) for each estimated age is

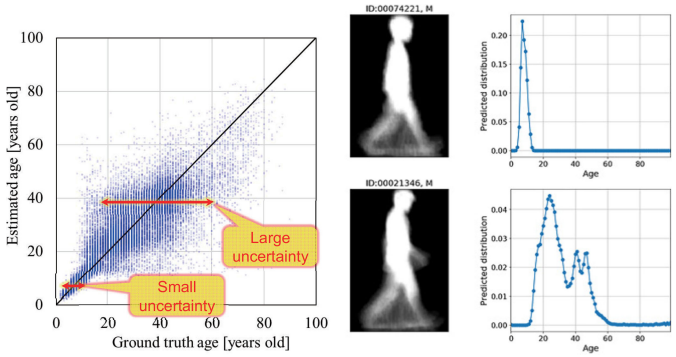


Fig. 1. Example of a scatter plot between the ground-truth age (horizontal axis) and estimated ages (vertical axis) using [21] (left), and probability distributions of estimated ages using the proposed method (right) given the input of a silhouette-based gait feature (middle). We obtain a relatively sharp age probability distribution (i.e., one with small uncertainty) for the upper subject (a child); however, we obtain a relatively spread probability distribution (i.e., one with large uncertainty) for the lower subject (an adult). The proposed method indicates that more age variations are likely to occur for the lower subject's gait feature than the upper one.

quite different between a child and an adult. Specifically, the uncertainty is small for a child (e.g., less than ± 2 years), but quite large for an adult (e.g., ± 20 years). In fact, it is relatively difficult even for humans to accurately determine the age of subjects from their gait.

As a trial, we would like readers to guess the ages of the two subjects in Fig. 1 (middle) given the input of silhouette-based gait features called the gait energy image (GEI) [29]. Regarding the upper subject, we expect that readers can determine a relatively fine-grained age estimation range (e.g., from 5 to 10 years old). Conversely, we expect that readers will struggle in the age estimation for the lower subject. Some readers may estimate it as 20 years old, whereas others may estimate it as 30 or 40 years old.[1] The age-dependent uncertainty difference between gait-based age estimations occurs for several reasons. First, the growth or decline of gait (and body shape) is relatively fast for children and the elderly, whereas it is quite slow for adults; hence, it is relatively difficult to determine the age difference in adulthood because of the slow change of the gait feature. Second, whereas facial images have texture cues to enable the estimation of age (e.g., wrinkles and dull skin appear more as age progresses), a gait image does not have such texture-based cues. Third, although a gait image still has some cues to enable age estimation, such as stoop and middle-age spread, they not only depend on age but also highly depend on individuality and individual lifestyle habits. For example, the gait of a person in his/her 40s, who maintains a slim body through good dietary and exercise habits may appear younger than his/her age, and vice versa; that is, there may be some subjects whose gait features are almost the same as those whose ages are different, which results in an uncertain age estimation result from the gait.

This type of uncertainty (or confidence) in age estimation plays an important role in some applications. We consider the scenario of a person search by age query. In the case of searching for a lost child of 5 years old, because the system is relatively confident in its estimation of human age for children, it would be sufficient to show a list of people whose estimated age is within a limited age range (e.g., 5 ± 2 years). Conversely, in the case of searching for a suspect in his/her 30s, it would be necessary to show many people whose estimated age is within a large age range (e.g., from 20s to 40s). To summarize, if the system is aware of the uncertainty (or confidence) of the estimated age, it can appropriately bound the age range of candidates for a search target.

Most existing approaches to gait-based age estimation output a single age as an estimation result; there is one exception [12] that outputs an estimated age and its uncertainty. The study [12] used a framework of Gaussian process regression (GPR) for gait-based age estimation, which outputs a Gaussian distribution for an estimated age (i.e., a mean and variance) given an input gait feature in addition to a training set of gait features and corresponding ages. The uncertainty (i.e., the variance of the estimated age) provided by the GPR framework is mainly derived from the degree of closeness between the test gait feature and the training gait features; that is, the uncertainty decreases as the test gait feature becomes relatively closer to one of the training gait features, and vice versa. In this sense, the GPR framework cannot manage age uncertainty derived from similar gait features with different ages, as mentioned above. Specifically, the system should ideally return a large uncertainty if a test gait feature is close to a cluster of similar training gait features but with different ages, whereas the GPR returns a small uncertainty, even in such a scenario.

To overcome such a difficulty, we propose a deep neural network-based approach to gait-based age estimation that outputs not a single estimated age but a probability distribution of the estimated age. The contributions of this work are two-fold.

1) *Uncertainty-aware gait-based age estimation using a label distribution:* Unlike the GPR framework, the proposed method can better cope with the above-mentioned similar gait features but with different ages. Specifically, we introduce a label distribution [30] as the output of an age estimation network given a gait feature as input. The age estimation network is trained to minimize the joint loss function of (1) the Kullback–Leibler (KL) divergence [31] for the uncertainty representation; and (2) mean absolute error (MAE) for the ordinary preserving property in the same manner as state-of-the-art facial age estimation [32]. Thus, we can assign probabilities to multiple ages, as shown in Fig. 1 (right).

2) *State-of-the-art accuracy for gait-based age estimation:* We achieved state-of-the-art accuracy for gait-based age estimation for standard evaluation metrics, such as the MAE and cumulative score (CS) on OU-ISIR Gait Database, Large Population Dataset with Age (*OULP-Age*) [28], which is the world's largest publicly available gait database that contains over 60,000 subjects with wide age diversity ranging from 2 to 90 years old.

In addition to the above-mentioned original contributions in our conference paper [33], we have made two extensions in this version. One is the more effective use of the proposed uncertainty-aware method in two applications: person search by age query and people counting by age group. The other is

[1]The answer is 30 years old.

the use of a more sophisticated backbone network for better accuracy.

Regarding the person search, we consider a toy example in which the two subjects in Fig. 1 are enrolled in a database, and a conventional uncertainty-unaware method estimates the ages of the upper and lower subjects in the figure as 8 and 33 years old, respectively. If we search the database for a person of 20 years old, a straightforward yet reasonable approach is to create a rank list based on the difference between the query age (i.e., 20 years old) and the estimated age for each person in the database. Using this method, the upper subject would be the first candidate (i.e., the first rank) because its difference (i.e., $20-8 = 12$ years) is smaller than that of the lower subject (i.e., $33 - 20 = 13$ years), even though the lower subject (i.e., an adult) is actually more likely to be 20 years old than the upper subject (i.e., a child). Conversely, the proposed uncertainty-aware method may create a rank list based on the likelihood (i.e., the probability) of the query age, and consequently, the lower subject would be the first candidate because its probability for the query age (i.e., 20 years old) is higher than that for the upper subject (i.e., over 0.03 and approximately zero for the lower and upper subjects, respectively, which are both read from Fig. 1 (right)). Thus, the proposed uncertainty-aware method contributes to a more efficient person search by age query.

People counting by age group is another potential application. For example, counting the number of visitors to facilities, such as shopping malls and museums, by age group is important for market research and providing services according to customer attributes, such as age and gender. A straightforward yet reasonable approach to count people by age group, is to estimate an age for each detected person and then create a histogram by casting one vote to an age group bin that includes the estimated age. A conventional gait-based age estimator, however, suffers from biases, for example, elderly subjects (i.e., over 60s) tend to be underestimated, as shown in Fig. 1 (left); hence, the age group bin for the elderly may receive fewer votes than the actual value. Conversely, the proposed uncertainty-aware method may cast multiple votes that are weighted by probability for each age group; hence, we have more opportunities to cast votes to the elder's bin. Consequently, we may mitigate the above-mentioned bias problem as a result of the estimated age probability distribution.

To summarize, the extensions of this version of the study are two-fold.

3) *Uncertainty-aware person search and people counting:* We did not quantitatively evaluate the goodness of the estimated uncertainty in the previous version of the study [33], whereas we attempt this in the present study through two applications: person search by age query and people counting by age group. Both applications use uncertainty, specifically, the age probability distribution using the proposed method, unlike conventional person search and people counting, which do not consider uncertainty. We show the effectiveness of the proposed uncertainty-aware method quantitatively through the two applications.

4) *More effective backbone network:* We used a relatively simple network architecture, that is, GEINet [34], as the backbone network in the previous version of the study [33], whereas we introduce GaitSet [35] as the backbone network in the extension in the present study, which has been proven to be more effective in gait recognition. Consequently, we further advance state-of-the-art gait-based age estimation.

II. Related Work

A. Gait-Based Age Estimation

Early stage studies on gait-based age analysis mainly considered age group classification problems (e.g., children vs. adults) using body joint-based gait parameters. For example, Davis [10] classified children and adults based on a biological motion cue (i.e., point light sources on the body joints) and Begg [11] classified young and elderly people based on the minimum foot clearance from the ground [11]. Mannami *et al.* [22] analyzed the differences between an image-based gait feature, such as GEI [29] (a.k.a. averaged silhouette [36]), and a frequency-domain feature [5] from multiple views among four age/gender groups: children, adult females, adult males, and the elderly.

Studies on gait-based age estimation have been conducted since 2010 using image-based gait features [5], [29] in conjunction with classical machine learning techniques. For example, Lu and Tan proposed ordinary preserving manifold learning in [13], and also proposed its extensions: ordinary preserving linear discriminant analysis (OPLDA) and ordinary preserving margin Fisher analysis (OPMFA) in [14]. Xu *et al.* [28] constructed the world's largest gait database called OULP-Age and evaluated multiple benchmark methods for gait-based age estimation, including [13], [14] and also a method using SVR [25]. Li *et al.* [23] proposed a multi-stage framework that combines age group classification and subsequent age regression for each age group using several machine learning techniques, such as an SVM, manifold learning, and SVR.

In addition to the above-mentioned classical machine learning-based approaches, researchers have started to use deep learning-based approaches for gait-based age estimation. For example, Sakata *et al.* [21] used DenseNet [37], which is a state-of-the-art deep neural network architecture, and validated its effectiveness in the gait-based age estimation task. In addition to age, other attributes, such as gender, age group, and/or identity, have been incorporated into multi-task learning frameworks [16], [26] and a multi-stage learning framework [38]. Moreover, a gait-based age estimation method that is robust against a carried object was proposed in [27].

