

## PAPER

## Health Indicator Estimation by Video-Based Gait Analysis

Ruochen LIAO<sup>†a)</sup>, Kousuke MORIWAKI<sup>†</sup>, *Nonmembers*, Yasushi MAKIHARA<sup>†</sup>,  
Daigo MURAMATSU<sup>†,††</sup>, *Members*, Noriko TAKEMURA<sup>†</sup>, *Nonmember*, and Yasushi YAGI<sup>†</sup>, *Member*

**SUMMARY** In this study, we propose a method to estimate body composition-related health indicators (e.g., ratio of body fat, body water, and muscle, etc.) using video-based gait analysis. This method is more efficient than individual measurement using a conventional body composition meter. Specifically, we designed a deep-learning framework with a convolutional neural network (CNN), where the input is a gait energy image (GEI) and the output consists of the health indicators. Although a vast amount of training data is typically required to train network parameters, it is unfeasible to collect sufficient ground-truth data, i.e., pairs consisting of the gait video and the health indicators measured using a body composition meter for each subject. We therefore use a two-step approach to exploit an auxiliary gait dataset that contains a large number of subjects but lacks the ground-truth health indicators. At the first step, we pre-train a backbone network using the auxiliary dataset to output gait primitives such as arm swing, stride, the degree of stoop, and the body width — considered to be relevant to the health indicators. At the second step, we add some layers to the backbone network and fine-tune the entire network to output the health indicators even with a limited number of ground-truth data points of the health indicators. Experimental results show that the proposed method outperforms the other methods when training from scratch as well as when using an auto-encoder-based pre-training and fine-tuning approach; it achieves relatively high estimation accuracy for the body composition-related health indicators except for body fat-relevant ones.

**key words:** gait analysis, body composition, deep learning, healthcare

## 1. Introduction

Body composition refers to the proportions of major components of the human body such as water, fat, and muscle. Common health indicators such as muscle mass and basal metabolic rate can be calculated from the body composition information. It is also known that the body composition changes and muscle mass decreases with aging [1]. The body composition information is therefore used for health management including prevention and mitigation of lifestyle-related diseases and for management of physical training [2]–[5].

Most existing commercially available body composition meters employ bioelectrical impedance [6], [7]. These methods involve sending a weak alternating current from the subject's hands and feet into the body, and analyzing the

impedance to measure the body composition and estimate related health indicators. Although these methods can accurately measure the indicators, it requires relatively time-consuming steps: removing socks, cleaning hands and feet, and standing on the body composition meter, which may take several minutes in total. Furthermore, the same machine cannot measure multiple people simultaneously.

Because body composition is closely related to human motor function [8]–[10] as well as body shape [11]–[14], there is a possibility that we can estimate body composition and the related health indicators by analyzing the human motor function and body shape, which could be observed from gait videos. We therefore raise a research question on the possibility of health indicator estimation from gait videos as an alternative for the body composition meter. One of the most promising approaches is visually observing the subject walking, i.e., assessing the gait. A gait video contains information pertaining to the subject's motor function and body shape. Thus, it may be possible to estimate the health indicators using video-based gait analysis techniques. Moreover, because gait is the most fundamental mode of locomotion, we have more opportunities to observe it than other actions (e.g., jumping or doing bending exercises).

Video-based gait analysis is the subject of a large body of literature, including but not limited to gait-based estimation of gender, age, and health status [15]–[19]. Generally, the related techniques require just one gait period of a gait video (i.e., approximately one second) for analysis, thus suggesting that the health indicators can also be obtained much faster than with body composition meters.

The above-mentioned studies on video-based gait analysis typically employ machine learning techniques (e.g., classifiers for gender classification, regression techniques for age estimation) and the amount of training data is crucial to the success of these algorithms (e.g., pairs consisting of the gait video and ground-truth age label for age estimation), in particular, in this deep learning era.

It is, however, relatively difficult to collect a sufficiently large amount of training data, particularly if medical data is required. For example, studying the relationship between gait and a certain disease requires establishing collaboration with a hospital and obtaining informed consent from patients. Furthermore, it requires the effort of collecting gait videos of the patients and asking a medical professional to annotate ground-truth labels for each gait video.

Manuscript received January 12, 2021.

Manuscript revised April 24, 2021.

Manuscript publicized July 9, 2021.

<sup>†</sup>The authors are with the Institute of Scientific and Industrial Research, Osaka University, Ibaraki-shi, 567-0047 Japan.

<sup>††</sup>The author is with Faculty of Science and Technology, Seikei University, Tokyo, 180-8633 Japan.

a) E-mail: liao@am.sanken.osaka-u.ac.jp

DOI: 10.1587/transinf.2020ZDP7502

Moreover, the number of instances related to a specific disease is considerably limited compared to healthy subject data. As a result, the researcher must work with small-scale data, and hence tend to rely on handcrafted gait features and classical machine learning or even rule-based methods [18], [20]–[22]. This is in contrast to general computer vision tasks, such as object recognition and human pose estimation, which rely on deep learning frameworks [23], [24] with the help of publicly available and well organized large-scale databases [25], [26].

Some large-scale publicly available databases exist for video-based gait analysis [27]–[29] containing over 10,000 subjects. Although these databases are not aimed at medical/healthcare-oriented applications, but at individual recognition [30]–[32]. We may be able to use their data to pre-train a backbone network for video-based gait analysis, and subsequently fine-tune the network for our task; i.e., health indicator estimation with a limited training data. This is a similar approach to computer vision research that fine-tunes networks pre-trained by standard datasets [25], [26] for specific tasks.

In this paper, we therefore propose a deep learning framework to estimate health indicators via video-based gait analysis even with limited ground-truth training data. The contributions of this work are summarized as follows.

### 1. Health indicator estimation from a walking video.

To our knowledge, this is the first work to estimate health indicators based entirely on a video of the subject walking. This potentially increases the efficiency and ease of health indicator measurement compared to the conventional body composition meter.

### 2. Pre-training and fine-tuning strategies with gait primitives relevant to health indicators.

To enable the use of a deep learning framework even with the limited number of training data for gait-based health indicators, we use pre-training and fine-tuning strategies. Presently, there has not been extensive research on a pre-training strategy before fine-tuning in video-based gait analysis. Thus, we demonstrate that pre-training with gait primitives relevant to health indicators (e.g., arm swing, stride, the degree of stoop, and body width) is beneficial for the subsequent fine-tuning for the health indicator estimation task. This achieves an improvement in accuracy compared with training from scratch or pre-training using a conventional reconstruction task with an auto-encoder.

## 2. Related Work

### 2.1 Video-Based Gait Analysis

There is a rich body of literature on video-based gait analysis which we address in this subsection. Several comprehensive surveys on this topic provide further details for the interested reader [33]–[36].

The majority of studies on this topic pertain to gait recognition, i.e., person identification from walking videos [37]–[43]. Other popular related topics include gen-

der classification [15] and age estimation [44].

Early studies in video-based gait analysis mainly focused on developing handcrafted gait features using both model-based approaches [38] and appearance-based approaches [41], [45]–[48], and applied traditional machine learning techniques (e.g., linear discriminant analysis and support vector machines) to the gait features for tasks such as person identification, gender classification, or age estimation. Publicly available gait video databases [40], [49]–[51] played an important role in training the machine learning models and providing a basis for performance evaluation, although database sizes were limited (i.e., at most a few hundreds subjects).

Subsequently, deep learning frameworks were applied to video-based gait analysis [17], [52]–[62] in a similar fashion to other computer vision topics such as face recognition, object recognition, semantic segmentation, and human pose estimation. Because deep learning frameworks generally require substantial training data, the aforementioned studies typically used larger-scale gait databases [27]–[29], [63] (e.g., over 60,000 subjects in [28], [29]) for training and performance evaluation.

In contrast, it is difficult to collect large-scale gait data in the field of medical/healthcare-oriented gait analysis, both because the number of eligible subjects is limited, and annotation by a medical professional is required. Thus, studies in this field typically rely on small-scale gait databases. As a result, handcrafted features and classical machine learning or rule-based techniques are often employed because they can achieve suitable performance without substantial training data. For example, Liao et al. [20] and Ajay et al. [22] conducted gait analysis studies for specific diseases by using a small number of gait data including only 20 to 30 subjects. Aoki et al. [21] and Matsuura et al. [18] tackled the more general topic of cognitive function estimation from gait for hundreds of subjects, which is a larger dataset than those used in [20], [22], but still too small to apply a deep learning framework.

There are some studies that have applied deep learning frameworks in the field of medical gait analysis. For example, Zhang et al. [64] and Camps et al. [65] applied deep learning frameworks with only 18 and 21 subjects, respectively, with the help of data augmentation to cover the shortage of training data. The data augmentation, however, does not necessarily reflect subject diversity.

