2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Vision-Based Gait Events Detection Using Deep Convolutional Neural Networks

Ankhzaya Jamsrandorj, Mau Dung Nguyen, Mina Park, Konki Sravan Kumar, Kyung-Ryoul Mun, Jinwook Kim

***Abstract*— Accurate gait events detection from the video would be a challenging problem. However, most vision-based methods for gait event detection highly rely on gait features that are estimated using gait silhouettes and human pose information for accurate gait data acquisition. This paper presented an accurate, multi-view approach with deep convolutional neural networks for efficient and practical gait event detection without requiring additional gait feature engineering. Especially, we aimed to detect gait events from frontal views as well as lateral views. We conducted the experiments with four different deep CNN models on our own dataset that includes three different walking actions from 11 healthy participants. Models took 9 subsequence frames stacking together as inputs, while outputs of models were probability vectors of gait events: *toe-off* and *heel-strike* for each frame. The deep CNN models trained only with video frames enabled to detect gait events with 93% or higher accuracy while the user is walking straight and walking around on both frontal and lateral views.**

I. INTRODUCTION

As a basic function of human mobility, gait can be defined as a unique and periodic manner of human walking [1]. Every individual has their inherent gait pattern caused by gender, age, body anthropometry, and psychological and pathological factors [2]. Therefore, accurate and reliable gait assessment is essential for appropriate clinical intervention. The gait assessment provides substantial clues in monitoring person's health status. It also helps evaluate diagnosis and prognosis for the people who have trauma, neurological diseases, musculoskeletal anomalies, and psychiatric disorder. Human gait strongly relies on temporal characteristics of an individual such as the step and stride, double limb support (DLS), single limb support (SLS) time, etc., and the gait temporal parameters can be segregated by detecting two main gait events, heel-strike (HS) and toe-off (TO), which are the moment the foot touch and leave the ground respectively [2].

For the HS and TO detection, many researchers have been used wearable sensor-based and vision-based methods. Wearable sensors such as inertial measurement units (IMU) and foot pressure sensors are widely used due to their high accuracy, and flexibility to handle compared to optoelectronic motion capture systems [3]–[10]. However, the wearable sensor-based methods require operation by trained professionals and high cooperation with the participants. Besides, wearing sensors on the human body may hinder a person's natural walk.

On the other hand, the vision-based methods rely only on video cameras for acquiring gait data from the participants without the aid of any other special sensors. However, most vision-based methods used for gait data acquisition focus on the classification of pathologies [11] and gait recognition for people identification [12]–[16], and only a few methods explore gait events detection.

The vision-based gait event detection mainly adopted appearance-based methods using silhouettes of a person to understand spatiotemporal changes between two or more subsequent frames while walking [17]–[21] and pose-based methods using depth images captured by depth-sensing cameras to acquire the person's skeletal structure and features such as leg length, normalized average stride length, and gait velocity [13], [22].

For the appearance-based gait event detection, Tang *et al*. proposed a new feature called consecutive silhouettes difference (CSD) maps by encoding several consecutive silhouettes to represent gait patterns, and detected TO events in the video using the CSD maps with a convolutional neural network [17]. Verlekar *et al* proposed a markerless 2D video-based system to estimate HS and TO events. Their system performs HS and TO events estimation in three steps. First, the system detects a time window that includes flat feet period where the person's feet are in complete contact with the floor by computing overlaps from the superimposed silhouettes. Then, it selects candidate frames of HS and TO events by analyzing the width of feet silhouettes. Finally, the system estimates HS and TO events using the frame numbers that include flat feet and the sets of candidate frames [21].

This type of method highly relies on human silhouettes that are extracted from original RGB frames. However, the silhouette extraction is affected by illumination changes and many silhouettes appear incomplete. Even when the extraction step is performed correctly, the shape of the silhouette can be easily altered depending on the view angles, the distance between the camera and human, and clothing conditions.

For the pose-based approach, Rocha *et al* proposed a system for fully automatic gait analysis based on a single RGB-D camera. The system recognized the three different

K.S. Kumar, K. Mun, and J. Kim are with the Center for Artificial Intelligence, Korea Institute of Science and Technology, Seoul, South Korea (corresponding author to provide phone: +82-2958-6785; fax: +82-2958-5769; e-mail: jwkim@imrc.kist.re.kr).

M. Park is with the Department of Brain & Cognitive Engineering, University of Korea, Seoul, South Korea.

A. Jamsrandorj and M.D. Nguyen are with the Department of Human Computer Interface & Robotics Engineering, University of Science & Technology, Daejeon, South Korea.

activities named walking, standing, and marching. Once the walking activity is detected, it estimates HS events by computing the distance between 3D ankles positions and the velocity of both ankles [22]. The researchers use pose-based methods to extract the invariant gait features for clothing and carrying conditions, but these methods fully depend on depth-sensing cameras or external pose machines to obtain human pose information.