All the above-mentioned approaches output a single value as the estimated age without taking uncertainty into consideration, whereas [12] used a GPR framework [39] for gait-based age estimation, which outputs a Gaussian distribution of the estimated age, specifically, its mean and variance (i.e., a type of uncertainty). As we addressed in Section I, the GPR framework, however, cannot well manage the uncertainty induced by similar gait features but with different ages. We provide more details of this issue with a preliminary simulation experiment

in Section III. Unlike the GPR framework, our method can better handle uncertainty using a label distribution framework, where the system outputs a discrete probability distribution over integer age labels as output given a gait feature as input.

The proposed method may be similar to a previous study in terms of the representation of integer age labels [15]. Specifically, Lu and Tan [15] casted an age estimation problem into a classification problem of integer-age classes (labels) using multi-label guided (MLG) subspace learning. Although they used integer age labels similarly to the proposed method, they still output a single integer age label as a result of a classification process, unlike the proposed method, which outputs an age probability distribution.

B. Face-Based Age Estimation Using Label Distribution

Because a face is the most frequently used biometric modality for age estimation, there is a rich body of literature on face-based age estimation. We briefly introduce existing work on face-based age estimation using label distribution because it is most closely related to our work.

Geng *et al.* [40] introduced the concept of label distribution learning into face-based age estimation, and proposed two learning algorithms: improved iterative scaling-learning from label distribution and a conditional probability neural network. Zhao and Wang [41] proposed strategic decision-making learning from the label distribution, which copes with different types of age label distribution, such as Gaussian-type, triangle-type, and box-type distributions. He *et al.* [42] proposed data-dependent label distribution learning, where training samples neighboring a test sample are first selected based on the face affinity graph, and the label distribution is then constructed based on the cross-age correlations among neighboring face samples.

Gao *et al.* [30] introduced deep learning-based approaches to facial age estimation with a label distribution called deep label distribution learning (DLDL), which uses KL-divergence to measure the distribution similarity. They also extended DLDL in [32] to cope with a regression problem in addition to the label distribution estimation simultaneously by introducing the joint loss function of an MAE and KL divergence.

As mentioned above, label distribution has already been used in the facial age estimation community. The gait analysis community has never used the useful label distribution for age estimation; hence, in this study, we use it for the purpose of gait-based age estimation for the first time, to the best of our knowledge.

C. Person Search on Image Data

Most person search methods on image data typically detect persons first, and then search a pool of detected persons for a target person that matches the query.

Query types used in person search mainly fall into three categories: image, attribute, and natural language. Image query-based approaches are the most direct approaches to search for a person, where the queries are given as images and/or biometric features extracted from the images. Given a query, person search is performed using person reidentification [43], [44], [45], [46], where color and texture information are mainly used, or biometric person authentication using a variety of modalities, such as face and gait.

Attribute query-based approaches do not require a query image, but instead use pre-defined attributes for the person search [47], [48]. For example, the person search attributes range from soft biometric-based attributes, such as age and gender [7], [8], [9], [16], [26], [32], to height and hair color computed from the image [48], which are estimated for detected persons and also specified as a query. Other types of attributes are action-based labels [49], [50], and clothes information associated with color, texture, and clothing types [48].

Natural language query-based approaches use a description in natural language as a query, and retrieve a person that matches the description [51], [52]. For example, in [52], at least eight words were used to describe the target person, and information associated with clothes and actions were described.

Although a variety of approaches to person search have been proposed, as mentioned above, the person search research community has never explicitly incorporated the uncertainty of estimated cues for each detected person (e.g., the uncertainty of age), to the best of our knowledge.

D. People Counting by Age

People counting by age group is typically performed by detecting people, estimating the age groups of the detected people, and counting them for each age group. For example, Ko *et al.* [53] proposed a method for people counting for four age groups using RGBD cameras installed on the front and ceiling. The age groups were estimated based on facial feature points and the texture of wrinkles.

Additionally, the industrial community has been developing a large number of age-specific people counting tools, such as People Face Analytics (Apexcount),[1] PeopleCounting System (MintM),[2] Demographic Analysis (V-Count),[3] Demographic Facia (GikenTrastem),[4] FieldAnalyst (NEC),[5] and TrueView People Counter (Cognimatics).[6] This shows that there is a high demand for people counting by age group.

III. Observation of the GPR-Based Approach [12]

Before moving on to the proposed method, we briefly mention the most closely related work, that is, the GPR-based approach to uncertainty-aware gait-based age estimation, which outputs the Gaussian distribution of the estimated age, that is, the expectation and variance, to clarify its drawback. Readers may refer to [12] for more details.

In the GPR-based approach, we assume that a training set $D = [X, \boldsymbol{y}]$ is given, where $X = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$ is a set of

[1] http://apexcount.com/people-face-analytics/

[2] https://mintm.com/people-counting-system/

[3] https://v-count.com/solutions/demographic-analysis/

[4] http://www.trastem.co.jp/eng/product/demographic.html

[5] https://www.nec-solutioninnovators.co.jp/en/sl/fieldanalyst/index.html

[6] https://netcam.cz/produkty/software-sprava-videa/pdf/trueview-people-counter-ps.pdf

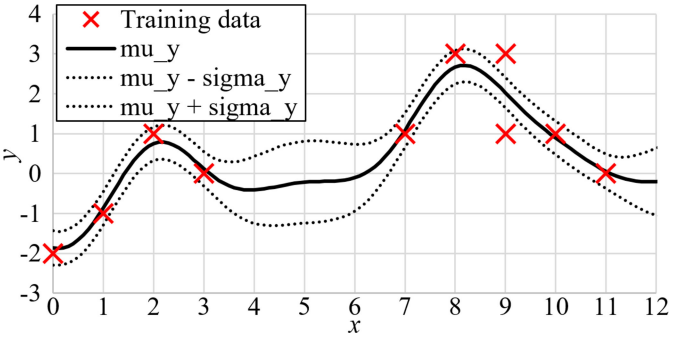


Fig. 2. Preliminary simulation experiment for GPR with 10 training data and hyperparameters $r = 1.0$ and $\Delta = 0.1$. The uncertainty range $\mu_y \pm \sigma_y$ increases if an input is more distant from any of the training data (e.g., $x = 5$). Conversely, uncertainty does not become large, even if an input is closer to the training data with multiple outputs (e.g., $x = 9$ with $y = 1$ and $y = 3$).

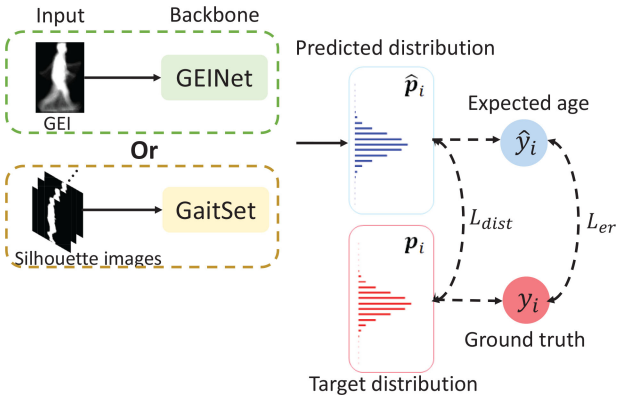


Fig. 3. Schematic of the proposed age estimation model with label distribution learning. Either GEINet [34] or GaitSet [35] is adopted as the backbone network.

N samples of gait features and $\boldsymbol{y} = [y_1, \ldots, y_N]$ is the set of corresponding ground-truth ages. Additionally, an affinity/similarity function, that is, an inner product between two feature vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, is defined, often using a nonlinear kernel function, such as a radial basis function (RBF):

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j; r) = \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2r^2}\right), \tag{1}$$

where $\|\cdot\|$ is the L_2 norm and r is a hyperparameter for the RBF kernel. Note that the function $k(\cdot, \cdot)$ measures the closeness between two input arguments, and it approaches 1 if the two inputs are more similar, whereas it approaches 0 if they are more different.

Thereafter, given an input gait feature $\boldsymbol{x}_*$, the GPR estimates the posterior probability distribution of age y_* corresponding to the gait feature $\boldsymbol{x}_*$ based on the training set D. According to Gaussian process theory [39], once we assume each age y_i in the training set D, follows a Gaussian distribution $\mathcal{N}(y; y_i, \Delta^2)$, where Δ^2 is the variance of age observation noise, and the posterior probability distribution $P(y_*|\boldsymbol{x}_*, D)$ also follows a Gaussian distribution $\mathcal{N}(y_*; \mu_y, \sigma_y^2)$, where mean μ_y and variance σ_y^2 are defined as

$$\mu_y = \boldsymbol{k}_*^T (K + S)^{-1} \boldsymbol{y} \tag{2}$$

$$\sigma_y^2 = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^T (K + S)^{-1} \boldsymbol{k}_* + \Delta^2, \tag{3}$$

respectively, where K is an $N \times N$ square matrix whose (i, j)-th component is $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $\boldsymbol{k}_*$ is an N-dimensional vector whose i-th row is $k(\boldsymbol{x}_i, \boldsymbol{x}_*)$, and S is an $N \times N$ diagonal matrix whose (i, i)-th component is age observation noise Δ^2.

From Eq. (3), we note that the variance (i.e., the uncertainty) is dependent not on the ground-truth ages $\boldsymbol{y}$ in the training set but relations, specifically, affinity/similarity/closeness between the input gait feature $\boldsymbol{x}_*$ and those in the training data X, which appear in K and $\boldsymbol{k}_*$ in the above-defined equations.

To better observe and understand this, we conducted a preliminary simulation experiment with 10 training data of one-dimensional feature vectors (i.e., a scalar value) and an estimation target (e.g., an age). The results are shown in Fig. 2. We can see that uncertainty increases as the input becomes more distant from any of the training data (e.g., $x = 5$). Moreover, in this simulation experiment, the training sample $x = 9$ had multiple outputs $y = 1$ and $y = 3$, which is analogous to subjects who have similar gait features but different ages. Even though it is preferable to represent uncertainty caused by different ages well in the sample, the uncertainty derived from the GPR did not become large. This is the drawback of the GPR framework.

Conversely, the label distribution learning framework can successfully cope with this type of uncertainty because it can be trained to output multiple ages. For example, given the sample $x = 9$ in the above-mentioned simulation experiment, we intuitively expect that probability 0.5 will be assigned to each of $y = 1$ and $y = 3$.

IV. Uncertainty-Aware Gait-Based Age Estimation

A. Overview

The proposed method attempts to estimate the age of a target person from the gait. We provide an overview of the proposed method in Fig. 3. Given a gait feature or silhouette sequence as input, the backbone network (i.e., an age estimation model) outputs the label distribution of the estimated age. GEINet [34] and GaitSet [35] are used as backbone networks to compare their performance. Once the label distribution is obtained, we can also compute the expectation of the label distribution as a single estimated age. We explain the details of the individual components in the following subsections.