### 2.2 Fine-Tuning in Deep Learning

Because deep learning frameworks usually require substantial training data that is not necessarily easy to collect, a researcher may decide to use a network pre-trained for general tasks on large-scale annotated databases (e.g., ImageNet [26], MS-COCO [25]), and fine-tune the network for a specific task such as object recognition [66], [67], action recognition [68], image retrieval [69], age estimation [70], person re-identification [71], and computer-aided diagnosis [72]–[75].

The networks pre-trained for general tasks with general image databases (e.g., for object recognition with ImageNet [26]), however, this data is not necessarily effective for fine-tuning for a specific task, particularly, when there are different domains between the general tasks and the specific tasks. To address this issue, some researchers prepare another training set for pre-training, whose domain/task is more relevant to the target task. For example, Yang et al. [70] pre-trained another network using facial images in addition to the network trained using a large-scale generic image database, and fine-tuned the two pre-trained networks for face-based age estimation. Wang et al. [76] pre-trained a network using pairs of real magnetic resonance images and manual segmentation annotations by radiologists, and fine-tuned the network for medical image segmentation. Gong et al. [77] used simulated positron emission tomography (PET) images to pre-train a network, and fine-tuned the network to improve the quality of PET images.

Similarly, we pre-train our network using large-scale databases for gait analysis, add some layers to the network, and then fine-tune the network for the target task. Specifically, we explore suitable tasks for pre-training in this work to achieve suitable fine-tuning for the subsequent health indicator estimation.

### 3. Health Indicators

We briefly introduce the health indicators used in our work. The indicators are summarized in Table 1. The health indicators are measured using a body composition meter InBody270 (InBody Japan Inc.), and the descriptions of the indicators are as follows:

- Weight, total body water (TBW), protein, minerals, body fat mass (BFM), soft lean mass (SLM), and skeletal

muscle mass (SMM) signify the body weight and the mass of each body composition factor measured directly. The sum of TBW, protein, minerals, and BFM is equal to the weight. SLM is the mass of all of the muscles, and SMM is the mass of muscles to move the skeleton, which is more directly related to the motor function. The percent body fat (PBF) is the ratio [%] of BFM to the body weight.

- Fat free mass (FFM) is calculated as the weight minus BFM. The basal metabolic rate (BMR), which is the energy a person consumes daily to sustain vital activities, can be calculated from the FFM using the Cunningham equation [78]:

$$v_{\text{BMR}} = 370 + 21.6 \times v_{\text{FFM}}.$$

Because the right-hand side of this equation is calculated using only FFMs and constants, we omit the BMR estimation.

- Body mass index (BMI) is calculated using weight  $w$  and manually inputted height  $h$  by the following equation:

$$v_{\text{BMI}} = w/h^2.$$

- InBody score (IBS) is an indicator originally defined by the manufacturer of the body composition meter, and is calculated by comparing the measured value and standard value of BFM and FFM. Obesity degree (OD) is the ratio of measured weight to standard weight. The standard weight is dependent on the height.
- Waist-hip ratio (WHR) is the estimated ratio of waist circumference and hip circumference. Visceral fat level (VFL) is a grading of visceral fat in the horizontal section of the abdomen around the navel. Both of these indicators are estimated by the body composition

**Table 1** Health indicators.

Indicator		Abbreviation	Unit	Measured value	
				Mean	SD
Weight		Weight	kg	60.5	12.3
Total Body Water		TBW	L	33.2	6.90
Protein		Protein	kg	8.93	1.91
Minerals		Minerals	kg	3.15	0.64
Soft Lean Mass		SLM	kg	42.7	8.92
Fat Free Mass		FFM	kg	45.3	9.44
Skeletal Muscle Mass		SMM	kg	24.9	5.76
Segmental	Mean of Arms	LBM_MA	kg	2.15	0.67
Lean Body	Mean of Legs	LBM_ML	kg	7.30	1.72
Mass	Trunk	LBM_T	kg	19.2	4.26
Limbs' Lean Body Mass		L_LBM	kg	18.9	4.71
Body Mass Index		BMI	kg/m <sup>2</sup>	22.2	3.53
InBody Score		IBS	N/A	72.3	5.40
Waist-Hip Ratio		WHR	N/A	0.812	0.0545
Obesity Degree		OD	%	103.3	16.0
Body Fat Mass		BFM	kg	15.2	6.95
Percent Body Fat		PBF	%	24.7	8.46
Segmental	Mean of Arms	BFM_MA	kg	1.00	0.59
Body Fat	Mean of Legs	BFM_ML	kg	2.54	1.06
Mass	Trunk	BFM_T	kg	7.13	3.62
Visceral Fat Level		VFL	Level	5.78	3.41

meter using the measured data.

- Segmental lean body mass (LBM)/body fat mass (BFM) refers to the muscle/body fat mass of each arm/leg and the trunk. The output of the body composition meter includes the value for each limb. Because we observe the subject walking from a side view in this work, it is difficult to distinguish motions derived from left and right arms/legs. We therefore use an average of the LBM/BFM over the left and right arms/legs.
- Skeletal muscle index (SMI) evaluates the motor function and is calculated using the total muscle mass of the limbs  $v_{L\_LBM}$  and the height  $h$  by the following equation:

$$v_{SMI} = v_{L\_LBM}/h^2.$$

This indicator is output by the body composition meter, but data acquisition errors for some subjects preclude us from including it in our work. We therefore substitute SMI with the summation of segmental LBM of the arms and legs to form the limb lean body mass (L\_LBM).

#### 4. Health Indicator Estimation Using Gait-Primitive Networks

##### 4.1 Overview

Figure 1 provides an overview of the proposed method to estimate the health indicators from a walking video. First, we design a deep neural network whose input is a gait template image (e.g., [41], [45]–[48]) and the output is a gait primitive (e.g., arm swing, body width). We pre-train this network using publicly available large-scale gait databases (e.g., [27]–[29]) and refer to it as a gait primitive network. Secondly, we design a deep neural network where the input is the gait template image and the output is the indicators. We achieve this by adding some layers to the gait primitive network and fine-tuning it with a limited number of training data. We describe the details in the following subsections.

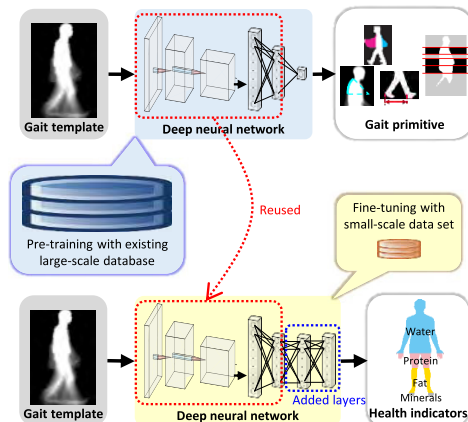


Fig. 1 Overview of the proposed method

##### 4.2 Gait Template Image

Gait recognition researchers have proposed a variety of gait template images such as gait energy image (GEI) [45] a.k.a. averaged silhouette [79], frequency-domain features [41], gait flow image [47], chrono-gait image [39], and masked GEI [48]. Among them, GEI is the most frequently used method in video-based gait analysis because it is simple yet effective.

The GEI is obtained by averaging cropped silhouette images, whose size is 88 by 128 pixels, over one gait period, as shown in Fig. 1 (see inputs for the network). It encodes both static body shape (e.g., body part width with white pixels) and dynamic motion (e.g., arm swing and leg movement with grey pixels). There are some literature reported correlations between muscle strength and walking ability, e.g., walking speed [80]–[82], which can be considered as correlations between muscle mass (i.e., SLM, SMM, etc.) and motion characteristics of gait; while others reported correlations between body fat mass (BFM) or percent body fat (PBF) and body shape parameters such as waist circumference [11]–[14]. In addition, the waist-hip ratio (WHR) itself is an indicator that describes body shape. This suggests that to estimate all of the indicators, the input data is required to contain both motion and shape information. The GEI is therefore suitable as the input for the network to estimate the health indicators as well as the gait primitives. The GEI is therefore suitable as the input for the network to estimate the health indicators as well as the gait primitives.

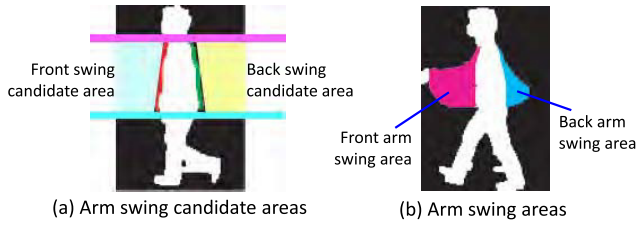
##### 4.3 Gait Primitives

The gait primitives refer to the fundamental components describing a subject's gait characteristics including motion characteristics, like arm swing, stride, and the degree of motion symmetry; pose characteristics, like the degree of stoop and pitch; and shape characteristics, like height and body width, etc. To effectively pre-train the network, the gait primitives must be relevant to the motor function and the body shape, which are subsequently relevant to the health indicators. Furthermore, it is preferable to automatically extract the gait primitives without laborious manual annotation for a large-scale gait database.

