Attempt for Estimation of Vertical Ground Reaction Force by Deep Learning with Time Factor from 2D Walking Images

Takeshi Mochizuki
Kochi University of Technology
Tosayamada, Kami, Kochi, 782-8502, Japan
e-mail: mochizuki.takeshi0094@gmail.com

Kyoko Shibata
Kochi University of Technology
Tosayamada, Kami, Kochi, 782-8502, Japan
e-mail: shibata.kyoko@kochi-tech.ac.jp

Abstract: **Ground reaction force data are useful for evaluating gait stability, but only specialized institutions can measure it because installed force plates are often used to measure it with high accuracy. Therefore, this report proposes an easy method for estimating ground reaction forces using images captured by a widely available device. In a previous report, we created an algorithm to estimate the ground reaction force from images using 2D Convolutional Neural Network (CNN), one of the deep learning. The results showed that if a deep learning model is created in advance, the estimation of vertical ground reaction forces can have an 8% to 14% error to body weight. To further improve accuracy, this report creates training data that include the time factor and performs vertical ground reaction force estimation by 3D CNN. The training data used in this report, the voxel data were created using images at the time of estimation and images prior to that time to incorporate the time factor. The results, estimation of ground reaction force resulted in a 15% error relative to body weight and did not improve accuracy. Since overlearning occurred in all deep learning models, we suppose that accuracy was not improved due to insufficient training data or bias.**

Keywords- Gait Analysis; Ground Reaction Force; Estimation; 3D CNN; Single Camera.

I. INTRODUCTION

Gait exercise is important for maintaining and improving health. However, a gait style that places the load on one leg only or that tends to place the load on the ground is less effective. For good health, it is necessary to keep in mind that gait should be stable on both sides of the body on a daily basis. In clinical practice, gait stability is determined from the time history of the ground reaction force waveforms for each leg during walking, and physicians and physical therapists with expertise in this area provide guidance on gait improvement based on the waveforms. Therefore, we believe that it is possible to diagnose gait stability from the ground reaction force waveform, and if individuals can easily and at any time know the ground reaction force waveform while walking, they will be aware of the need to improve their gait, which will contribute to extending their healthy life span.

The method used to accurately measure ground reaction force waveforms consists of installed force plates. However, installed force plates are expensive, are only available in specialized facilities and cannot be installed at the individual level. This means that ground reaction force values cannot be obtained on a daily basis. In addition, walking movement becomes more deliberate because one must step on an installed force plate. This occurs the problem that normal gait cannot be measured [1].

As an alternative to an installed force plate, this research group has proposed a method to derive the ground reaction force from the acceleration obtained by a wearable inertial sensor using the balance between inertial force and ground reaction force, as reported by Isshiki et al. [2]. This method estimates the combined ground reaction force of the left and right legs, making it difficult to use in clinical settings where ground reaction force for each leg is desired. In addition, the method does not consider individual differences because it uses the mass of each body segment and the position of the center of gravity of each body segment calculated from statistical values as parameters used to derive the ground reaction force. To address this problem, Liu et al. [3] used individual kinematic data and deep learning to estimate ground reaction forces without using statistical values. They estimated ground reaction forces using angular data obtained from optical motion capture and deep learning and showed that they can be estimated at 2% to 8% error to body weight for stair walking. Since personal kinematics data is used, individual differences can be considered. However, because it uses optical motion capture, it cannot be used in everyday life. In contrast, Sakamoto et al. [4] reported an example of estimating ground reaction forces using only data obtained by wearable sensors and deep learning, without using kinematics. They estimated ground reaction forces in a standing static posture and reported that it can be estimated with an average estimation error of 7.6%. However, since the input data used were myopotential, acceleration, and angular acceleration acquired by wearable sensors, the system was not easy to wear and was not simple to use.

On the other hand, Yagi et al. [5] is an example of an attempt to analyze gait from videos that can be easily captured, although it is not an estimation of ground reaction forces. Gait analysis was performed using OpenPose [6], which detects skeletal information from videos and images, and was able to obtain stride length and walking speed from videos captured with an RGB camera. Using only a camera without a wearable sensor for sensing, gait analysis is achieved with fewer burdens on the user. However, as Yagi et al. also mentioned, the method using OpenPose causes errors in the skeletal information acquired from OpenPose, which also causes errors in the estimation results.

To address these issues, this study aims to establish an algorithm to estimate triaxial ground reaction force waveforms for each left and right legs using only cameras that are easy to sense. The proposed algorithm does not use statistics-based kinetic theory, dedicated software to detect skeletal information from videos and images, or other unfamiliar sensor systems, such as wearable sensors, when estimating ground reaction forces. Only videos and images will be used for ground reaction force estimation to eliminate the physical burden on the user during sensing.