B. Input Data

The first step to achieve gait-based age estimation is to prepare input data for the age estimation model. For this purpose, first, we need to capture gait data, and then extract an efficient gait feature from the gait data. We can consider multiple sensors, such as image sensors (cameras), a depth sensor, and an inertial sensor. Among them, cameras are the most popular sensors, and are already installed in many places (e.g., CCTV in public spaces). Therefore, we focus on the gait image sequence captured using cameras as gait data.

However, the texture and color, which are important information contained in the captured RGB images, are not

the main cues that include useful age information for gait-based age estimation. For examples, children and adults may wear the same color and type of clothing, and people (e.g., a suspect) may also change clothes every day. Additionally, the faces, which contain rich age-related texture and color features, cannot be well observed in surveillance scenarios, particularly when observed at a distance from a camera, due to low resolution or back observation view. On the other hand, the body shapes and walking styles show apparent differences among different age groups (e.g., large head-to-body ratio of children, small stride of the elderly), which are also clearly reflected in silhouette images. We therefore use silhouettes instead of RGB images. We then extract the GEI [29] from the obtained gait silhouette sequence, which is the most widely used image-based gait feature, and use it as input to GEINet. Conversely, a size-normalized silhouette sequence [5] is directly fed into GaitSet, which is actually interpreted as a set of silhouettes in the network.

C. Network Structure

As the backbone network, we first use a convolutional neural network (CNN) designed for gait recognition called GEINet [34] and modify it for age estimation. The modified GEINet is composed of two sequential triplets, which include convolution, pooling, and normalization layers, fully connected layers with normalization, and another fully connected layer with the softmax activation function. Readers may refer to the layer configurations in the conference version of this paper [33].

The second backbone network we use is GaitSet [35], which is a state-of-the-art set-based gait recognition network that exhibits more effective gait recognition performance than GEINet. The structure of GaitSet is much more complex than that of GEINet. It contains a CNN for frame-level feature extraction, a set pooling operation for set-level feature aggregation, and a module called horizontal pyramid mapping for discrimination learning (see the original paper [35] for more details). A fully connected layer with the softmax activation function is added after the original architecture of GaitSet to output the estimated age label distribution.

The number of units for the last fully connected layer in both GEINet and GaitSet is set to the number of bins for discrete label distribution, as explained in Section IV-D.

D. Output Representations

Two major output/ground-truth representations exist: a scalar value for regression-based methods [13], [14], [21], [23], [38] and a one-hot vector for classification-based methods [15]. Different from these approaches, in this paper, we incorporate the idea of label distribution [32], [40]. A label distribution-based method assumes that the ground-truth is neither a scalar value nor a one-hot vector, but a discrete age distribution. Fig. 4 shows the conceptual difference between these three representations.

Let K be the number of bins for a discrete probability distribution of integer ages; we set the minimum and maximum ages to 0 and $K-1$, respectively. Let $y_i \in \mathbb{R}$ be the ground-truth age

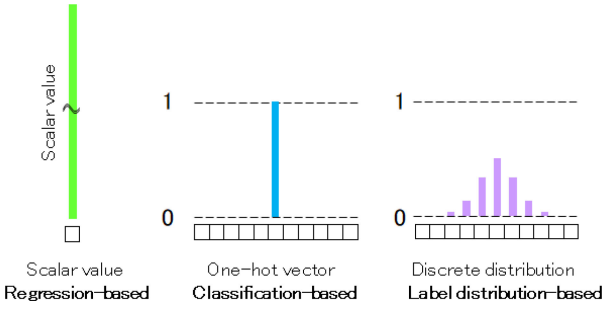


Fig. 4. Conceptual difference between output representations.

of the i-th training data. Additionally, let $\boldsymbol{p}_i$ and $p_{i,k}$ be assigned a discrete probability distribution and probability for age k associated with the i-th data, respectively. In the proposed method, we set $p_{i,k}$ so that it follows a Gaussian distribution whose mean and standard deviation are the ground-truth age y_i and σ, respectively:

$$p_{i,k} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(k-y_i)^2}{2\sigma^2}\right\}, \tag{4}$$

where σ is a hyperparameter that controls the uncertainty of the ground-truth age.

E. Loss Function

Because the ground-truth label is described by a discrete probability distribution, we can consider two criteria for evaluating the goodness of the trained parameters. The first criterion measures the similarity between the assigned target probability distribution and the estimated probability distribution. The other criterion is the dissimilarity between the ground-truth age and the expected age calculated from the estimated probability distribution.

Let $\hat{\boldsymbol{p}}_i = [\hat{p}_{i,0}, \ldots, \hat{p}_{i,K-1}]^T \in \mathbb{R}^K$ be the estimated discrete probability distribution for the i-th training data ($i = 1, \ldots, N$), where N is the number of training data. Note that the integer age for the k-th age label is k and its probability is $\hat{p}_{i,k}$.

For the first criterion, we consider KL divergence [31] between two distributions and set the loss function L_{KL} as follows:

$$L_{\mathrm{KL}} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{K-1} p_{i,k} \ln \frac{p_{i,k}}{\hat{p}_{i,k}}. \tag{5}$$

For the other criterion, we calculate the expected age from the estimated distribution, and measure the MAE between the ground-truth age and the expected age as follows:

$$L_{\mathrm{MAE}} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i|, \tag{6}$$

$$\text{where } \hat{y}_i = \sum_{k=0}^{K-1} k\hat{p}_{i,k}. \tag{7}$$

Finally, we define a joint loss function L using the two loss functions as

$$L = \lambda_{\mathrm{KL}} L_{\mathrm{KL}} + \lambda_{\mathrm{MAE}} L_{\mathrm{MAE}}, \tag{8}$$

where λ_{KL} and λ_{MAE} are hyperparameters that balance the two loss functions.

V. Applications to Person Search and People Counting by Age

Two typical applications of gait-based age estimation are person search and people counting in surveillance scenarios, for example, searching for a lost child of a known age, and estimating customers' statistics by age group for marketing research in a shopping mall. In this section, we introduce the details of applications to person search and people counting using the estimated age label distribution using the proposed method.

Note that we assume that person detection is a preprocessing step for the two applications that has been completed because person detection is out of scope of this study. Specifically, we assume that we have already detected N persons from video sources, such as CCTV footage, and constructed a database composed of the N persons with their estimated ages.

A. Person Search by Age

Because the database has already been constructed, the remaining task for person search is to create a ranking list of all the persons in the database given a query, which will be shown to the user who wants to search for a target person. In the person search application, the greater the number of persons with a true match to the given query located at high ranks, the more effective the system.

As application scenarios, we consider two types of queries: a specific age query and an age group (i.e., age range) query. For example, parents may look for their lost child in a shopping mall by specifying his/her exact age (e.g., 5 years old); and the police may search for an escaped suspect with a possible age range (e.g., around 30 to 40 years old) provided by an eyewitness.

1) Person Search by Age Query (Uncertainty-Unaware Baseline): Before describing the approach to person search using the proposed uncertainty-aware method, we introduce a baseline version that uses a single estimated age without uncertainty to clarify the differences. Let $\hat{y}_i$ be a single estimate age for the i-th person in the database. Given a query age $y^q \in \{0, 1, \ldots, M\}$, where M is the maximum age for the query, a straightforward yet reasonable approach is to first compute the dissimilarity to the query age for each person as $d_i = |\hat{y}_i - y^q|$, and then create a ranking list of all persons in a database by sorting the dissimilarities d_i in ascending order.

Proposed uncertainty-aware method: As shown in the example in Section I, the uncertainty-aware method is beneficial for person search because it can provide a more appropriate ranking list. To achieve this, the discrete probability distribution estimated by the proposed method is used to construct the ranking list, which is different from the uncertainty-unaware baseline. Let $\hat{\boldsymbol{p}}_i = [\hat{p}_{i,0}, \ldots, \hat{p}_{i,K-1}]^T \in \mathbb{R}^K$ be the estimated age label distribution for the i-th person in the database. We then select the probability (i.e., a type of likelihood or similarity) for the query age y^q as $s_i = \hat{p}_{i,y^q}$ and subsequently create a ranking list by sorting the probability s_i in descending order.

2) Person Search by Age Group Query: In this subsection, we describe a method for another type of query, that is, age group query, whose age range is defined as $[y^q_{\min}, y^q_{\max})$ $(y^q_{\min}, y^q_{\max} \in \{0, 1, \ldots, M\}, y^q_{\min} < y^q_{\max})$.

Uncertainty-unaware baseline: Similar to the age query case, first, we describe an uncertainty-unaware baseline. A straightforward yet reasonable approach is to first consider a representative age $y^q_{\mathrm{rep}} = (y^q_{\min} + y^q_{\max} - 1)/2$ of the age group and then compute the dissimilarity to the representative age for each person as $d_i = |\hat{y}_i - y^q_{\mathrm{rep}}|$. Finally, we create a ranking list by sorting the dissimilarity d_i in ascending order.

Proposed uncertainty-aware method: Because the age group query contains multiple age labels and any of the multiple ages is considered as a true match, we compute the probability of a sum event (i.e., the multiple ages) based on the estimated age label distribution rather than using the probability of the representative age y^q_{rep}. Specifically, we define the probability of the i-th person as a summation over the age group:

$$s_i = \sum_{y=y^q_{\min}}^{y^q_{\max}-1} \hat{p}_{i,y}. \tag{9}$$

Finally, we create a ranking list by sorting the probability s_i in descending order.

3) Performance Measures for Person Search: We introduce a standard performance measure for person search (e.g., precision-recall curve) in this subsection to make this paper self-contained.