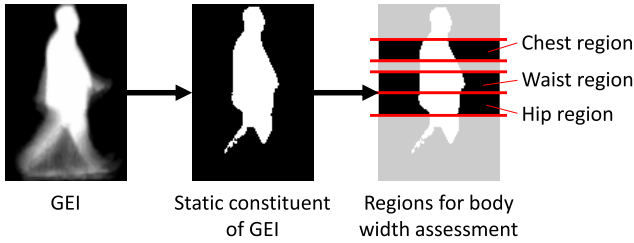
Considering these criteria, we select four gait primitives: forward arm swing, backward arm swing, back straightness, and stride length, as proposed in [83]. These gait primitives are relevant to motion characteristics of walking, and are measurable automatically from a cropped silhouette sequence. Although relatively simple handcrafted methods are used to extract these gait primitives [83], their reliability has been demonstrated through an experience-based long-run exhibition of video-based gait analysis conducted in a science museum, where over 70,000 visitors joined the demonstration over approximately one year [84], [85].

We briefly explain each of the gait primitives in the





**Fig. 2** Measurement of forward and backward arm swing. Reprinted with permission from Springer [83], COPYRIGHT (2009).



**Fig. 3** Measurement of body width. The average row width of the silhouette in each region is calculated as the body width

following paragraphs and refer interested readers to [83] for further details.

- **Forward arm swing (FAS) and backward arm swing (BAS)**

The front-end and back-end lines of the torso are extracted from a cropped silhouette at a single support phase and then forward/backward arm swing candidate regions are set. Areas swept by silhouettes in the forward/backward arm swing candidate region are counted as the degree of forward/backward arm swing, as shown in Fig. 2

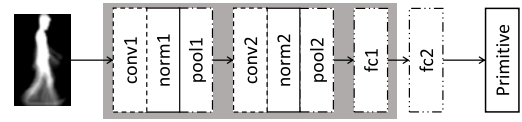
- **Back straightness (BS)**

The back-end line of the torso is extracted and the slope of the back-end line is regarded as the back straightness. Specifically, if the line is more vertical, the back is straighter, and a less vertical result signifies stoop.

- **Stride length (SL)**

The gait period is obtained as the time shift that maximizes the auto-correlation of the cropped silhouette sequence along the temporal axis [41]. The total traveling distance during a walking sequence is computed by subtracting the start position from the end position, and the walking speed is then computed by dividing the total traveling distance by the elapsed time during the walking sequence. Finally, the stride length is regarded as the traveling distance over half of a gait period, and hence it is calculated as the product of the walking speed and half of the gait period.

Furthermore, since the above-mentioned four gait primitives only contain motion and pose characteristics and do not contain shape characteristics, we designed an automatically extracted shape-oriented primitive, i.e., body width shown in Fig. 3. We first extract an almost static foreground part of the GEI by thresholding with 75% of



**Fig. 4** Structure of the gait primitive network that takes a GEI as input and outputs the gait primitive. The abbreviations conv, norm, pool, and fc refer to convolutional layer, normalization layer, pooling layer, and fully connected layer, respectively. This convention is consistent throughout this paper. The layers with gray background rectangles are used in fine-tuning.

**Table 2** Layer configurations for GEINet. Act. denotes the activation function.

Layer	#Kernels	Size/stride	Act.	Pooling
conv1	81	$5 \times 5 \times 1/1$	ReLU	
pool1		$3 \times 3/2$		Max pooling
conv2	45	$7 \times 7 \times 81/1$	ReLU	
pool2		$2 \times 2/2$		Max pooling

the maximum intensity and compute the average width of the static part for different height ranges: a chest region ( $24 \leq y \leq 40$ ), a waist region ( $48 \leq y \leq 64$ ), and a hip region ( $64 \leq y \leq 80$ ).

#### 4.4 Pre-Training the Gait Primitive Network

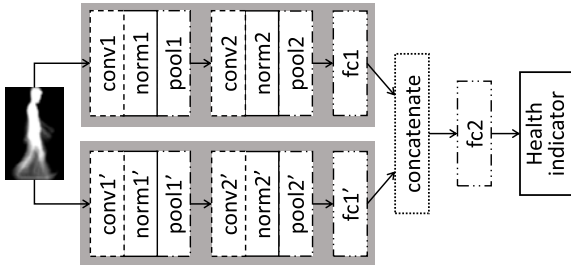
We used GEINet [54] as the backbone for our gait primitive network. GEINet is a standard convolutional neural network (CNN) and is utilized in tasks such as person identification [54], gender classification, and age estimation [17].

The structure of the gait primitive network is almost the same as the original GEINet [54], as shown in Fig. 4, where the leading six layers are two sequential triplets of convolution, pooling, and normalization layers, followed by two fully connected layers. While the last layer outputs 1,024 units in the original GEINet for person identification, the last layer (fc2) in our gait primitive network outputs only one unit, because each gait primitive is defined as a scalar value (i.e., one dimension). The penultimate layer (fc1) outputs 1024 units. The other detailed configurations for the convolution and pooling layers are shown in Table 2, where the number and size of the kernels are slightly different from the original GEINet [54] to suit our health indicator estimation task.

Once the gait primitive network structure is designed, we pre-train each gait primitive network using a training set composed of pairs of GEIs and each corresponding gait primitive so as to minimize a loss function, i.e., mean absolute error (MAE) between estimated gait primitives through the gait primitive network and gait primitives measured by the handcrafted method in [83] (i.e., a sort of the ground-truth gait primitives).

#### 4.5 Fine-Tuning the Gait Primitive Network for Health Indicator Estimation

We fine-tune the gait primitive networks for health indicator estimation. Because the output from the last layer (fc2)



**Fig. 5** Network structure for health indicator estimation in conjunction with two gait primitive networks. The outputs from the penultimate layers of the two gait primitive networks are concatenated and fed into a fully connected layer to estimate health indicators. The two sets of layers with gray background rectangles are from the pre-trained primitive networks.

of the gait primitive network is just a single unit (i.e., one dimension), the output itself does not contain sufficient information for health indicator estimation. Instead, we utilize the output of the penultimate layer (fc1), which comprises 1024-dimensions, for health indicator estimation. The output from the penultimate layer is fed into a fully connected layer (fc2) to estimate one health indicator. Specifically, we fine-tune the layers from the gait primitive network (i.e., conv1, conv2, fc1) while we train the newly added fully connected layer fc2 from scratch. For this purpose, we first define the MAE for the health indicator as  $L_{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$ , where  $N$  is the number of training data,  $\hat{y}_i$  and  $y_i$  are the estimated health indicator through the network (i.e., an output of fc2) and the ground-truth of the health indicator for the  $i$ -th sample. We then use the  $L_{MAE}$  as the loss function to minimize during network training.

We also use multiple gait primitive networks for further improvement, as shown in Fig. 5. We concatenate the outputs from the penultimate layers of the multiple primitive networks, and the concatenated feature is fed into a fully connected layer fc2. Thereafter, we train the network similarly to the aforementioned case of the single gait primitive network.

Some of the health indicators, such as the body weight and the mass of various body components, are closely related to the subject's height. However, the input to the network (i.e., the GEI) lacks height information because we use height-normalized cropped silhouettes. The health indicators that are represented by mass in kilograms (e.g., weight and SLM) and volume in liters (e.g., TBW) are also normalized by the height, specifically, divided by the cubic height, and height-normalized health indicators are set as estimation targets. We require the subject's height information for this normalization, and hence we assume that the subject provides the height information or we automatically measure the subject's height from a captured image with a calibrated camera and a ground plane constraint [86]. Once we estimate the height-normalized health indicators, we obtain the original health indicators by un-normalizing them (i.e., multiplying by the cubic height).

## 5. Experiments

### 5.1 Dataset

We used OU-ISIR Gait Database, Large Population Dataset with Age (OULP-Age) [29] for pre-training the gait primitive networks. More specifically, we extracted a subset of 40,000 subjects out of 63,846 subjects.

We conducted experiments to collect the data used for training and evaluating health indicator estimation because there is no such publicly available database. We conducted the experiments twice in March 2019 and September 2019, and recruited a total of 332 subjects (167 females and 165 males). The ages of the subjects range from 8 to 71 years old, with a mean age of 34.6 years and a standard deviation of 15.6 years. We obtained informed consent from the subjects to use the data for research purposes.