As the limitations mentioned above on the vision-based methods, feature extraction processes are laborious, time-consuming, and resource-intensive for accurate gait data acquisition. Besides, little contrast among the consecutive silhouettes and skeletons when captured at frontal views makes the vision-based gait event detection methods only available in lateral view. Therefore, a non-contact method that is not interrupting the natural gait assessment pattern, as well as a view-invariant vision-based method, is essential for effective and feasible vision-based gait event detection.

Hence, this study aimed to present an efficient, practical, and multi-view approach for gait events detection using the raw RGB frames and deep convolutional neural networks. We mainly focused on detecting gait events from frontal views as well as lateral views.

II. MATERIALS AND METHODS

A. Experimental protocol to obtain the new dataset

This study is based on work supported by the Institutional Review Board of the Korea Institute of Science and Technology. All participants provided written informed consent and no violation was made during the experiment.

Eleven subjects (*8 males and 3 females; age = 24.2* ±3.8 *years, height = 170.7* ±6.4 *cm, and weight = 71.7* ±16.1 *kg*) without any presence or history of neurological disorders participated in this experiment. All the participants neither reported inconvenience in their daily walking nor had gait disturbance identifiable by naked eyes. As shown in Fig.1, the data was collected in an indoor environment from 6 different viewing angles: 3 frontal views and 3 lateral views. The capture area has a size of 3.68 meters in horizontal and 2.0 meters in vertical. The data from a subject is simultaneously captured by timely-synchronized 6 RGB cameras with a resolution of Full HD and a frame rate of 60Hz.

There are three different walking actions in the dataset: Walking straight (WS), Walking around (WA), and Walking on Treadmill (WoT). Subjects were asked to perform each action 2 times for a minute (30 seconds for WoT). The dataset details are summarized in Table I and in total, it includes 1.18M frames.

For WS action, each subject was requested to walk straight and turn back at the specific point of the area and repeat this during the given time, whereas the subjects walked around on the capture area without a given direction in WA action. For WS and WA actions, walking was performed at a self-selected pace, but in the WoT action, we set the constant speed for all subjects at 3.5m/s.

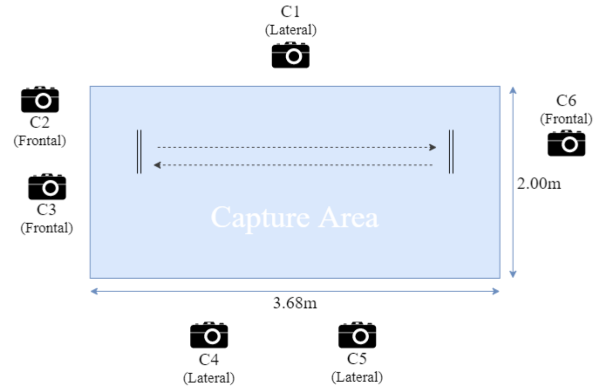

Fig. 1. Experimental setup with six RGB cameras.

Every video is equipped with an annotation, containing the frame numbers that are manually labeled in four gait events: *right heel-strike, right toe-off, left heel-strike, and left toe-off* and human bounding boxes on every frame.

B. Gait events detector development

1) Target construction for gait events detection: We manually labeled the frame numbers that contain HS and TO events of both legs for the whole dataset. Therefore, these gait events should be clearly defined. In the medical field, HS event is defined as the moment that foot touches the ground and TO event is defined as the moment that foot leaves the ground, shown as in Fig.2(a).

The gait event in the video is a micro-event that is labeled at one frame. Because neighboring frames of each HS and TO event have similar pixel contents but should have different labels, we needed to determine them to avoid making confusion to our networks. Therefore, we smooth our raw labels to produce probability distributions fitted around the frames in which gait events occur. In other words, the target is four one-dimensional Gaussian distribution curves

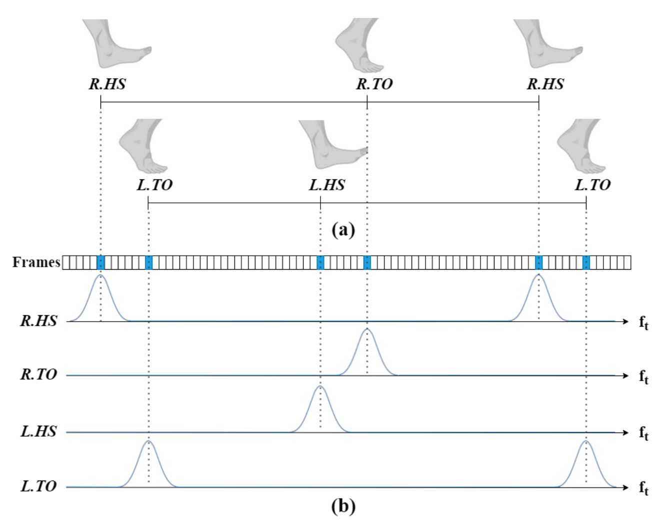

Fig. 2. (a) Definitions used to annotate gait events on video frames and (b) Target construction for gait events detection.