First, we introduce a true/false match indicator for the query. In the case of an age query, the indicator l_i for the i-th person is defined as

$$l_i = \begin{cases} 1, & y_i = y^q \text{ (true match)} \\ 0, & \text{otherwise (false match)}, \end{cases} \tag{10}$$

where y_i is the ground-truth age of the i-th person. In case of the age group query, it is defined as

$$l_i = \begin{cases} 1, & y^q_{\min} \le y_i < y^q_{\max} \text{ (true match)} \\ 0, & \text{otherwise (false match)}. \end{cases} \tag{11}$$

Next, the precision at rank $k \in \{1, \ldots, N\}$ is defined as the ratio of the number of true positive samples to the total number of true positive and false positive samples in the top-k candidates based on the ranking list. With the true/false match indicator l_i, the precision at k can be computed as

$$\mathrm{precision}(k) = \frac{\sum_{j=1}^{k} l_j}{k}. \tag{12}$$

The recall at rank k is defined as the ratio of the number of true positive samples to the total number of true positive and false negative samples, which can be converted using the indicator to

$$\mathrm{recall}(k) = \frac{\sum_{j=1}^{k} l_j}{\sum_{j=1}^{N} l_j}. \tag{13}$$

A precision-recall curve is then introduced, which is drawn using pairs of precision and recall over ranks. We further introduce the average precision (AP) as a single numerical criterion

to summarize the curve, which is an approximation of the area under the curve for the precision-recall curve, and which is defined as

$$\mathrm{AP} = \sum_{k=1}^{N} \mathrm{precision}(k) \times \Delta\mathrm{recall}(k), \tag{14}$$

where $\Delta\mathrm{recall}(k) = \mathrm{recall}(k) - \mathrm{recall}(k-1)$.

Finally, the overall performance of all queries is measured using the mean AP (mAP), which is computed as the average AP of all query ages/age groups.

B. People Counting by Age

1) People Counting by Age Group: In this subsection, we describe people counting by age group, which is beneficial for several applications, such as marketing research. For this purpose, we prepare a histogram of the age groups and then cast a vote for each detected person to the corresponding age group bin that includes his/her estimated age.

Specifically, we define the j-th age group bin as $[y_{j,\min}, y_{j,\max})$ $(y_{j,\min}, y_{j,\max} \in \{0, 1, \ldots, M\}$, $y_{j,\min} < y_{j,\max}, j = 1, \ldots, G)$, where M is the maximum age and G is the number of age group bins. Then, once the vote of the i-th detected person $(i = 1, \ldots, N)$ for the j-th age group bin $v_{i,j}$ is computed based on his/her age estimation result, the histogram of the age group is simply computed by summation as

$$h_j = \sum_{i=1}^{N} v_{i,j}, \tag{15}$$

where h_j is the histogram for the j-th age group bin. We can also normalize the histogram among all age group bins with the total people count N as $\bar{h}_j = h_j/N$, where $\bar{h}_j$ is the proportion of the people count of the j-th age group bin to the total people count. The key difference in the people count by age group is how to set the vote $v_{i,j}$; hence, we describe it in the following paragraphs.

Uncertainty-unaware baseline: A straightforward yet reasonable baseline involves setting a binary vote to an age group bin that includes the estimated age. Formally, the vote $v_{i,j}$ of the i-th person for the j-th age group bin is

$$v_{i,j} = \begin{cases} 1, & y_{j,\min} \le \hat{y}_i < y_{j,\max} \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

Proposed uncertainty-aware method: Different from the uncertainty-unaware baseline, which suffers from biases because of the severe underestimation of elderly people, even a single person can vote for multiple age group bins as a result of the proposed label distribution learning-based method. To achieve this, we set a weight for each age group bin by taking the uncertainty of age estimation into consideration, which mitigates such a bias problem. The vote $v_{i,j}$ of the i-th person for the j-th age group bin is formally defined using the probability of the sum event of multiple ages included in the age group bin as

$$v_{i,j} = \sum_{y=y_{j,\min}}^{y_{j,\max}-1} \hat{p}_{i,y}. \tag{17}$$

2) Performance Measure for People Counting by Age Group: To evaluate the performance of people counting, we compare the estimated normalized histogram of all age group bins $\{\bar{h}_j\}$ with that of the ground truth $\{\bar{h}_j^{\mathrm{gt}}\}$. We adopt intersection over union (IoU) to measure the differences between the estimated and ground-truth normalized histograms:

$$\mathrm{IoU} = \frac{\sum_{j=1}^{G} \min\left(\bar{h}_j, \bar{h}_j^{\mathrm{gt}}\right)}{\sum_{j=1}^{G} \max\left(\bar{h}_j, \bar{h}_j^{\mathrm{gt}}\right)}, \tag{18}$$

where $\bar{h}_j^{\mathrm{gt}}$ is the ground-truth ratio of the people count of the j-th age group bin.

VI. Experiments

A. Dataset

We used the *OULP-Age* [28] to evaluate the performance of the proposed method. *OULP-Age* is the world's largest gait database and includes gait images, in addition to the ground truth of age and gender. It consists of 63,846 gait images (31,093 males and 32,753 females) with an age range of 2 to 90 years old. GEIs and silhouette sequences of 88×128 pixels extracted for a side-view gait are provided for each subject. Following the original setting in [35], we further resized the silhouette sequences to 64×64 as the input for the GaitSet backbone network. Based on the predefined protocol, we divided the database into a training set composed of 31,923 subjects (15,596 males and 16,327 females) and a test set composed of 31,923 subjects (15,407 males and 16,426 females).

B. Training

We trained the network to minimize the joint loss function (Eq. (8)) using adaptive moment estimation (Adam) [54] with a batch size of 128 and 100 epochs. We set the initial learning rate to 0.001 for the GEINet backbone network,[†] and 0.0001 for the GaitSet backbone network. We set the number of age labels to $K = 101$, that is, the minimum and maximum ages were 0 and 100 years old, respectively. We set the hyperparameter σ for label distribution to 1.0, and both balancing parameters λ_{KL} and λ_{MAE} of the joint loss function to the same value: 1.0. We set the number of input frames for the GaitSet backbone network to 30 in both the training and test phases.

C. Evaluation Measure

We evaluated the accuracy of gait-based age estimation using two criteria. One is the MAE between the estimated age (i.e., an expectation of the age label distribution) and the ground-truth age, which is computed similarly to the MAE-based loss function (Eq. (6)).

[†]The conference version of this paper [33] used TensorFlow, but we re-implemented GEINet using PyTorch for an entirely fair comparison between GEINet and GaitSet for the same framework.

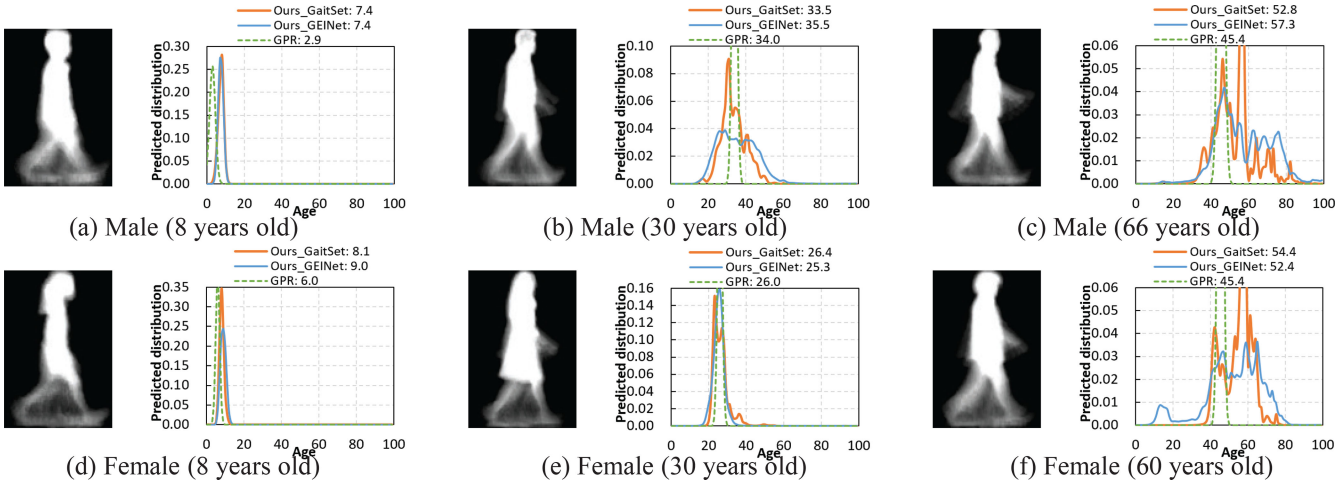


Fig. 5. Pairs of GEIs and estimated age distributions using the GPR-based method [12] (depicted in green), and the proposed method using the GEINet (depicted in cyan) and GaitSet (depicted in orange) backbone networks for males/females and children/adults/the elderly. The digits shown after each method are the corresponding estimated ages. In each caption, the ground-truth gender and age (in parentheses) are provided. The GPR-based method returned similar ranges for the uncertainty (variance), regardless of age. Conversely, the proposed method using both GEINet and GaitSet successfully returned small and large uncertainties for children and adults/the elderly, which coincides with the intuition regarding the gait-based age estimation accuracy addressed in Section I, and also with the scatter plots in Fig. 6. Best viewed in color.

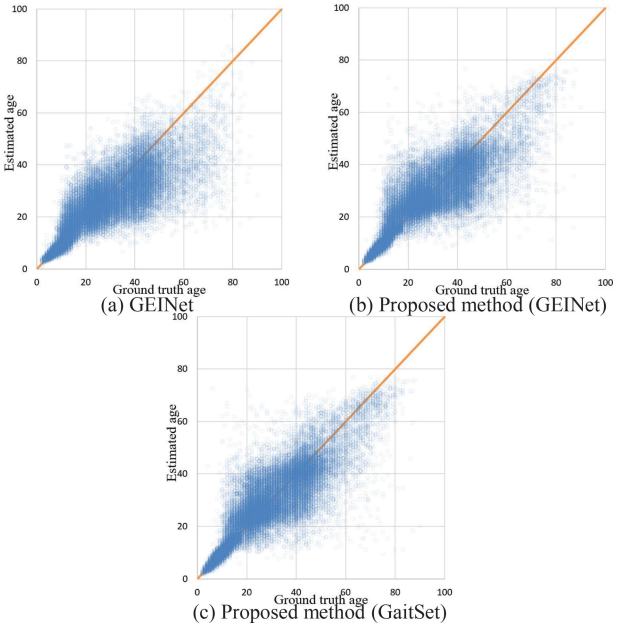


Fig. 6. Scatter plots between the ground-truth age and estimated age for GEINet [34], and the proposed method using both GEINet and GaitSet. Diagonal lines indicate equal lines of ground-truth age and estimated age.

The other is the CS, which is the error tolerance ratio. Specifically, we define the number of test samples whose absolute error between the estimated age and the ground-truth age is less than or equal to y as $n(y)$, and then the CS of the y-year absolute error as

$$\mathrm{CS}(y) = \frac{n(y)}{N}. \tag{19}$$

We evaluated the performance of person search by age using the AP for each query age/age group (Eq. (14)) and mAP over all queries, and evaluated the accuracy of people counting by age using the IoU between the estimated and ground-truth people statistics (Eq. (18)).