We asked each subject to walk on a pre-defined course of approximately 8 meters at his/her natural walking speed in two sessions: in the morning and in the afternoon. We thus obtained 653 valid gait videos after removing those with unnatural walking actions (e.g., touching the head or hands on the waist while walking). A camera was set up approximately five-meters away from the walking course to capture the walking subject from a side view. The captured walking video's frame size was 640 by 480 pixels and it contained 30 frames per second (fps) for 2.5 seconds (75 frames in total). We extracted silhouettes from the original walking video, cropped the height-normalized silhouettes, and then extracted the GEIs.

We also measured the health indicators of each subject using the InBody270 body composition meter. Because the body composition meter requires the subject's height to measure the health indicators, we asked each subject his/her height in advance. Because of time limitations, each subject's health indicators were measured only in one session, we thus associated the single health indicator measurement with both GEIs extracted in the two walking sessions for each subject.

### 5.2 Setup

We employed 20-fold cross-validation to fine-tune and evaluate performance. We randomly divide the entire dataset into 20 groups with approximately equal number of subjects. We use one group as the validation set, and the rest as the training set in the fine-tuning. That is, 315 or 316 subjects were used as the training set and 16 or 17 subjects were used as the validation set.

We evaluated the accuracy using the relative errors of the health indicators, defined as the ratio of the MAE of the estimated health indicator to the mean of the measured (ground-truth) values. This scheme enables us to evaluate the accuracy despite the varying scales of the different health indicators.

In the training stage, the size of mini-batches was 128

and the initial learning rate was set to 0.001. The number of epochs was set to 200 and 250 for pre-training and fine-tuning, respectively.

### 5.3 Benchmarks

Because this is the first work to address health indicator estimation from gait, there are no benchmarks for this specific task. We therefore prepare a suite of benchmarks for this purpose.

Firstly, to verify the effectiveness of our pre-training and fine-tuning strategy with the gait primitive networks, we trained the entire network for health indicator estimation from scratch, denoted as DL (scratch). Note that this is equivalent to the proposed method without pre-training.

Secondly, to verify the effectiveness of the gait primitives for pre-training, we pre-trained the network using an auto-encoder (see Fig. 6) to minimize the reconstruction errors of the GEIs. The pre-trained encoder part of the network, whose structure is identical to that of the gait primitive network except for the absence of the last layer (fc2), was used for fine-tuning in the same way as the proposed method, and is denoted as DL (auto-encoder + fine-tuning).

Finally, we employed support vector regression (SVR) as a family of classical machine learning. The classical

machine learning usually works better even with less number of training data than deep learning-based approaches, while its capability is usually inferior to the deep learning-based approaches. We therefore investigated the trade-off between the number of training data and the capability by comparing the proposed method with SVR.

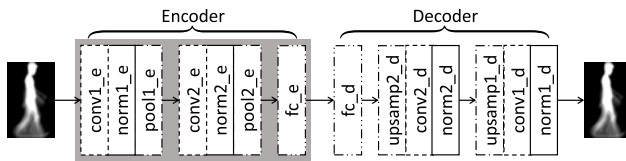
### 5.4 Ablation Studies on Gait Primitives

We conducted ablation studies to determine the best use schema of the primitive networks.

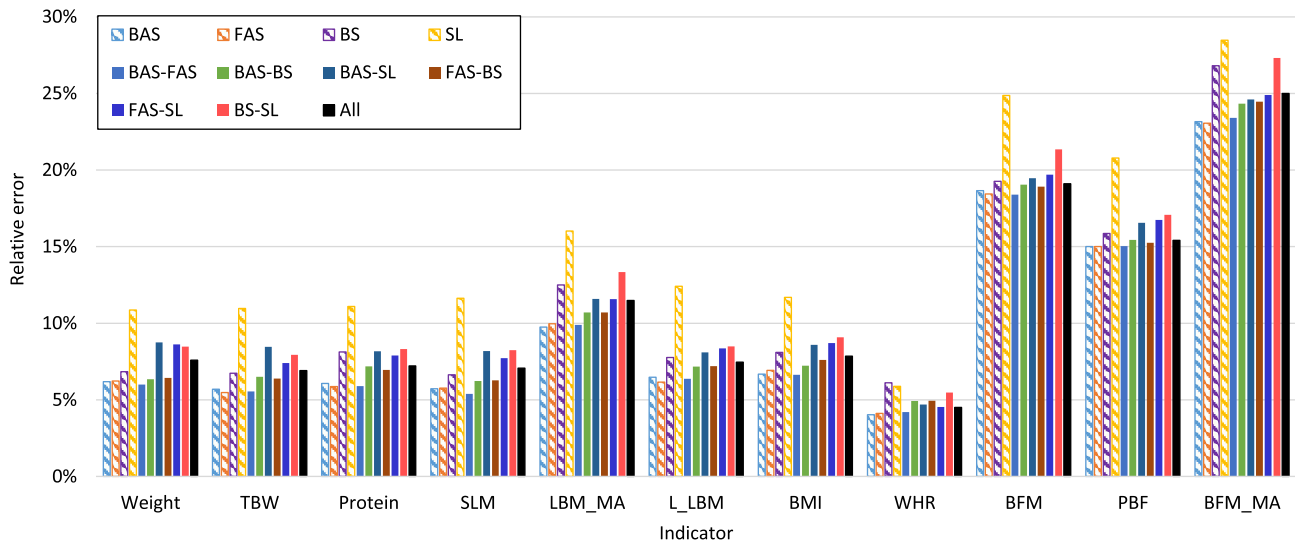
For the gait primitives with motion and pose characteristics, we report the relative errors for particular health indicators when each single gait primitive is used, each combination of two gait primitives are used, and all four gait primitives are used in Fig. 7. We see that the estimation error increases significantly when we pre-train the networks with the step length gait primitive. The estimation errors using the other gait primitives did not show significant variation. Furthermore, pre-training using all of the gait primitives does not outperform dual network combinations (e.g., FAS + BAS). Therefore, considering cost-effectiveness, we selected the dual network combination yielding the highest accuracy — forward/backward arm swing — as the best schema.

On the other hand, for the gait primitives with shape characteristics, we similarly report the relative errors when pre-training using single, dual or all the three regional body widths as gait primitives in Fig. 8. We found that combining multiple primitive networks resulted in marginal improvement in accuracy when using gait primitives with shape characteristics. Therefore, we chose the body width of the hip region, with the lowest error among the single primitives, as the best schema.

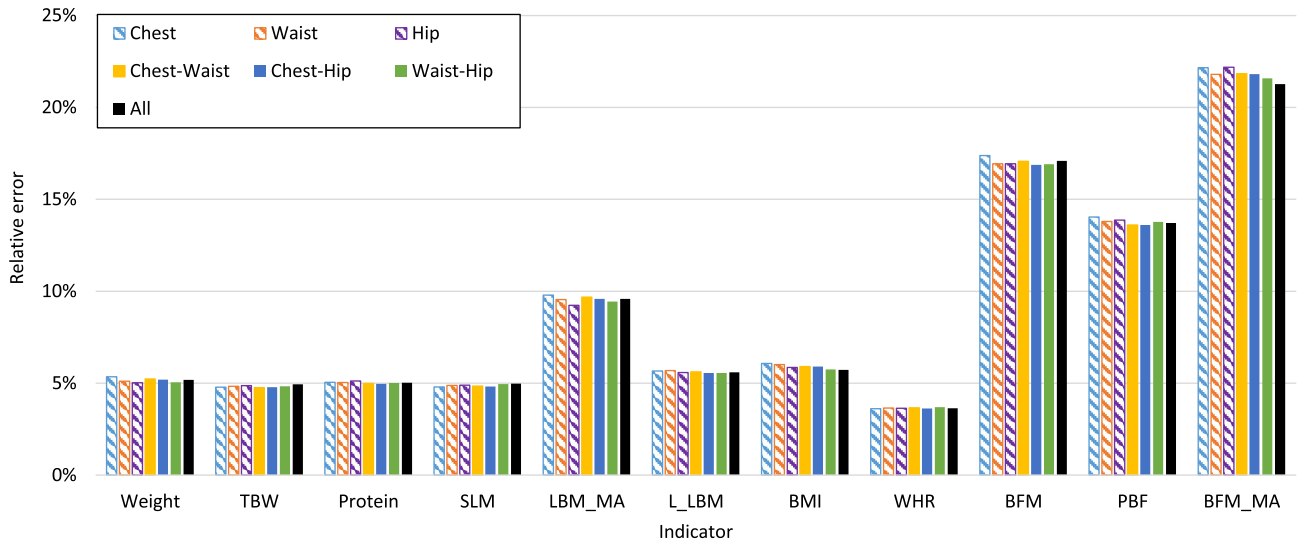
We compare the original MAEs and relative errors of



**Fig. 6** Network structure of the auto-encoder. The abbreviation upsamp indicates an upsampling layer. We used the encoder part (layers with gray background rectangles) for fine-tuning.