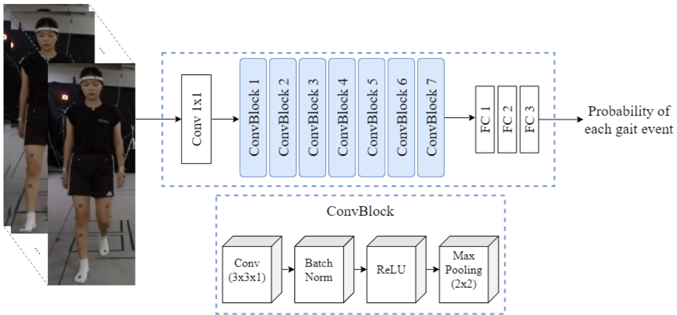

Fig. 3. The architecture of the CNN model with seven Convolutional Blocks.

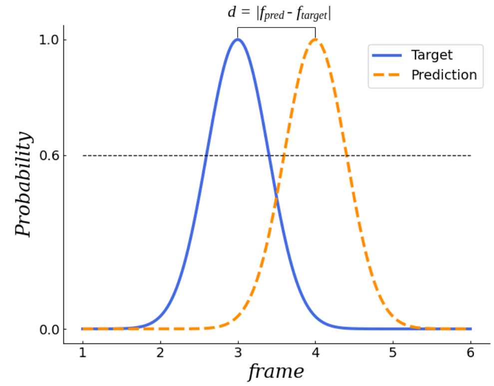

Fig. 4. Graphical demonstration of frame-error that used to calculate the precision of the prediction. The frame difference between the target frame of gait events and the prediction frame is noted as d. To discretize the probability distribution into precise frame numbers that include gait events, the peak values were picked at a threshold of 0.6.

(probability vectors) with means associated with the four gait events, respectively (see Fig.2(b)).

2) Network Architectures: The architecture of the CNN model used in this experiment is depicted in Fig.3. The input of the model is a stack of 9 subsequent image frames and output is the probability of each gait event for each frame. There is the 1×1 Convolutional layer, followed by seven Convolutional blocks wired sequentially and each Convolutional block consists of a 3×3 convolutional layer, batch-normalization, Rectified Linear Unit (ReLU) function, and max-pooling layer with 2×2 size and 2×2 stride. On top, we used three fully connected (FC) layers and the last FC layer has thirty-six neurons corresponding to nine input frames by four gait events [23]. For the activation function, we adopted the ReLU function except for the Sigmoid function at the output layer. We also ran experiments with three well-known models named ResNet18, ResNet50, and DenseNet121 to compare performances. For these three networks, the last FC layer was removed and added a new FC layer to adapt to our output.

TABLE I
DATASET SUMMARY FOR GAIT EVENTS DETECTION

No of Subjects	Action Name	Time (seconds)	Frame Rate	No of Views	Total Frames
11 (male 8 and female 3)	Walking Straight Walking Around Waking On Treadmill	60 x 2 60 x 2 30 x 2	60fps	6	475200 475200 237600
	Dataset size				**1188000**

C. Implementation details

1) Data processing: The input tensors were constructed from 9 subsequent frames with the ground truth human bounding boxes and each event probability target for all nine frames. We preprocessed raw frames by cropping the person in the frame based on the ground truth human bounding box, then pad left and right to get a square image and resizing it to a size of 448×448. The preprocessed frames were stacked together as input to four networks. The dataset was separated by subjects into three parts: 5 subjects for the training set (432k frames), 3 subjects for the validation set (259.2k frames), and the rest subjects for the test set (259.2k frames). We used data from WS and WA actions for the training, validation, and testing phase in the experiments and the WoT action was used only for an additional test.

2) Training scheme: We implemented the experiments using PyTorch. The weights of the simple CNN model were initialized from scratch while initializing three well-known models by pretraining on the ImageNet for gait events detection. We trained the four networks with Mean Squared Error (MSE) loss using adaptive moment estimation (Adam) optimization with momentum 0.9 and a batch size of 16. The initial learning rate was set to 0.0001 and applied a learning rate schedule by reducing the learning rate 2 times after every 3 epochs.