D. Qualitative Evaluation of the Label Distribution

First, as the most important aspect of the analysis, we show and compare the age distributions using the GPR-based method [12], and the proposed method using both the GEINet and GaitSet backbone networks in Fig. 5 to verify that the age estimation uncertainty is well represented.

Regarding the GPR-based method, we can clearly see that the uncertainty (i.e., the variance of the Gaussian distribution, depicted as green curves in Fig. 5) does not change much among children, adults, and the elderly, despite the fact that they should change to reflect the differences between ages in the uncertainty, as described and shown in Fig. 1 (i.e., a small uncertainty for children, and a large uncertainty for adults and the elderly). This is because the uncertainty obtained using the GPR-based approach mainly depends only on the closeness of an input gait feature to any of the training gait features, and hence cannot appropriately handle similar gait features with different ages, as we discussed in Section III.

Conversely, the proposed label distribution learning-based approach can successfully cope with age-dependent uncertainty. Specifically, the proposed method returns a sharp distribution (i.e., small uncertainty) for children (Fig. 5 (left)), and spread distributions (i.e., large uncertainty) for adults (Fig. 5 (middle)) and the elderly (Fig. 5 (right)), which is consistent with a common insight on gait-based age estimation, in addition to the scatter plots between the ground-truth and estimated ages (see Fig. 6). Additionally, we can see that the ground-truth age for each subject is covered by relatively high probabilities in the estimated age distribution, and in most cases, the estimated probability for the ground-truth age using GaitSet is higher than that using GEINet, which is consistent with the statistical results in Section VI-E.

TABLE I
MAEs [YEARS] AND CSs [%] AT 1-, 5-, AND 10-YEAR ABSOLUTE ERRORS. BOLD AND ITALIC BOLD INDICATE THE BEST AND SECOND-BEST PERFORMANCES, RESPECTIVELY. "-" INDICATES NOT PROVIDED IN THE ORIGINAL PAPERS. THE LAST SIX METHODS ARE DEEP LEARNING-BASED APPROACHES

Method	MAE	CS(1)	CS(5)	CS(10)
MLG [15]	10.98	16.7	43.4	60.8
GPR ($k = 10$) [12]	8.83	9.1	38.5	64.7
GPR ($k = 100$) [12]	7.94	10.5	43.3	70.2
GPR ($k = 1000$) [12]	7.30	10.7	46.3	74.2
SVR (Liner)	8.73	7.9	38.2	67.6
SVR (Gaussian)	7.66	9.4	44.2	73.4
OPLDA [14]	8.45	7.7	37.9	67.6
OPMFA [14]	9.08	7.0	34.9	64.1
ADGMLR [23]	6.78	18.4	54.0	76.2
DenseNet [21]	5.79	22.5	55.9	80.4
Multi-task [16]	5.47	-	-	-
Multi-stage [38]	5.48	***25.3***	***62.6***	82.0
GEINet [34]	6.22	17.2	55.9	79.2
Ours (GEINet)	***5.43***	23.5	61.7	***82.5***
Ours (GaitSet)	**5.01**	**25.9**	**65.1**	**84.4**

Consequently, we confirmed that the proposed method is suitable for uncertainty-aware gait-based age estimation.

Moreover, we note that the subject in Fig. 5(e) has a smaller uncertainty than the subject in Fig. 5(b), but they are both 30 years old. This may be because the outfit of subject (e), which is, generally speaking, a type of generation-specific fashion style, could be a cue to narrow the possible age range. As such, we can see that the proposed method is potentially capable of handling not only age-dependent uncertainty but also sample-dependent uncertainty.

E. Comparison With State-of-the-Art Methods

We compared the proposed method with the benchmarks of existing gait-based age estimation. As a baseline algorithm, we used a method using GPR with the RBF kernel [12] and an active set method in the same manner shown in [55], where the k nearest neighbors for each test sample were used for GPR and $k = 10, 100, 1000$ were evaluated. We also evaluated existing gait-based age estimation methods using conventional machine learning techniques: SVR with linear and Gaussian kernels, denoted by SVR (linear) and SVR (Gaussian), MLG [15], OPLDA [14], OPMFA [14], and AGDMLR [23]. Additionally, we used a slightly modified version of GEINet [34] as a deep learning-based approach to age estimation. Specifically, the original GEINet outputs class (i.e., subject) likelihoods; hence, the number of nodes at the last layer is equal to the number of subjects. Conversely, the modified version of GEINet outputs an age; hence, the last layer has only a single node. Moreover, we also evaluated other state-of-the-art deep learning approaches [16], [21], [38] to gait-based age estimation.

The MAEs and CSs for 1-, 5-, and 10-year tolerances are summarized in Table I. The results showed that the deep learning-based methods (i.e., GEINet [34], DenseNet [21], multi-task [16], multi-stage [38], and the proposed method) significantly outperformed the other conventional machine learning-based methods. Additionally, the proposed method also outperformed the state-of-the-art deep learning-based approaches [16], [21], [34], [38]. Compared with a relatively simple backbone network (i.e., GEINet), the proposed method worked better with a more state-of-the-art backbone (i.e., GaitSet). Conversely, the proposed method clearly outperformed a regression-based method under the same backbone network, that is, GEINet [34] by a large margin (e.g., the proposed method improved the MAE and CS(1) by 0.79 years and 6.3%, respectively). This is because the proposed method with the age label distribution has a more powerful and flexible expression capability than the regression-based method, which outputs a single age value. Specifically, the regression-based model (e.g., GEINet [34]) is highly affected by outlier subjects who look much younger/older than their age, whereas the proposed method can mitigate the effect of outliers by assigning probabilities to multiple age labels.

TABLE II
MAEs [YEARS] FOR THE SENSITIVITY ANALYSIS OF THE STANDARD DEVIATION σ OF THE GROUND-TRUTH LABEL DISTRIBUTION

σ	0.1	0.2	0.5	1.0	2.0	3.0	5.0	10.0
GEINet	5.49	5.45	**5.40**	***5.43***	5.47	5.51	5.58	5.64
GaitSet	5.23	5.13	***5.03***	**5.01**	5.07	5.11	5.19	5.25

As further analysis, Fig. 6 shows scatter plots between the ground-truth age and the estimated ages for deep learning-based methods with the same backbone network, that is, GEINet [34] as a regression-based approach, and the proposed method using both GEINet and GaitSet as label distribution-based approaches. In all cases, we can observe a common property in the gait-based age estimation, that is, the age estimation uncertainty is small for children, whereas it is large for adults and the elderly. We took a closer look at the differences among them. The estimated ages for GEINet [Fig. 6(a)] deviated more than those for the proposed method [Figs. 6(b) and (c)] for all age ranges, particularly children under 15 years old. Additionally, the proposed method significantly improved the estimation accuracy for subjects over 60 years old compared with GEINet. Specifically, most subjects over 60 years old were underestimated (i.e., biased toward the younger direction) for GEINet. The proposed method successfully mitigated the underestimate, where the GaitSet backbone network [Fig. 6(c)] improved more than the GEINet backbone network [Fig. 6(b)]. This led to the performance improvement of quantitative criteria such as the MAE and CS.

F. Sensitivity Analysis

Because the standard deviation σ of the label distribution is a key hyperparameter for the proposed method, we analyze its sensitivity to gait-based age estimation accuracy.

Table II shows the MAEs when changing the standard deviation σ. According to Table II, the MAE became worse when σ is greater than 1.0 or less than 0.5. This is partly because the ground-truth age in *OULP-Age* was given in the integer domain. In fact, the age annotation of *OULP-Age* was provided by each subject's self-declaration in a long-run exhibition of video-based gait analysis [55]; hence, unless the subject provided incorrect information, the rounding error for integer age annotation (e.g., both a just 10 years-old subject and a 10

TABLE III
MAEs [YEARS] AND MCE FOR DIFFERENT SETTINGS OF BALANCING PARAMETERS λ_{KL} AND λ_{MAE}

Balancing parameter		Backbone			
λ_{KL}	λ_{MAE}	GEINet		GaitSet	
		MAE	MCE	MAE	MCE
1	0	5.78	-3.09	5.35	-3.06
0	1	5.82	-8.60	5.29	-6.36
1	1	***5.43***	**-3.02**	**5.01**	**-3.00**
1	2	***5.43***	***-3.06***	***5.02***	***-3.08***
1	5	**5.42**	-3.17	**5.01**	-3.16
1	10	5.45	-3.41	5.09	-3.40
1	100	5.47	-5.38	5.16	-4.35
1	1000	5.51	-7.12	5.22	-6.09

years + 11 months + 30 days-old subject input 10 years-old.) is ideally less than 1 year, and the mean absolute error of annotation is around 0.5 years. Therefore, it is reasonable to set σ in the range of 0.5 to 1.0. Furthermore, the results of $\sigma = 0.5$ and $\sigma = 1.0$ are similar to each other for both the GEINet and GaitSet backbones. Additionally, the results of the GaitSet backbone network are better than those of the GEINet backbone network for all σ values, which is consistent with the results in Section VI-E. Moreover, we confirmed that the tendency of the sensitivity analysis of the GaitSet backbone is similar to that of the GEINet.

G. Effects of Loss Functions

Moreover, we evaluated the effects of the two loss functions, that is, the KL divergence-based loss function L_{KL} and MAE-based loss function L_{MAE} with both the GEINet and GaitSet backbone networks.

Table III shows the results for different settings of balancing parameters. The results show that the MAE became worse when we made either of the loss functions invalid (i.e., set the balancing parameter to 0); that is, both the KL divergence-based and MAE-based loss play important roles in a complementary manner to each other. Moreover, the MAEs were quite similar for $\lambda_{\mathrm{MAE}} = 1, 2, 5$, and became worse when λ_{MAE} is larger than 10. This is because the balance between the MAE-based loss and KL divergence-based loss gets worse if λ_{MAE} is too large, which results in the accuracy of the estimated age label distribution being too much sacrificed, and further leads to the worse performance of age estimation.

We further validated the above-mentioned point by evaluating the performance not only of mean (or expectation of) age with MAE but also of age label distribution estimation itself (i.e., including uncertainty). For this purpose, we introduced the mean cross-entropy (MCE) between the estimated and ground-truth distribution as a criterion. Following [56], we defined the ground-truth label distribution as a delta function of the ground-truth age; hence, the MCE is computed as [56]

$$\mathrm{MCE} = \frac{1}{N}\sum_{i=1}^{N} \log \hat{p}_{i,y_i}, \tag{20}$$

where N is the number of test samples, and $\hat{p}_{i,y_i}$ is the estimated probability of the ground-truth age label y_i for the i-th sample. The MCE will be greater if the estimated probability of the ground-truth age is greater.