**Fig. 7** Comparison of the relative errors among the gait primitives with motion/pose characteristics. Hatched colored bars indicate a single gait primitive, while solid colored bars indicate combinations of two gait primitives. The solid black bar indicates all four gait primitives.



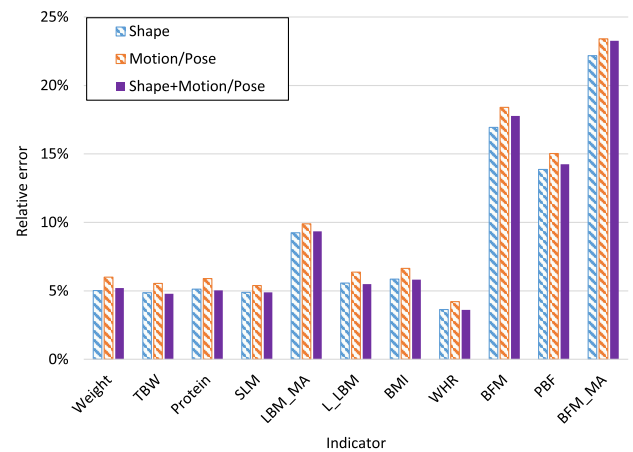
**Fig. 8** Comparison of the relative errors among the gait primitives with shape characteristics. Same as Fig. 7, hatched colored bars indicate a single gait primitive, while solid colored bars indicate combinations of two gait primitives. The solid black bar indicates all three gait primitives.

**Table 3** MAEs and relative errors of the proposed method of using gait primitive with motion/pose characteristics and shape characteristics, respectively

Indicator	Mean MAE		Relative error	
	Motion/pose	Shape	Motion/pose	Shape
Weight	3.63	3.04	6.00%	5.02%
TBW	1.84	1.62	5.54%	4.86%
Protein	0.53	0.46	5.90%	5.12%
Minerals	0.18	0.16	5.69%	4.95%
SLM	2.30	2.09	5.39%	4.89%
FFM	2.45	2.20	5.41%	4.86%
SMM	1.53	1.41	6.15%	5.63%
LBM_MA	0.21	0.20	9.89%	9.24%
LBM_ML	0.43	0.37	5.94%	5.02%
LBM_T	1.15	1.06	6.00%	5.50%
L_LBM	1.20	1.05	6.37%	5.58%
BMI	1.47	1.30	6.64%	5.85%
IBS	3.89	3.34	5.38%	4.62%
WHR	0.03	0.03	4.21%	3.63%
OD	6.51	5.84	6.30%	5.65%
BFM	2.80	2.57	18.40%	16.93%
PBF	3.72	3.43	15.03%	13.87%
BFM_MA	0.23	0.22	23.41%	22.18%
BFM_ML	0.43	0.41	16.95%	16.12%
BFM_T	1.51	1.37	21.14%	19.14%
VFL	1.51	1.40	26.14%	24.27%

the two best schemas in Table 3. The comparison shows that pretraining using gait primitives with shape characteristics results in smaller errors in all indicators than using those with motion characteristics. We also tried to combine the primitive networks of the two best schemas (see Fig. 9), but no significant improvement in accuracy was obtained. Therefore, we decided to use the gait primitives with shape features, i.e., hip region body width, as the proposed method.

On the other hand, we see that the proposed method with the body shape-oriented gait primitive can relatively



**Fig. 9** Comparison of the respective best schemes for gait primitives with motion/pose characteristics and shape characteristics, and their combinations. The “shape” one is body wide in the hip region and the “motion/pose” one is the combination of forward and backward arm swing.

accurately estimate the health indicators other than body fat-relevant ones (i.e., the relative errors are approximately 5%). However, the indicators related to body fat show higher errors (e.g., from 15 to 30% error) than the other indicators. We hypothesize that this may be related to the subjects’ clothing — participants in the experiment conducted in March generally wore thicker clothing, and their body shape would have been obscured and not properly represented by the silhouette. Furthermore, since the body composition meter needs to ensure that the sum of TBW, protein, minerals and BFM output is exactly equal to the weight, it is possible that the measured weight minus the first three values will be used as the BFM measurement; and the deviation of the measured weight of the subjects will also vary due to the different clothing. Further validation is therefore



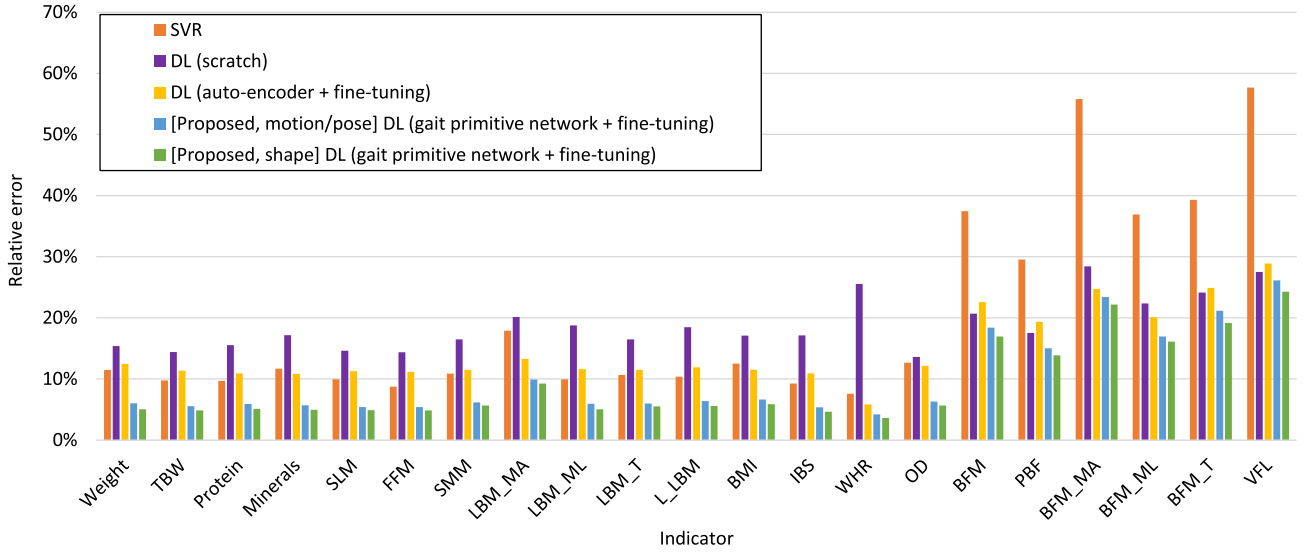


Fig. 10 Comparison of the relative errors with the benchmarks.

still needed in future.

### 5.5 Comparison with Benchmarks

We evaluated the relative errors among the benchmarks as shown in Fig. 10. SVR (i.e., a classical machine learning technique) underperforms compared to deep learning from scratch for body fat-related indicators, but these two methods are comparable for the other indicators. Among the deep learning methods, deep learning frameworks trained from scratch show higher estimation errors for most of the indicators as compared to the method that is fine-tuned with auto-encoder. Furthermore, the proposed deep learning framework fine-tuned with the gait primitive network yields the highest accuracy. This indicates the effectiveness of the proposed pre-training and fine-tuning strategy compared to training from scratch when using a limited number of training samples for the target task. Also, we observe that it is essential to pre-train the network with features relevant to the target task (i.e., gait primitives for health indicator estimation) rather than pre-training in a general way (i.e., using an auto-encoder). Without proper pre-training the deep learning methods lose their effectiveness and may become inferior to classical machine learning methods (e.g., while SVR yields less than 10% error for TBW, pre-training with auto-encoder yields over 10% error).

### 5.6 Sensitivity Analysis of the Number of Training Samples

Because we claim the effectiveness of the proposed method for a relatively small number of training samples for the target task, we conducted a sensitivity analysis of the number of training samples on the accuracy of health indicator estimation. For this purpose, we evaluated the relative errors of the proposed method for selected health indicators as we decreased the number of training samples, as shown in

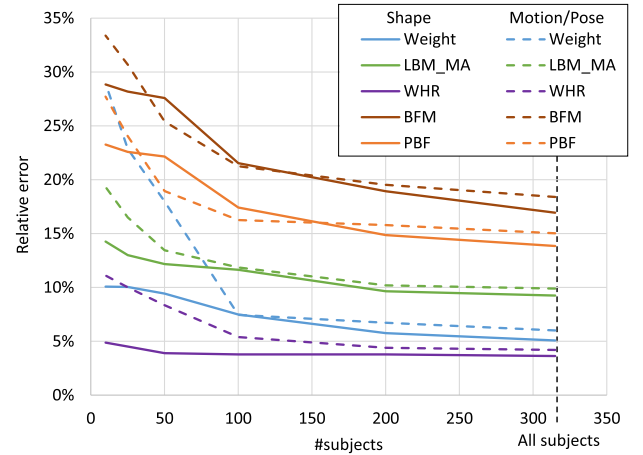


Fig. 11 Sensitivity analysis of the number of training samples on the relative errors.