As a methodology to achieve this, we have proposed a ground reaction force estimation method using a Convolutional Neural Network (CNN), which is a type of deep learning that excels in image classification, in our previous report [7]. The system creates in advance a deep learning model capable of estimating triaxial ground reaction forces in natural and abnormal walking on level ground using CNN from walking images captured by an RGB camera, so that the user only needs to capture walking images to perform the estimation.

In the previous report [7], the estimation of the ground reaction force from the load response phase to the front swing phase was performed using only walking images obtained from an RGB camera for detection. Accuracy was verified using cross-validation for five volunteers. The results showed that in a laboratory environment, vertical ground reaction forces can be estimated with an error of approximately 8% to 14% error to body weight and an average Pearson's correlation coefficient of 0.80. However, Dongwei Li et al. [3] and Sakamoto et al. [4], mentioned above, estimated it at 2 to 8% error to body weight, even though operating conditions were different. Therefore, in this report, we consider improving the accuracy of the proposed method by targeting 5%, which is a similar level of accuracy of estimation. In the previous report, we improved the estimation accuracy by converting color images to black-and-white images and reducing the image resolution to eliminate the influence of clothing color when creating training data. To further improve the accuracy of estimation, the training data created in the previous report does not include time factor, even though gait is a continuous motion. In this report, we consider learning time factor as well. As a first step of verification to improve accuracy, voxel data containing time factor are used as training data. In this report, vertical ground reaction forces are used.

In the next section, we describe how to create training data that includes time factor and how to create a deep learning model devised in this report. Section III presents the results of the vertical ground reaction forces estimated by the deep learning model created, Section IV discusses the reasons for the lack of improvement in accuracy, and Section V conclusions close the report.

II. METHOD

A. Deep Learning Model

When the user uses the system, he/she simply takes a walking video and inputs it to the system without prior preparation. To achieve this, a deep learning model must be created in advance. A deep learning model is a learning model that outputs ground reaction force values normalized by body weight when voxel data created from a walking image are input. The structure of the 3D CNN in the deep learning model consists of an input layer, followed by two convolution layers, a pooling layer, and a dropout layer to prevent overlearning. The process was repeated from the convolution layer to the dropout layer. After smoothing, it was passed through the fully connected layer one layer at a time, and then the output layer. The convolution layer uses the Relu function as the activation function, and the all-coupled layer uses the Softmax function.

B. Experimental Methods

An experiment was conducted to obtain data to be used for training and validation. The same experimental design as previously reported [7] was used to see the difference in accuracy of the training data generation method. In the experiment, 1 force plate unit (manufactured by Tec Gihan Co., Ltd., TF-6090-C 1 unit) was used for training data for deep learning models and an iPad Pro as a camera were used. Five healthy male volunteers (age 22±1, height 1.73±0.05 [m], weight 61±13 [kg]) participated in the experiment. A 10 step walk path was prepared, and the camera was placed 1.0 [m] from the floor, 3.5[m] from the center of the force plate, and perpendicular to the walking path. In order to collect training and validation data efficiently, videos were shot at 1080p HD/60fps and then converted to images at different frame rates. Participants were asked to walk as usual, and 50 trials were filmed during the sixth step, the stance phase of one gait, when the volunteers were walking normally and the left foot on the front side touched the force plate.

C. Deep Learning Models Creation Methods

Preprocessing is applied to the data obtained from experiments to create training data and validation data. The acquired walking videos are converted into images for each frame rate using the Python module OpenCV. Although the obtained image is a color image, it is converted to a monochrome image to reduce the influence of clothing color on the estimation and converted to 40 x 40 pixels by the bilinear interpolation method. Only the stance phase is extracted by checking the image and matching the time when the foot touches the force plate with the time when the ground reaction force value begins to output due to the foot touching the force plate. The ground reaction force values are normalized by the respective body weight to eliminate differences in values due to body weight and are set to true values. After that, 3D data for input to deep learning is created using the image at the time of the estimation and the images from four images before that time. The number of outputs of the deep learning model are 150, ranging from 0.01 to 1.50 in increments of 0.01. Of the data from the five volunteers, four are used as training data and one as validation data, and the training data is created so that all volunteers become validation data. The number of training and validation data is shown in Table 1 because the number of acquired images is different for each experimental collaborator. the structure of the 3D CNN is as described in Section 2-1, and the parameter values are shown in Table 2 after a trial-and-error process. The deep learning model is created using the deep learning library Keras with reference to the Keras Documentation [8]. EarlyStopping was used as censoring condition, the training error was used as the monitor, and auto was used as the mode. The deep learning model is created by training on the created training data and is terminated by the censoring condition, and estimation is performed on the validation data.