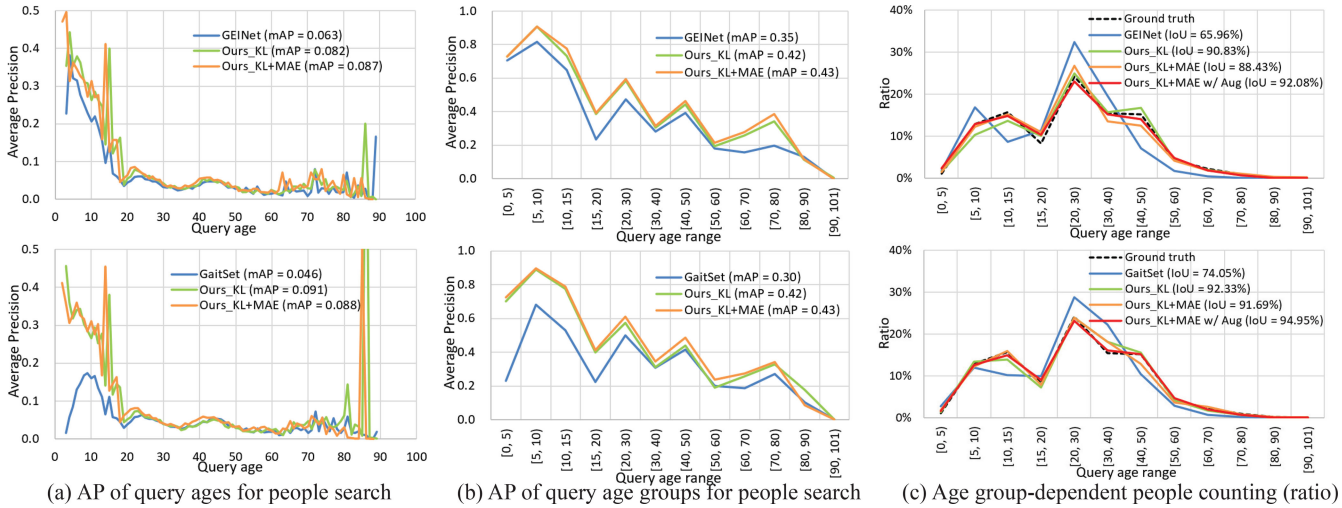
The results in Table III show that the MCE became worse with the increase of λ_{MAE}, which means the decrease of the accuracy of the estimated age label distribution. When $\lambda_{\mathrm{MAE}} = 1000$, both MAE and MCE got closer to the results of using only the MAE-based loss (i.e., the second row). As a result, the setting of $\lambda_{\mathrm{KL}} = \lambda_{\mathrm{MAE}} = 1$ yielded the best overall performance of age estimation and age label distribution estimation. Additionally, the GaitSet backbone network achieved better results than the GEINet backbone network in almost all cases, which shows the effectiveness of a powerful backbone network for age estimation.

H. Simulation Experiments for Person Search and People Counting by Age

We conducted simulation experiments for person search and people counting by age on *OULP-Age*. As introduced in Section V, we investigated two application scenarios for person search, that is, person search by age query [Fig. 7(a)] and person search by age group query [Fig. 7(b)], whereas we considered only the age group for people counting [Fig. 7(c)]. Similar to [57], we set the query age groups to [0, 5), [5, 10), [10, 15), [15, 20), [20, 30), [30, 40), . . . , [80, 90), [90, 101), where the age interval was set to 5 years for children and teenagers, and 10 years for adults, considering the changes in the growth rate of a human body. We used age estimation results from both the GEINet and GaitSet backbone networks for evaluation. For comparison, we also evaluated the corresponding regression-based models, that is, GEINet [34] and GaitSet [35], as baseline methods. Additionally, we did not use the expected ages (Eq. (7)) in the two applications (i.e., no MAE-related criteria were included); hence, we also report the results of models trained with only the KL divergence-based loss function L_{KL} for comparison.

Regarding person search, the proposed uncertainty-aware methods clearly outperformed the regression-based methods, as shown in Fig. 7. Specifically, for the person search by age query, the APs of the regression-based methods were worse than those of the uncertainty-aware methods for query ages of less than 20 years old, whereas the performances were similar to each other for those over 20 years old. Conversely, the results of the person search by age group query with the uncertainty-aware methods were always better than the regression-based methods for all query age groups. This is because the uncertainty of age estimation was larger for adults compared with children; hence, the exact age query (i.e., uncertainty of ± 1 years) for adults was quite difficult for both the regression-based and the uncertainty-aware methods. Conversely, the age group query considered a reasonable uncertainty for children and adults, which resulted in the superior performance of the uncertainty-aware methods.

Regarding people counting by age group, we evaluated the normalized histogram of the age groups (i.e., people statistics), as shown in Fig. 7(c). As a result, the regression-based method tended to assign more votes to young and middle-age people (i.e., 20–40 years old) and fewer votes to the



(a) AP of query ages for people search (b) AP of query age groups for people search (c) Age group-dependent people counting (ratio)

Fig. 7. Results of person search and people counting using the estimated age using regression-based methods (i.e., GEINet and GaitSet), and the estimated label distribution using the proposed method trained with only the KL divergence loss, and both the KL divergence and MAE losses. Additionally, the results of people counting using the model trained with the augmented training set are also shown in (c) (depicted in red). The results shown in the top row are based on the GEINet backbone network, whereas those in the bottom row are based on the GaitSet backbone network.

elderly compared with the ground truth. This is because the ages estimated by the regression-based method were biased toward the population's mean to avoid large errors in case of misestimation, which was also seen in the scatter plots [Fig. 6(a)]. Conversely, the proposed uncertainty-aware methods were essentially similar to the ground-truth people statistics because they had more opportunities to cast a vote to the elder's bin because of the label distribution representation.

The estimation accuracies were different between models trained with and without the MAE loss, as shown in Section VI-G, whereas the performances of person search and people counting with the model trained only with KL divergence loss were quite similar to those with both KL divergence and MAE losses. This is understandable because the age estimation accuracy is evaluated using the MAE; hence, the inclusion of the MAE loss naturally yielded better results. Conversely, the performance of uncertainty-aware person search and people counting was never computed using an MAE-related metric (e.g., absolute difference from the query or age group bin's representative age) but was computed using the probability distribution instead; hence, the inclusion of the MAE loss did not necessarily result in better performance. Additionally, the GaitSet backbone network obtained slightly better results than the GEINet backbone network in most cases, which is essentially consistent with the results of age estimation in Section VI-G.

I. Discussion

1) Analysis of Person Search by Age Query: Because of the large uncertainty of age estimation for adults, which largely increases the difficulty of person search by age query, the proposed method yielded relatively low APs for query ages in the adult group, as shown in Fig. 7(a). In order to take a closer look, we further discuss the results of precision and recall by taking the real application scenarios into consideration.

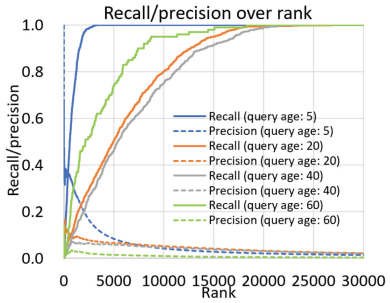
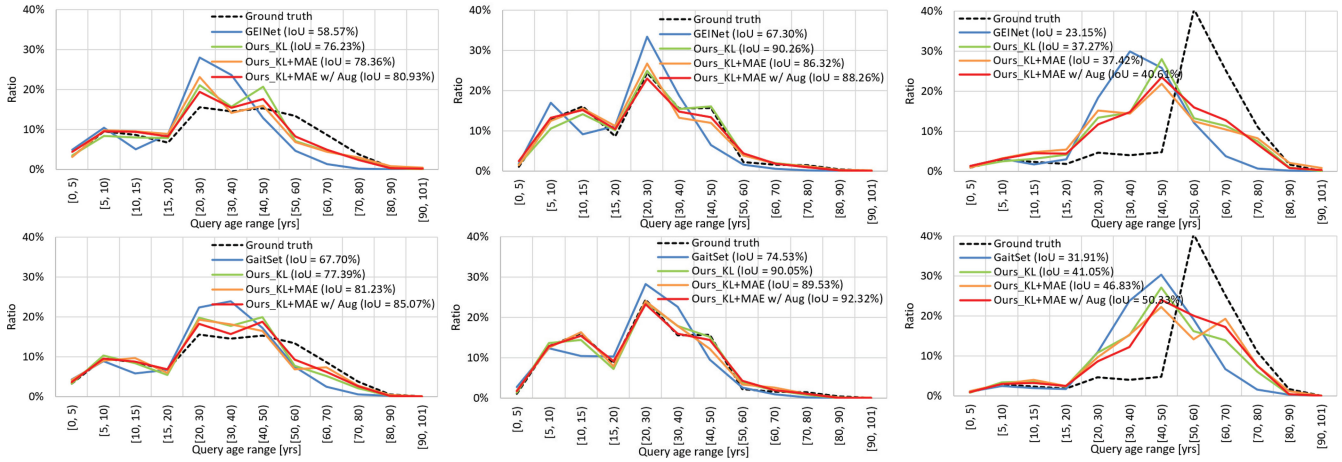


Fig. 8. Precision and recall over ranks for the query age of 5, 20, 40, and 60 years old obtained by the proposed method using the GaitSet backbone with both the KL divergence and MAE losses.

To check the precision and recall at each rank, we chose four typical query ages, i.e., 5, 20, 40, 60 years old to compare the performance. Figure 8 shows the results of precision and recall over ranks for the proposed method using the GaitSet backbone with both the KL divergence and MAE losses. According to the results, a list of approximately 3,000, 25,000, 23,000, and 18,000 candidates was created for the query age of 5, 20, 40, and 60 years old, respectively, when all persons whose ground-truth ages matches the query age (i.e., true match) were included in the list (i.e., the rank for 100% recall). Consider the application to search for a suspect with his/her exact age, because the total database for searching contains 31,923 persons, the proposed method eliminated about 20% of candidates even in the worst case (i.e., query age of 20 years old), which still reduced considerable time efforts for criminal investigators. If we consider a slightly lower recall (e.g., 80% recall), the search work on 31,923 persons would be greatly reduced to 1,500, 10,000, 12,000, 6,000 persons for query age of 5, 20, 40, and 60 years old, respectively.