Fig. 11. This experiment indicates that the proposed method can adequately accomplish the target task when the number of training samples exceeds 100.

### 5.7 Discussion

Although the proposed method still has some limitations, such as a large error in estimation of body fat-related indicators, its advantages over existing body composition meters still exist. In our data collection experiments, it took about 2 minutes per subject to measure health indicators using a body composition meter, including preparation time (e.g., taking off and putting on shoes and socks). On the other hand, it took only 6 to 8 seconds per subject to capture a gait video, i.e., much faster than the body composition meter. Moreover, while the body composition meter requires subject's contact to the device, the gait video capturing does not require the subject's contact, which is

preferable under the situation of the current COVID-19 pandemic.

## 6. Conclusion

We proposed a method to estimate health indicators related to body composition using video-based gait analysis by a deep learning framework. To address the challenge of the small sample size, we pre-trained the gait primitive networks with a large-scale gait database, and fine-tuned them with a limited number of health indicator training samples. We confirmed that the proposed method achieved fairly low relative errors (approximately 5%) for the health indicators other than body fat-relevant ones, and outperformed the benchmarks.

Future research directions include incorporating age and gender information for health indicator estimation because the correlation between these factors is well established. In addition, the temporal information of gait is not exploited by the estimation due to using GEI as input. Therefore, we will try to use gait features that contain temporal information, e.g., silhouette sequences. Furthermore, although we used only side-view gait images in this study, we believe that estimation using data from other view angles is also possible. Since the multi-view large-scale gait database [29] has been already available, we will try to extend the proposed method to allow the estimation of health indicators using an arbitrary-view gait video.

## Acknowledgments

This work was supported by JSPS Grants-in-Aid for Scientific Research (A) JP18H04115, JST-Mirai Program Grant Number JPMJMI17DH, Japan, and Innovation Platform for Society 5.0 from Japan Ministry of Education, Culture, Sports, Science and Technology (Code: S004541).

## References

- [1] G.B. Forbes and J.C. Reina, "Adult lean body mass declines with age: Some longitudinal observations," *Metabolism*, vol.19, no.9, pp.653–663, Sept. 1970.
- [2] A.S. Jackson and M.L. Pollock, "Practical Assessment of Body Composition," *The Physician and Sportsmedicine*, vol.13, no.5, pp.76–90, May 1985.
- [3] R. Baumgartner, "Body composition in healthy aging," *Annals of the New York Academy of Sciences*, vol.904, pp.437–448, 2000.
- [4] K. Sanada, M. Miyachi, I. Tabata, K. Suzuki, K. Yamamoto, H. Kawano, C. Usui, and M. Higuchi, "Differences in body composition and risk of lifestyle-related diseases between young and older male rowers and sedentary controls," *Journal of Sports Sciences*, vol.27, no.10, pp.1027–1034, Aug. 2009.
- [5] J.S. Brach, E.M. Simonsick, S. Kritchevsky, K. Yaffe, and A.B. Newman, "The Association Between Physical Function and Lifestyle Activity and Exercise in the Health, Aging and Body Composition Study," *Journal of the American Geriatrics Society*, vol.52, no.4, pp.502–509, 2004.
- [6] R.F. Kushner and D.A. Schoeller, "Estimation of total body water by bioelectrical impedance analysis," *The American Journal of Clinical Nutrition*, vol.44, no.3, pp.417–424, Sept. 1986.
- [7] H.C. Lukaski, W.W. Bolonchuk, C.B. Hall, and W.A. Siders, "Validation of tetrapolar bioelectrical impedance method to assess human body composition," *Journal of Applied Physiology*, vol.60, no.4, pp.1327–1332, April 1986.
- [8] B. Sternfeld, L. Ngo, W.A. Satariano, and I.B. Tager, "Associations of Body Composition with Physical Performance and Self-reported Functional Limitation in Elderly Men and Women," *American Journal of Epidemiology*, vol.156, no.2, pp.110–121, July 2002.
- [9] A. Özkan, G. Kayihan, Y. Köklü, N. Ergun, M. Koz, G. Ersöz, and A. Dellal, "The Relationship Between Body Composition, Anaerobic Performance and Sprint Ability of Amputee Soccer Players," *Journal of Human Kinetics*, vol.35, pp.141–146, Dec. 2012.
- [10] D. Stodden, S. Langendorfer, and M.A. Robertson, "The Association Between Motor Skill Competence and Physical Fitness in Young Adults," *Research Quarterly for Exercise and Sport*, vol.80, no.2, pp.223–229, June 2009. Publisher: Routledge.
- [11] P. Suchanek, I. Kralova Lesna, O. Mengerova, J. Mrazkova, V. Lanska, and P. Stavek, "Which index best correlates with body fat mass: BAI, BMI, waist or WHR?," *Neuro Endocrinology Letters*, vol.33, Suppl 2, pp.78–82, Jan. 2012.
- [12] P. Brambilla, G. Bedogni, M. Heo, and A. Pietrobelli, "Waist circumference-to-height ratio predicts adiposity better than body mass index in children and adolescents," *International Journal of Obesity*, vol.37, no.7, pp.943–946, July 2013. Number: 7 Publisher: Nature Publishing Group.
- [13] P.T. Katzmarzyk and C. Bouchard, "Where is the beef? Waist circumference is more highly correlated with BMI and total body fat than with abdominal visceral fat in children," *International Journal of Obesity*, vol.38, no.6, pp.753–754, June 2014. Number: 6 Publisher: Nature Publishing Group.
- [14] S.M. Camhi, G.A. Bray, C. Bouchard, F.L. Greenway, W.D. Johnson, R.L. Newton, E. Ravussin, D.H. Ryan, S.R. Smith, and P.T. Katzmarzyk, "The Relationship of Waist Circumference and BMI to Visceral, Subcutaneous, and Total Body Fat: Sex and Race Differences," *Obesity*, vol.19, no.2, pp.402–408, 2011.
- [15] 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.
- [16] Y. Makihara, H. Mannami, and Y. Yagi, "Gait Analysis of Gender and Age Using a Large-Scale Multi-view Gait Database," *Computer Vision—ACCV 2010*, ed. R. Kimmel, R. Klette, and A. Sugimoto, *Lecture Notes in Computer Science*, pp.440–451, Springer Berlin Heidelberg, 2011.
- [17] A. Sakata, Y. Makihara, N. Takemura, D. Muramatsu, and Y. Yagi, "Gait-Based Age Estimation Using a DenseNet," *Computer Vision—ACCV 2018 Workshops*, ed. G. Carneiro and S. You, *Lecture Notes in Computer Science*, pp.55–63, Springer International Publishing, 2019.
- [18] T. Matsuura, K. Sakashita, A. Grushnikov, F. Okura, I. Mitsugami, and Y. Yagi, "Statistical Analysis of Dual-task Gait Characteristics for Cognitive Score Estimation," *Scientific Reports*, vol.9, no.1, pp.1–12, Dec. 2019.
- [19] R. Chang, L. Guan, and J.A. Burne, "An automated form of video image analysis applied to classification of movement disorders," *Disability and Rehabilitation*, vol.22, no.1-2, pp.97–108, Jan. 2000.
- [20] R. Liao, Y. Makihara, D. Muramatsu, I. Mitsugami, Y. Yagi, K. Yoshiyama, H. Kazui, and M. Takeda, "A video-based gait disturbance assessment tool for diagnosing idiopathic normal pressure hydrocephalus," *IEEE Transactions on Electrical and Electronic Engineering*, vol.15, no.3, pp.433–441, 2020.
- [21] K. Aoki, T.T. Ngo, I. Mitsugami, F. Okura, M. Niwa, Y. Makihara, Y. Yagi, and H. Kazui, "Early Detection of Lower MMSE Scores in Elderly Based on Dual-Task Gait," *IEEE Access*, vol.7, pp.40085–40094, 2019. Conference Name: IEEE Access.
- [22] J. Ajay, C. Song, A. Wang, J. Langan, Z. Li, and W. Xu, "A pervasive and sensor-free Deep Learning system for Parkinsonian gait analysis," 2018 IEEE EMBS International Conference on