III. RESULTS

Fig. 1 shows the correlation between the estimated value for the 2,013 voxel data of the deep learning model III and the true value that was estimated most accurately. The values estimated using the deep learning model are shown on the horizontal axis as estimated values, and the values obtained using the force plate and normalized by body weight are shown on the vertical axis as true values. Some voxel data are estimated with good accuracy when the true value is larger than 0.70, but not when the value is smaller than 0.70.

Table III shows Pearson's correlation coefficients between the estimated and true values for each deep learning model, the average mean absolute error calculated by multiplying the body weight by the ground reaction force value [N], and the ratio of the mean absolute error to the body weight. The average Pearson's correlation coefficient for the deep learning model was 0.59, showing no improvement in accuracy. Even the deep learning model with the smallest mean absolute error had an error of 10% relative to body weight, and the average mean absolute error for all deep learning models was 15% error relative to body weight, with no improvement in accuracy.

IV. DISCUSSION

In this report, the target estimation accuracy was set at 5% of error to body weight, but the accuracy was 15% error to body weight, showing no improvement in accuracy. Fig. 2 shows the accuracy percentage of correct answers for the training data and the correct answers are shown on the vertical axis, and the accuracy percentage of correct answers for the validation data during the training of the deep learning model III. The percentage of number of epochs is shown on the horizontal axis. The graph shows that the rate of correct answers for the training data improves with each successive training, but the rate of correct answers for the validation data does not. This trend was observed for all deep learning models. This is thought to be caused by overlearning. There are two possible causes of overlearning. The first is the lack of training data, which is a sufficient cause since the current training data are only about 8,000 images each, and we expect improvement by increasing the training data through future experiments. However, to obtain training data from experimental data, a huge amount of experiments must be conducted, and it is difficult to increase training data from experiments because the burden on volunteers, time, and cost are too great. Furthermore, there is no publicly available data set that can be used. Second, there is a bias in the training data. Table 4 shows the percentage of training data per estimation interval for each deep learning model. For all deep learning models, the proportion of training data in the interval between 0.70 and 1.20 accounts for about 80% of the training data, indicating that the training data are biased. In Fig. 1, it can be read that the model is able to estimate in the range where the estimated value is greater than 0.70, but not in the range where the estimated value is smaller than 0.70. This suggests that the accuracy did not improve due to bias in the training data. On the basis of these results, we expect that the number and bias of the training data are problematic. Therefore, if the same amount of training data can be generated for all intervals, such as by expanding the data only

TABLE I NUMBER OF TRAINING AND VALIDATION DATA.

Deep learning models number	I	II	III	IV	V
Training data	B,C, D,E	A,C, D,E	A,B, D,E	A,B, C,E	A,B, C,D
Number of voxel data for the training data.	8252	8046	8255	8153	8366
Validation data	A	B	C	D	E
Number of voxel data for the validation data	2016	2222	2013	2115	1902

TABLE II 3D CNN LEARNING CONDITIONS.

		Set value
Convolution layer	Filter size	5×5×2
	Stride	1
	Channels	256
Pooling layer	Filter size	5×5×2
	Stride	1
Dropout		0.3
Fully connected layer		128
Batch size		100
Epoch		500

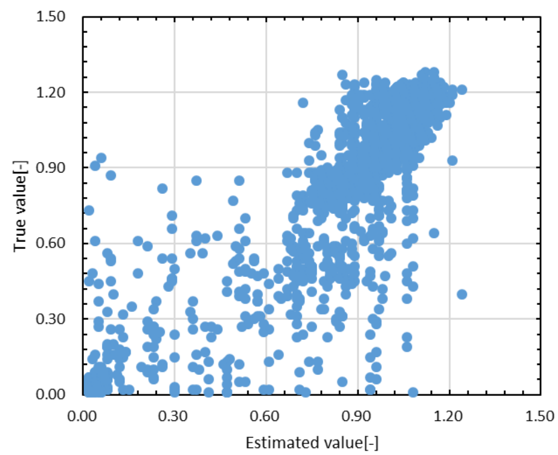

Fig.1 Normalized ground reaction force estimates versus true values.

TABLE III RESULTS FOR ALL DEEP LEARNING MODELS.