Another possible application scenario is searching for a known person (e.g., a family member such as a lost child or a wandering grandparent), where the query age may be more biased towards children and the elderly groups. In that case,



(a) Uniform distribution (b) Younger-biased (c) Elderly-biased

Fig. 9. Age group-dependent people counting (ratio) with different test distributions. Each column shows the results of each test subset. We compared the results of regression-based methods (i.e., GEINet [34] and GaitSet [35]), and the proposed method using the same backbone networks that were trained with only the KL divergence loss, and both the KL divergence and MAE losses, as well as using the model trained with the augmented training set. The results shown in the top row are based on the GEINet backbone network, whereas those in the bottom row are based on the GaitSet backbone network.

the recall and precision are relatively better compared with the adults, which results in much fewer search efforts.

Besides, in this work, we used the size-normalized silhouettes as input data, which may discard a couple of useful information containing age-related cues. To improve the accuracy of person search as well as the performance of age estimation, a possible way is to include normalization factors such as the height and walking speed (e.g., heights of children are smaller, and walking speeds of the elderly are slower). Additionally, rather than using the appearance-based representations that entangle the body shape and motion factors, we may also improve the performance by disentangling them via disentangled representation learning [58] or human model fitting [59], which helps to maximize the use of each body shape and motion factor for better generalization capability.

2) People Counting Over Various Statistics: We then analyzed the effects of the statistical distribution on the performance of people counting. For this purpose, we trained the proposed method still using the original training set, whereas the following three test subsets with different subject distributions were prepared for evaluation: a subset of 8,012 subjects with uniform age group distribution as much as possible, a younger-biased subset of 7,833 subjects, and an elderly-biased subset of 2,756 subjects. Specifically, because of the limited number of subjects over 50 years old, we set up the first test subset by randomly choosing 300 subjects for each gender and age group in 5-year intervals under 50 years old while keeping all the subjects over 50 years old in the entire test set to maintain a uniform distribution as much as possible. Regarding the second subset that contained more young subjects, we randomly selected a quarter of subjects from the entire test set for each gender and age group in 5-year intervals under 50 years old, and randomly selected 30 subjects for each gender and age group in 5-year intervals over 50 years old.[‡] To ensure that the third subset contained more elderly subjects, we randomly selected 30 subjects for each gender and age group in 5-year intervals under 50 years old, whereas we used all subjects over 50 years old in the entire test set.

The results are shown in Fig. 9. Generally, the accuracy of people statistics estimation is highly dependent on the distribution similarity between the training set and test subset; hence, the younger-biased subset achieved the best performance among the three subsets, whereas the elderly-biased subset performed relatively worse than the others. Similar to the results in Section VI-H, the proposed uncertainty-aware methods yielded much better performances than the regression-based methods for all three subsets, where the GaitSet backbone network also yielded somewhat better performance than the GEINet backbone network.

Although the estimated people statistics of the uniform subset still had differences from the ground truth, the difference in counting ratios among all age ranges was relatively smaller compared with the original test set [see Fig. 7(c)], which shows a trend of uniform distribution, to some extent. Because more young subjects were contained in the younger-biased subset, whose subject distribution was essentially similar to that of the training set and the entire test set, its estimation accuracy was quite close to that of the entire test set. Conversely, more subjects were counted as middle-aged for the elderly-biased subset, which had a relatively large estimation error from the ground-truth. This was caused by the significant difference between the age distribution of the training and test sets.

Therefore, the prior distribution of training data is quite important for the performance of people counting with different statistics. To improve the accuracy of the elderly-biased subset, one possible solution is to train the proposed method using a training set with a uniform age distribution as much as possible. However, because of the bias toward young subjects in *OULP-Age*, an extremely uniform training distribution may result in much fewer training samples compared with the original training set. Conversely, to maintain a large amount of

[‡]All subjects over 80 years old were used because of a lack of data (i.e., fewer than 30 subjects for each gender and age group).

training data, a considerable bias may still exist in the training distribution in terms of age. Consequently, it is necessary to consider a trade-off between a sufficient number of training samples and an appropriately uniform age distribution when constructing the training set.

3) Balancing Training Data via Augmentation: As mentioned in Section VI-I2, the non-uniform training distribution of *OULP-Age* causes the model to be biased towards young subjects. Therefore, we balanced the training set of each age group through conventional data augmentation methods to compare the performance. More specifically, we applied random rotation between -8° and 8° in ±2° intervals, rescaling between 90% and 110% in 5% intervals, and translation between −5 and 5 pixels in ±2-pixel intervals to generate the augmented samples. Additionally, we made the most of the subjects in the original training set to maximize the subject diversity after augmentation. Finally, we prepared a training set containing 90,000 samples, where 2,500 samples are included in each gender and age group in 5-year intervals.

Using the proposed method trained by the augmented training set, the MAE of the model using the GEINet backbone was reduced from 5.43 year to 5.41 year, while that using the GaitSet backbone was reduced from 5.01 year to 4.91 year, which shows the slight performance improvement of age estimation. We also showed the corresponding results of people counting on the original test set in Fig. 7(c) (depicted in red). It is obvious that the estimated people statistics became quite closer to the ground-truth, which obtained approximately 95% IOU with the model using the GaitSet backbone.

We again evaluated the performance of people counting on three test subsets introduced in Section VI-I2, and the results are shown in Fig. 9 (depicted in red). Similarly to the original test set, the accuracies of estimated people statistics were all improved on three test subsets, and the improvements on the uniform subset and elderly-biased subset were relatively larger than that on the younger-biased subset. This is understandable because the augmented training set increased much more children and elderly samples compared with the young people, which is more favorable for the uniform and elderly-biased subsets. On the other hand, although the accuracy on the elderly-biased subset was improved to some extent, there is still a large margin between the estimation results and the ground-truth. This is due to too few elderly subjects in the original dataset (e.g., only 21 subjects in the 81 to 90 age group), which greatly limits the subject diversity even after data augmentation, and hence, easily leads to overfitting of the trained model. Therefore, it is important to capture sufficient subjects to maintain the balance between generations in the dataset.

4) Considerations for Real-World Applications: Although *OULP-Age* is the world's largest gait database with age annotations, the gait videos were captured under well-controlled indoor conditions, which makes the silhouette extraction much easier than those obtained in real-world scenarios. Occlusions, illumination, and complex backgrounds, which are difficult variations often exist in real captured scenes, have great influences on the silhouette segmentation results, and may further affect the age estimation performance of the proposed method. Fortunately, recent state-of-the-art deep learning-based semantic segmentation methods, such as RefineNet [60] and Mask R-CNN [61], achieved significant improvement in human segmentation in complex scenes. In fact, some recent gait recognition involves semantic segmentation (e.g., [58]) and also involves silhouette extraction in an end-to-end gait recognition framework (e.g., [62]), which could be the support of future applications of gait recognition in the real world. Therefore, we can also consider combining the proposed method with these segmentation methods when applying to real applications.

VII. Conclusion

In this paper, we described an uncertainty-aware gait-based age estimation method. Unlike existing uncertainty-aware approaches, such as the GPR-based method [12], the proposed method can successfully and naturally handle similar gait features with different ages by introducing a label distribution learning framework. Experiments on the world's largest gait database *OULP-Age* showed that the proposed method can represent the age estimation uncertainty well and outperformed or was comparable with state-of-the-art methods. Additionally, experiments on two applications, that is, person search and people counting by age, showed the effectiveness of the estimated uncertainty in a quantitative manner.

One future research avenue is to make the estimated age distribution smoother. The probabilities between adjacent age labels sometimes change abruptly (Fig. 5), which is unreasonable in reality. We will therefore include frameworks to retain the ordinary properties of age labels (e.g., a smoothness loss or adaptive setting of the uncertainty of the ground-truth age distribution considering the degree of shortage of training samples for the age). Additionally, because this study focused only on side-view gait, another research direction is to make the age estimator robust against various covariates, such as view, carrying status, and walking speed.

Acknowledgment

The authors thank Maxine Garcia, Ph.D., from Edanz Group (https://en-author-services.edanzgroup.com/) for editing a draft of this manuscript.

References

[1] M. S. Nixon, T. N. Tan, and R. Chellappa, *Human Identification Based on Gait* (International Series on Biometrics). Heidelberg, Germany: Springer-Verlag, Dec. 2005.

[2] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon, "On using gait in forensic biometrics," *J. Forensic Sci.*, vol. 56, no. 4, pp. 882–889, 2011.

[3] S. Stevenage, M. Nixon, and K. Vince, "Visual analysis of gait as a cue to identity," *Appl. Cogn. Psychol.*, vol. 13, no. 6, pp. 513–526, Dec. 1999.

[4] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, Feb. 2005.

[5] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, "Gait recognition using a view transformation model in the frequency domain," in *Proc. 9th Eur. Conf. Comput. Vis.*, Graz, Austria, May 2006, pp. 151–163.

[6] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proc. 18th Int. Conf. Pattern Recognit.*, vol. 4. Hong Kong, Aug. 2006, pp. 441–444.

[7] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu, "A study on gait-based gender classification," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1905–1910, Aug. 2009.

[8] L. Lee and W. Grimson, "Gait analysis for recognition and classification," in *Proc. 5th IEEE Conf. Face Gesture Recognit.*, vol. 1, 2002, pp. 155–161.

[9] X. Li, S. Maybank, S. Yan, D. Tao, and D. Xu, "Gait components and their application to gender recognition," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 2, pp. 145–155, Mar. 2008.

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

[11] R. Begg, "Support vector machines for automated gait classification," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 5, pp. 828–838, May 2005.

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

[13] J. Lu and Y.-P. Tan, "Ordinary preserving manifold analysis for human age estimation," in *Proc. IEEE Comput. Soc. IEEE Biometr. Council Workshop Biometr.*, San Francisco, CA, USA, Jun. 2010, pp. 1–6.

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

[15] J. Lu and Y.-P. Tan, "Gait-based human age estimation," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 4, pp. 761–770, Dec. 2010.

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

[17] D. Zhang, Y. Wang, and B. Bhanu, "Ethnicity classification based on gait using multi-view fusion," in *Proc. IEEE Comput. Soc. IEEE Biometr. Council Workshop Biometr.*, San Francisco, CA, USA, Jun. 2010, pp. 1–6.