- Biomedical Health Informatics (BHI), pp.108–111, March 2018.
- [23] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems* 25, ed. F. Pereira, C. Burges, L. Bottou, and K. Weinberger, pp.1097–1105, Curran Associates, Inc., 2012.
  - [24] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21–26, 2017, pp.1302–1310, IEEE Computer Society, 2017.
  - [25] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C.L. Zitnick, “Microsoft coco: Common objects in context,” *Computer Vision – ECCV 2014*, ed. D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Cham, pp.740–755, Springer International Publishing, 2014.
  - [26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” *Proc. 22nd IEEE Conf. on Computer Vision and Pattern Recognition*, 2009.
  - [27] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, “Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition,” *IPSI Transactions on Computer Vision and Applications*, vol.10, no.1, p.4, Feb. 2018.
  - [28] M.Z. Uddin, T.T. Ngo, Y. Makihara, N. Takemura, X. Li, D. Muramatsu, and Y. Yagi, “The ou-isir large population gait database with real-life carried object and its performance evaluation,” *IPSI Transactions on Computer Vision and Applications*, vol.10, no.1, p.5, May 2018.
  - [29] C. Xu, Y. Makihara, G. Ogi, X. Li, Y. Yagi, and J. Lu, “The OU-ISIR Gait Database comprising the Large Population Dataset with Age and performance evaluation of age estimation,” *IPSI Transactions on Computer Vision and Applications*, vol.9, no.1, p.24, Dec. 2017.
  - [30] H. Chao, Y. He, J. Zhang, and J. Feng, “Gaitset: Regarding gait as a set for cross-view gait recognition,” *Proc. 33rd AAAI Conf. on Artificial Intelligence (AAAI 2019)*, 2019.
  - [31] C. Fan, Y. Peng, C. Cao, X. Liu, S. Hou, J. Chi, Y. Huang, Q. Li, and Z. He, “Gaitpart: Temporal part-based model for gait recognition,” *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
  - [32] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, “Gait recognition via semi-supervised disentangled representation learning to identity and covariate features,” *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
  - [33] M.S. Nixon, T.N. Tan, and R. Chellappa, *Human Identification Based on Gait*, Int. Series on Biometrics, Springer-Verlag, Dec. 2005.
  - [34] Y. Makihara, D.S. Matovski, M.S. Nixon, J.N. Carter, and Y. Yagi, *Gait Recognition: Databases, Representations, and Applications*, pp.1–15, John Wiley & Sons, Inc., 1999.
  - [35] P. Connor and A. Ross, “Biometric recognition by gait: A survey of modalities and features,” *Computer Vision and Image Understanding*, vol.167, pp.1–27, 2018.
  - [36] C. Wan, L. Wang, and V.V. Phoha, “A survey on gait recognition,” *ACM Comput. Surv.*, vol.51, no.5, pp.89:1–89:35, Aug. 2018.
  - [37] S. Niyogi and E. Adelson, “Analyzing and recognizing walking figures in xyt,” *Proc. 7th IEEE Conf. on Computer Vision and Pattern Recognition*, pp.469–474, 1994.
  - [38] D. Cunado, M. Nixon, and J. Carter, “Automatic extraction and description of human gait models for recognition purposes,” *Computer Vision and Image Understanding*, vol.90, no.1, pp.1–41, 2003.
  - [39] L. Wang, T. Tan, H. Ning, and W. Hu, “Silhouette analysis-based gait recognition for human identification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.25, no.12, pp.1505–1518, Dec. 2003.
  - [40] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, 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. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
  - [41] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, “Gait recognition using a view transformation model in the frequency domain,” *Proc. 9th European Conference on Computer Vision*, Graz, Austria, pp.151–163, May 2006.
  - [42] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, “Support vector regression for multi-view gait recognition based on local motion feature selection,” *Proc. IEEE computer society conference on Computer Vision and Pattern Recognition 2010*, San Francisco, CA, USA, pp.1–8, June 2010.
  - [43] J. Lu and Y.-P. Tan, “Uncorrelated discriminant simplex analysis for view-invariant gait signal computing,” *Pattern Recognition Letters*, vol.31, no.5, pp.382–393, 2010.
  - [44] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi, “Gait-based age estimation using a whole-generation gait database,” *Proc. Int. Joint Conf. on Biometrics (IJCB2011)*, Washington D.C., USA, pp.1–6, Oct. 2011.
  - [45] 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. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
  - [46] C. Wang, J. Zhang, L. Wang, J. Pu, and X. Yuan, “Human identification using temporal information preserving gait template,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.34, no.11, pp.2164–2176, Nov. 2012.
  - [47] T.H.W. Lam, K.H. Cheung, and J.N.K. Liu, “Gait flow image: A silhouette-based gait representation for human identification,” *Pattern Recognition*, vol.44, no.4, pp.973–987, April 2011.
  - [48] K. Bashir, T. Xiang, and S. Gong, “Gait recognition without subject cooperation,” *Pattern Recognition Letters*, vol.31, no.13, pp.2052–2060, Oct. 2010.
  - [49] S. Yu, D. Tan, and T. Tan, “A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition,” *18th International Conference on Pattern Recognition (ICPR’06)*, vol.4, pp.441–444, Aug. 2006. ISSN: 1051-4651.
  - [50] J. Shutler, M. Grant, M. Nixon, and J. Carter, “On a large sequence-based human gait database,” *Proc. 4th Int. Conf. on Recent Advances in Soft Computing*, Nottingham, UK, pp.66–71, Dec. 2002.
  - [51] Y. Makihara, H. Mannami, A. Tsuji, M. Hossain, K. Sugiura, A. Mori, and Y. Yagi, “The ou-isir gait database comprising the treadmill dataset,” *IPSI Transactions on Computer Vision and Applications*, vol.4, pp.53–62, April 2012.
  - [52] T. Wolf, M. Babae, and G. Rigoll, “Multi-view gait recognition using 3D convolutional neural networks,” *2016 IEEE International Conference on Image Processing (ICIP)*, pp.4165–4169, Sept. 2016. ISSN: 2381-8549.
  - [53] T. Yeoh, H.E. Aguirre, and K. Tanaka, “Clothing-invariant gait recognition using convolutional neural network,” *2016 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp.1–5, Oct. 2016.
  - [54] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, “GEINet: View-invariant gait recognition using a convolutional neural network,” *2016 International Conference on Biometrics (ICB)*, pp.1–8, June 2016.
  - [55] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, “A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.39, no.2, pp.209–226, Feb. 2017. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
  - [56] M. Alotaibi and A. Mahmood, “Improved gait recognition based on specialized deep convolutional neural network,” *Computer Vision and Image Understanding*, vol.164, pp.103–110, Nov. 2017.
  - [57] W. Liu, C. Zhang, H. Ma, and S. Li, “Learning Efficient Spatial-Temporal Gait Features with Deep Learning for Human Identification,” *Neuroinformatics*, vol.16, no.3, pp.457–471, Oct. 2018.
  - [58] M. Babae, L. Li, and G. Rigoll, “Gait Energy Image