Deep learning models number	I	II	III	IV	V	Average
Pearson's correlation coefficient	0.65	0.78	0.85	0.02	0.63	0.59
Mean absolute error [N]	105.5	101.5	63.5	120.3	69.2	92.0
Mean absolute error for body weight [%]	16	14	10	19	15	15

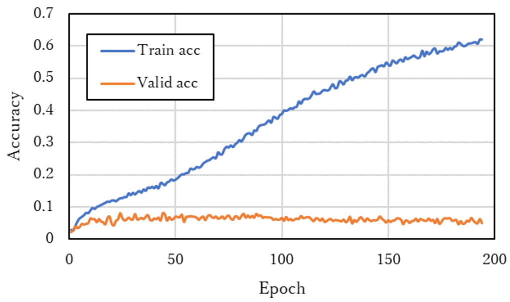

Fig.2 Accuracy rates of training and validation data for deep learning model III.

generated for all intervals, such as by expanding the data only for the intervals where the amount of data was small, the accuracy can be expected to improve.

V. CONCLUSION AND FUTURE WORK

In this report, we examined how to improve the accuracy of the ground reaction force estimation algorithm using only the RGB camera for sensing, which is the proposed method. The method of creating training data was changed, that is voxel data including images at the time of estimation and images up to four images before that time were created and used as training data. No improvement in accuracy was found. It is suggested that overlearning occurs during the training of any deep learning models. We suppose that the overlearning is due to the small amount of training data and bias. Therefore, creating a large amount of unbiased training data is expected to eliminate overlearning and improve accuracy. Another improvement is to incorporate a layer of recurrent neural network into the deep learning model used, in addition to CNN, so that time factor can be learned and accuracy can be improved.

In the future, our aim is to develop a system that can capture images and estimate three directions ground reaction forces using only a tablet device. If this is realized, it will be possible to evaluate gait on a daily by observing ground reaction force waveforms, which will support people to be aware of gait improvement and contribute to extending healthy life expectancy.

REFERENCES

[1] J.Perry, and J. M. Burnfield, GAIT ANALYSIS Normal and Pathological Function, Ishiyaku Publishers, Inc., pp. 243-249, 2007, (in Japanese)

[2] A. Isshiki, Y. Inoue, K. Shibata, and M. Sonobe, "Estimation of Floor Reaction Force During Walking Using Physical Inertial Force by Wireless Motion Sensor," HCI Int'l, vol. 714, pp. 249-254, May 2017, DOI: 10.1007/978-3-319-58753- 0_37, 2017, pp.22-33, ISSN:1348-711

[3] D. Liu, M. He, M. Hou, and Y. Ma, "Deep learning based ground reaction force estimation for stair walking using kinematic data, Measurement," volume 198, July 2022, 111344

[4] S. Sakamoto, D. Owaki, and M. Hayashibe, "Ground Reaction Force Estimation from EMG and IMU Sensor Using Recurrent Neural Network," The Japan Society of Mechanical Engineers Tohoku Branch 55th Annual Meeting and Lecture, Section ID: 108_paper, 2020, (in Japanese)

[5] K. Yagi, Y. Sugiura, K. Hasegawa, and H. Saito, "Gait Measurement at Home Using a Single RGB Camera," Gait&Posture Volume 76, February 2020, Pages 136-140

[6] OpenPose, https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/, 2023.10.13

[7] T. Mochizuki, and K. Shibata, "Estimation of Floor Reaction Forces by Convolutional Neural Network Using Walking Image without Depth Information : Evaluation of Generalization Ability," 2023 JSME Information, Intelligence and Precision Equipment Division, IIPB-4-12, 2023, (in Japanese)

[8] keras Documentation, https://keras.io, 2023.10.13

TABLE IV PERCENTAGE OF TRAINING DATA PER ESTIMATION INTERVAL.

Estimation interval	0.01~0.10	0.11~0.20	0.21~0.30	0.31~0.40	0.41~0.50	0.51~0.60
Deep learning model I	5.7%	2.2%	2.1%	2.3%	2.8%	2.9%
Deep learning model II	6.1%	2.0%	1.9%	2.1%	3.2%	3.0%
Deep learning model III	5.7%	2.1%	2.0%	2.2%	3.0%	2.7%
Deep learning model IV	6.3%	2.0%	1.7%	2.0%	3.0%	2.9%
Deep learning model V	6.0%	2.1%	2.2%	2.3%	3.4%	2.6%

0.61~0.70	0.71~0.80	0.81~0.90	0.91~1.00	1.01~1.10	1.11~1.20	1.21~1.30	1.31~1.40	1.41~1.50
3.3%	11.2%	26.4%	17.1%	11.1%	11.8%	1.0%	0.0%	0.0%
3.2%	11.6%	21.9%	16.0%	12.4%	15.3%	1.4%	0.0%	0.0%
3.4%	11.1%	25.9%	17.4%	11.8%	12.2%	0.6%	0.0%	0.0%
3.2%	12.6%	23.4%	14.6%	13.0%	14.3%	1.1%	0.0%	0.0%
2.7%	9.0%	26.7%	16.8%	11.0%	13.9%	1.3%	0.0%	0.0%