[18] M. Lemke, T. Wendorff, B. Mieth, K. Buhl, and M. Linnemann, "Spatiotemporal gait patterns during over ground locomotion in major depression compared with healthy controls," *J. Psychiatric Res.*, vol. 34, nos. 4–5, pp. 277–283, 2000.

[19] N. F. Troje, "Decomposing biological motion: A framework for analysis and synthesis of human gait patterns," *J. Vis.*, vol. 2, no. 2, pp. 371–387, 2002.

[20] A. Shehata, Y. Hayashi, Y. Makihara, D. Muramatsu, and Y. Yagi, "Does my gait look nice? human perception-based gait relative attributes estimation by dense trajectory analysis," in *Proc. 5th Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2019, pp. 1–14.

[21] A. Sakata, Y. Makihara, N. Takemura, D. Muramatsu, and Y. Yagi, "Gait-based age estimation using a DenseNet," in *Proc. Int. Workshop Attention Intention Understand. (AIU)*, Dec. 2018, pp. 55–63.

[22] H. Mannami, Y. Makihara, and Y. Yagi, "Gait analysis of gender and age using a large-scale multi-view gait database," in *Proc. 10th Asian Conf. Comput. Vis.*, Queenstown, New Zealand, Nov. 2010, pp. 975–986.

[23] 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. [Online]. Available: https://doi.org/10.1007/s11042-018-6049-7

[24] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A discriminant analysis for undersampled data," in *Proc. 5th Annu. ACM Workshop Comput. Learn. Theory*, 1992, pp. 144–152.

[25] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004, doi: 10.1023/B:STCO.0000035301.49549.88.

[26] M. J. Marìn-Jimènez, F. M. Castro, N. Guil, F. de la Torre, and R. Medina-Carnicer, "Deep multi-task learning for gait-based biometrics," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 106–110.

[27] 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. Biometr. Theory Appl. Syst. (BTAS)*, Sep. 2019, pp. 1–9.

[28] 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. [Online]. Available: https://doi.org/10.1186/s41074-017-0035-2

[29] 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.

[30] B. Gao, C. Xing, C. Xie, J. Wu, and X. Geng, "Deep label distribution learning with label ambiguity," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2825–2838, Jun. 2017.

[31] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, 1951.

[32] B.-B. Gao, H.-Y. Zhou, J. Wu, and X. Geng, "Age estimation using expectation of label distribution learning," in *Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI)*, Jul. 2018, pp. 712–718. [Online]. Available: https://doi.org/10.24963/ijcai.2018/99

[33] A. Sakata, Y. Makihara, N. Takemura, D. Muramatsu, and Y. Yagi, "How confident are you in your estimate of a human age? Uncertainty-aware gait-based age estimation by label distribution learning," in *Proc. 4th Int. Joint Conf. Biometr. (IJCB)*, 2020, pp. 1–10.

[34] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "GeiNet: View-invariant gait recognition using a convolutional neural network," in *Proc. 8th IAPR Int. Conf. Biometr. (ICB)*, Halmstad, Sweden, Jun. 2016, pp. 1–8.

[35] 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.

[36] Z. Liu and S. Sarkar, "Simplest representation yet for gait recognition: Averaged silhouette," in *Proc. 17th Int. Conf. Pattern Recognit.*, vol. 1, Aug. 2004, pp. 211–214.

[37] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[38] A. Sakata, N. Takemura, and Y. Yagi, "Gait-based age estimation using multi-stage convolutional neural network," *IPSJ Trans. Comput. Vis. Appl.*, vol. 11, no. 1, p. 4, 2019. [Online]. Available: https://doi.org/10.1186/s41074-019-0054-2

[39] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[40] X. Geng, C. Yin, and Z. H. Zhou, "Facial age estimation by learning from label distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2401–2412, Oct. 2013.

[41] W. Zhao and H. Wang, "Strategic decision-making learning from label distributions: An approach for facial age estimation," *Sensors*, vol. 16, no. 7, p. 994, Jun. 2016. doi: 10.3390/s16070994.

[42] Z. He *et al.*, "Data-dependent label distribution learning for age estimation," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3846–3858, Aug. 2017.

[43] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with *k*-reciprocal encoding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 3652–3661.

[44] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2285–2294.

[45] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Bian, and Y. Yang, "Progressive learning for person re-identification with one example," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2872–2881, Jun. 2019.

[46] Z. Zhong, L. Zheng, Z. Luo, S. Li, and Y. Yang, "Invariance matters: Exemplar memory for domain adaptive person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 598–607.

[47] A. Schumann, A. Specker, and J. Beyerer, "Attribute-based person retrieval and search in video sequences," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Surveillance (AVSS)*, 2018, pp. 1–6.

[48] M. Halstead, S. Denman, C. Fookes, Y. Tian, and M. S. Nixon, "Semantic person retrieval in surveillance using soft biometrics: AVSS 2018 challenge II," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Surveillance (AVSS)*, 2018, pp. 1–6.

[49] W. Chen and J. J. Corso, "Action detection by implicit intentional motion clustering," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 3298–3306.

[50] G. Gkioxari and J. Malik, "Finding action tubes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 759–768.

[51] S. Li, T. Xiao, H. Li, B. Zhou, D. Yue, and X. Wang, "Person search with natural language description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 5187–5196.

[52] M. Yamaguchi, K. Saito, Y. Ushiku, and T. Harada, "Spatio-temporal person retrieval via natural language queries," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 1462–1471.

[53] G. Ko, Y. Lee, N. Moon, N. Koki, Y.-S. Lee, and N. Moon, "People counting system by facial age group," *J. Inst. Electron. Eng. Korea*, vol. 51, no. 2, pp. 69–75, 2014.

[54] D. P. Kingma and J. Ba. (2014). *Adam: A Method for Stochastic Optimization*. [Online]. Available: http://arxiv.org/abs/1412.6980
[55] Y. Makihara *et al.*, "Gait collector: An automatic gait data collection system in conjunction with an experience-based long-run exhibition," in *Proc. Int. Conf. Biometr. (ICB)*, Jun. 2016, pp. 1–8.
[56] C. Xu *et al.*, "Real-time gait-based age estimation and gender classification from a single image," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 3460–3470.
[57] 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.
[58] Z. Zhang, L. Tran, X. Yin, Y. Atoum, and X. Liu, "Gait recognition via disentangled representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4710–4719.
[59] X. Li, Y. Makihara, C. Xu, Y. Yagi, S. Yu, and M. Ren, "End-to-end model-based gait recognition," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, Nov. 2020, pp. 3–20.
[60] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5168–5177.
[61] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988.
[62] C. Song, Y. Huang, Y. Huang, N. Jia, and L. Wang, "GaitNet: An end-to-end network for gait based human identification," *Pattern Recognit.*, vol. 96, Dec. 2019, Art. no. 106988.

Chi Xu received the Ph.D. degree in engineering from the Nanjing University of Science and Technology, China, in 2021. She worked as a Visiting Researcher in 2016, and a specially appointed Researcher (part-time) from 2017 to 2020 with the Institute of Scientific and Industrial Research, Osaka University, Japan, where she is currently a specially appointed Researcher (full-time). Her research interests are gait recognition, machine learning, and image processing.

Atsuya Sakata received the B.S. and M.S. degrees from Osaka University, Japan, in 2018 and 2020, respectively. His research interest is computer vision including video-based gait analysis.

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. He is currently a Professor with the Institute for Advanced Co-Creation Studies, Osaka University. His research interests are computer vision, pattern recognition, and image processing including gait recognition, pedestrian detection, morphing, and temporal super resolution. 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, the Prizes for Science and Technology, Research Category in 2014. He has served as an Associate Editor in Chief of *IEICE Transactions on Information and Systems*, an Associate Editor of *IPSJ Transactions on Computer Vision and Applications*, the Program Co-Chair of the 4th Asian Conference on Pattern Recognition (ACPR 2017), the Area Chairs of ICCV 2019, CVPR 2020, and ECCV 2020. He is a member of IPSJ, IEICE, RSJ, and JSME.

Noriko Takemura received the B.S., M.E., and Ph.D. degrees in engineering from Osaka University in 2006, 2007, and 2010, respectively, where she is currently an Associate Professor with the Institute for Datability Science. Her research interests include gait recognition, ambient intelligence, and emotion estimation. She is a member of the SICE, RSJ, and VRSJ.

Daigo Muramatsu (Member, IEEE) received the B.S., M.E., and Ph.D. degrees in engineering from Waseda University, Tokyo, Japan, in 1997, 1999, and 2006, respectively. He is currently a Professor with the Department of Computer and Information Science, Faculty of Science and Technology, Seikei University. His research interests are pattern recognition, and biometrics including gait recognition. He is a member of IPSJ and IEICE.

Yasushi Yagi (Senior Member, IEEE) received the Ph.D. degree from Osaka University in 1991, where he is a Professor with the Institute of Scientific and Industrial Research. In 1985, he joined the Product Development Laboratory, Mitsubishi Electric Corporation, where he worked on robotics and inspections. He became a Research Associate in 1990, a Lecturer in 1993, an Associate Professor in 1996, and a Professor in 2003 with Osaka University. He was also the Director of the Institute of Scientific and Industrial Research, Osaka University from 2012 to 2015, and the Executive Vice President of Osaka University from 2015 to 2019. His research interests are computer vision, medical engineering, and robotics. He was awarded the ACM VRST2003 Honorable Mention Award, the IEEE ROBIO2006 Finalist of T. J. Tan Best Paper in Robotics, the IEEE ICRA2008 Finalist for Best Vision Paper, the MIRU2008 Nagao Award, and the PSIVT2010 Best Paper Award. The international conferences for which he has served as the Chair includes FG1998 (Financial Chair), OMINVIS2003 (Organizing Chair), ROBIO2006 (Program Co-Chair), ACCV2007 (Program Chair), PSVIT2009 (Financial Chair), ICRA2009 (Technical Visit Chair), ACCV2009 (General Chair), ACPR2011 (Program Co-Chair), and ACPR2013 (General Chair). He has also served as the Editor of IEEE ICRA Conference Editorial Board from 2007 to 2011. He is the Editorial Member of *International Journal of Computer Vision* and the Editor-in-Chief of *IPSJ Transactions on Computer Vision and Applications*. He is a Fellow of IPSJ and a member of IEICE and RSJ.

Jianfeng Lu received the B.S. degree in computer software and the M.S. and Ph.D. degrees in pattern recognition and intelligent systems from the Nanjing University of Science and Technology, China, where he is currently a Professor. His research interests include image processing, pattern recognition, and data mining.