- Reconstruction from Degraded Gait Cycle Using Deep Learning,” *Computer Vision–ECCV 2018 Workshops*, ed. L. Leal-Taixé and S. Roth, Lecture Notes in Computer Science, Cham, pp.654–658, Springer International Publishing, 2019.
- [59] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, “Gait recognition invariant to carried objects using alpha blending generative adversarial networks,” *Pattern Recognition*, vol.105, p.107376, Sept. 2020.
- [60] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, “Gait-based human age estimation using age group-dependent manifold learning and regression,” *Multimedia Tools and Applications*, vol.77, no.21, pp.28333–28354, Nov. 2018.
- [61] H. Zhu, Y. Zhang, G. Li, J. Zhang, and H. Shan, “Ordinal distribution regression for gait-based age estimation,” *Science China Information Sciences*, vol.63, no.2, p.120102, Jan. 2020.
- [62] T. Liu, X. Ye, and B. Sun, “Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition,” *2018 Chinese Automation Congress (CAC)*, pp.3477–3481, Nov. 2018.
- [63] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, “The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition,” *IEEE Transactions on Information Forensics and Security*, vol.7, no.5, pp.1511–1521, Oct. 2012. Conference Name: IEEE Transactions on Information Forensics and Security.
- [64] S. Zhang, S. Poon, K. Vuong, A. Sneddon, and C. Loy, “A Deep Learning-Based Approach for Gait Analysis in Huntington Disease,” *Studies in Health Technology and Informatics*, vol.264, pp.477–481, Aug. 2019.
- [65] J. Camps, A. Samà, M. Martín, D. Rodríguez-Martín, C. Pérez-López, J.M. Moreno Arostegui, J. Cabestany, A. Català, S. Alcaine, B. Mestre, A. Prats, M.C. Crespo-Maraver, T.J. Counihan, P. Browne, L.R. Quinlan, G.O. Laighin, D. Sweeney, H. Lewy, G. Vainstein, A. Costa, R. Annicchiarico, A. Bayés, and A. Rodríguez-Molinero, “Deep learning for freezing of gait detection in Parkinson’s disease patients in their homes using a waist-worn inertial measurement unit,” *Knowledge-Based Systems*, vol.139, pp.119–131, Jan. 2018.
- [66] T.-Y. Lin, A. RoyChowdhury, and S. Maji, “Bilinear CNN Models for Fine-Grained Visual Recognition,” *2015 IEEE International Conference on Computer Vision (ICCV)*, pp.1449–1457, Dec. 2015. ISSN: 2380-7504.
- [67] A.K. Reyes, J.C. Caicedo, and J.E. Camargo, “Fine-tuning Deep Convolutional Networks for Plant Recognition,” *CLEF (Working Notes)*, vol.1391, pp.467–475, 2015.
- [68] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, “Dynamic Image Networks for Action Recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.3034–3042, June 2016. ISSN: 1063-6919.
- [69] F. Radenović, G. Tolias, and O. Chum, “Fine-Tuning CNN Image Retrieval with No Human Annotation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.41, no.7, pp.1655–1668, July 2019. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [70] X. Yang, B.-B. Gao, C. Xing, Z.-W. Huo, X.-S. Wei, Y. Zhou, J. Wu, and X. Geng, “Deep Label Distribution Learning for Apparent Age Estimation,” *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp.344–350, Dec. 2015.
- [71] T. Matsukawa and E. Suzuki, “Person re-identification using CNN features learned from combination of attributes,” *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp.2428–2433, Dec. 2016.
- [72] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, “An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification,” *IEEE J. Biomed. Health Inform.*, vol.21, no.1, pp.31–40, Jan. 2017. Conference Name: IEEE Journal of Biomedical and Health Informatics.
- [73] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, and J. Liang, “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?,” *IEEE Trans. Med. Imag.*, vol.35, no.5, pp.1299–1312, May 2016. Conference Name: IEEE Transactions on Medical Imaging.
- [74] Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang, “Fine-Tuning Convolutional Neural Networks for Biomedical Image Analysis: Actively and Incrementally,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.4761–4772, July 2017. ISSN: 1063-6919.
- [75] X. Liu, C. Wang, J. Bai, and G. Liao, “Fine-tuning Pre-trained Convolutional Neural Networks for Gastric Precancerous Disease Classification on Magnification Narrow-band Imaging Images,” *Neurocomputing*, vol.392, pp.253–267, 2020, doi: <https://doi.org/10.1016/j.neucom.2018.10.100>.
- [76] G. Wang, W. Li, M.A. Zuluaga, R. Pratt, P.A. Patel, M. Aertsen, T. Doel, A.L. David, J. Deprest, S. Ourselin, and T. Vercauteren, “Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning,” *IEEE Trans. Med. Imag.*, vol.37, no.7, pp.1562–1573, July 2018. Conference Name: IEEE Transactions on Medical Imaging.
- [77] K. Gong, J. Guan, C.-C. Liu, and J. Qi, “PET Image Denoising Using a Deep Neural Network Through Fine Tuning,” *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol.3, no.2, pp.153–161, March 2019. Conference Name: IEEE Transactions on Radiation and Plasma Medical Sciences.
- [78] J.J. Cunningham, “Body composition as a determinant of energy expenditure: a synthetic review and a proposed general prediction equation,” *The American Journal of Clinical Nutrition*, vol.54, no.6, pp.963–969, Dec. 1991. Publisher: Oxford Academic.
- [79] Z. Liu and S. Sarkar, “Simplest representation yet for gait recognition: averaged silhouette,” *Proc. 17th International Conference on Pattern Recognition*, 2004. *ICPR 2004*, pp.211–214, vol.4, Aug. 2004. ISSN: 1051-4651.
- [80] S. Volpato, L. Bianchi, F. Lauretani, F. Lauretani, S. Bandinelli, J.M. Guralnik, G. Zuliani, and L. Ferrucci, “Role of Muscle Mass and Muscle Quality in the Association Between Diabetes and Gait Speed,” *Diabetes Care*, vol.35, no.8, pp.1672–1679, Aug. 2012. Publisher: American Diabetes Association Section: Original Research.
- [81] S.A. Ross and J.R. Engsberg, “Relationships Between Spasticity, Strength, Gait, and the GMFM-66 in Persons With Spastic Diplegia Cerebral Palsy,” *Archives of Physical Medicine and Rehabilitation*, vol.88, no.9, pp.1114–1120, Sept. 2007.
- [82] H. Nakao, T. Yoshikawa, T. Mimura, T. Hara, K. Nishimoto, and S. Fujimoto, “Influence of Lower-extremity Muscle Force, Muscle Mass and Asymmetry in Knee Extension Force on Gait Ability in Community-dwelling Elderly Women,” *Journal of Physical Therapy Science*, vol.18, no.1, pp.73–79, 2006.
- [83] M. Okumura, Y. Makihara, S. Nakamura, S. Morishima, and Y. Yagi, “The Online Gait Measurement for the Audience-Participant Digital Entertainment,” *Proc. Invited Workshop on Vision Based Human Modeling and Synthesis in Motion and Expression*, no.IW1-3, pp.1–10, Xi’an, China, Sept. 2009.
- [84] Y. Makihara, T. Kimura, F. Okura, I. Mitsugami, M. Niwa, C. Aoki, A. Suzuki, D. Muramatsu, and Y. Yagi, “Gait collector: An automatic gait data collection system in conjunction with an experience-based long-run exhibition,” *2016 International Conference on Biometrics (ICB)*, pp.1–8, June 2016.
- [85] “Laboratory for New Media 15th Exhibition “Let’s Walk! The first step for innovation” | Miraikan Permanent Exhibition.”
- [86] D. Zhou, Y. Dai, and H. Li, “Ground-plane-based absolute scale estimation for monocular visual odometry,” *IEEE Trans. Intell. Transp. Syst.*, vol.21, no.2, pp.791–802, 2020.





**Ruochen Liao** received an M.S. degree in information science and technology from Osaka University, Japan in 2014. He is currently pursuing a Ph.D. degree in computer science at Osaka University. His research interests are computer vision, healthcare, and the topic of gait analysis.



**Kousuke Moriwaki** received his M.S. degree in information science and technology from Osaka University, Japan in 2020. He is currently working at a Japanese technology company. His research interests are the topics of gait recognition and neural networks in computer vision.



**Yasushi Makihara** received his B.S., M.S., and Ph.D. degrees in Engineering from Osaka University in 2001, 2002, and 2005, respectively. He is currently a professor of the Institute of Scientific and Industrial Research, Osaka University. His research interests are computer vision, pattern recognition, and image processing, including gait recognition, pedestrian detection, morphing, and temporal super resolution. He is a member of IPSJ, IEICE, RSJ, and JSME.



**Daigo Muramatsu** received his B.S., M.E., and Ph.D. degrees in electrical, electronics, and computer engineering from Waseda University, Tokyo, Japan, in 1997, 1999, and 2006, respectively. He is currently a professor of Faculty of Science and Technology, Seikei University. His current research interests include gait recognition, signature verification, and biometric fusion. He is a member of the IEEE, IEICE, and the IPSJ.



**Noriko Takemura** received B.S., M.S., and Ph.D. degrees in engineering from Osaka University, Osaka, Japan, in 2006, 2007, and 2010, respectively. She is currently an associate professor at the Institute of Dataability Science, Osaka University. Her current research interests include gait recognition, ambient intelligence, and emotion estimation. Dr. Takemura is a member of the SICE, RSJ and VRSJ.



**Yasushi Yagi** is the Executive Vice President of Osaka University. He received his Ph.D. degree from Osaka University in 1991. In 1985, he joined the Product Development Laboratory, Mitsubishi Electric Corporation, where he worked on robotics and inspections. He became a Research Associate in 1990, a Lecturer in 1993, an Associate Professor in 1996, and a Professor in 2003 at Osaka University. He was also Director of the Institute of Scientific and Industrial Research, Osaka University from

2012 to 2015. International conferences for which he has served as Chair include 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 the IEEE ICRA Conference Editorial Board (2007–2011). He is an editorial member of IJCV, and the Editor-in-Chief of IPSJ Transactions on Computer Vision & Applications. He was awarded the ACM VRST2003 Honorable Mention Award, IEEE ROBIO2006 Finalist of T.J. Tan Best Paper in Robotics, IEEE ICRA2008 Finalist for Best Vision Paper, MIRU2008 Nagao Award, and the PSVIT2010 Best Paper Award. His research interests are computer vision, medical engineering, and robotics. He is a fellow of IPSJ and a member of IEICE, RSJ, and IEEE.