D. Evaluation metrics

The target for gait events detection was designed as probability distributions for each gait event and the detection was assessed with two metrics. The first, A Smooth Percentage Correct Events (**SPCE**) was adopted [23]. It calculates differences between target and prediction probabilities for every frame and an event was considered as correct if this probability difference is less than a threshold of 0.25. The second metric is Average Precision (**AP**) which calculates the prediction's accuracy. As shown in Fig.4, for each probability distribution, first, the peak values were picked at the threshold of 0.6 to precise the frame numbers that contain gait events, and then calculated the frame difference between the target frame number and the prediction frame number is noted as d. An event is treated as a correct event if the frame difference is less than a threshold of 4 frames in the experiments.

III. RESULTS

We trained the four different models and tested the performances of the models separately on frontal and lateral views data. The example distributions of the probabilities for each gait event were illustrated in Fig.5 and Fig.6. It has been

TABLE II
THE PERFORMANCES OF FOUR CNN MODELS ON OUR TEST SET FOR WS ACTION.

Models	Pretrain	Frontal		Side	
		SPCE	*AP*	*SPCE*	*AP*
ResNet18	ImageNet	0.90	0.91	0.90	0.93
ResNet50	ImageNet	0.89	0.90	0.90	0.93
DenseNet121	ImageNet	0.90	0.91	0.91	0.92
Simple CNN	None	0.88	0.88	0.90	0.91

TABLE III
THE PERFORMANCES OF FOUR CNN MODELS ON OUR TEST SET FOR WA ACTION.

Models	Pretrain	Frontal		Side	
		SPCE	*AP*	*SPCE*	*AP*
ResNet18	ImageNet	0.93	0.95	0.94	0.99
ResNet50	ImageNet	0.93	0.94	0.93	0.96
DenseNet121	ImageNet	0.93	0.94	0.94	0.98
Simple CNN	None	0.92	0.93	0.93	0.95

demonstrated that models can detect the precise frames in which HS and TO events occur.

Table II and Table III summarized the performances of the deep CNN models on WS and WA actions, respectively. We used a simple deep CNN model for gait events detection. Even though the model learned from scratch without pre-trained weights, the detection performances were near with performances of well-known three deep CNN models.

As shown in Table II and Table III, the ResNet18 model achieved the best performances on our test set. For WS action, the average precisions (AP) of the ResNet18 model were 0.91 and 0.93 from frontal and lateral views, whereas for WA action, APs were 0.95 and 0.99 on the data captured under the frontal and lateral views, respectively. We also analyzed performances of TO and HS events detection separately, and for all deep CNN models, performances of the TO event detection were higher than detection performances of the HS event. As shown in Table VI, models achieved APs of 0.95 and 0.90 on right and left TO events detection on average, whereas detection performances were APs of 0.88 and 0.88 on right and left HS events, respectively for WS action. Table VII summarized TO and HS events detection performances of deep CNN models for WA action and APs were 0.97 and 0.98 on right and left TO events detection on average, whereas APs were 0.95 and 0.96 on right and left HS events detection on average for four deep networks, respectively. The highest values of SPCE were obtained by ResNet18 and DenseNet121 on both frontal and lateral views for all actions.

We also tested the performances of the models on the WoT action that was not used during training. As shown in Table IV, all models achieved more than APs of 0.95, which means that models can detect gait events from unseen data in the training phase.

IV. DISCUSSION

This study proposed a novel vision-based gait event detection method using deep CNN models which works accurately and effectively in both lateral and frontal views of the mobile cameras respectively. It is worth noting that the deep CNN models trained with 9 sequential video frames achieved promising performances regardless of the camera view angles. Additionally, it should also be noted that this method does not require a lot of laborious experimental setup and protocols, time-consuming preprocessing and resource-intensive processes, and empirical knowledge for gait feature engineering.

Models acquired more than APs of 0.90 except the simple CNN model on WS action, and all models detected gait events with the APs of 0.93 and 0.95 or higher on frontal and lateral views for WA action, respectively. We investigated the reason that the performances on WS action are lower than the performances on WA action. For our dataset, WS action consists of two different stages: *walking straight stage* and *turning back stage* at each end of the straight corridor. We tried to segregate these two stages and check the model performance separately, and the result was shown in Table V. When subjects walk straight, the detection was excellent, whereas some of detections were failed during the turning back motion. We observed that the HS and TO motions at turning-back stage were very different from those at straight walking. Unlike a typical gait pattern, it starts from toe-strike and ends with toe-off or heel-off. In addition, lack of the number of data, intra-variability among the subjects as well as self-occlusions between the left and right limbs during the turning-back stage may hinder the accurate gait event detection with the deep CNN models. Besides, as shown in Table VI and Table VII, we can conclude that performances of HS events detection were lower than detection performances of TO events at turning-back motion. Despite relatively low detection accuracy during turning-back, our approach still can be applied for gait analysis during walking straight or walking around motions, and we will enhance our approach to detect gait events at turning back motion in future works.

TABLE IV
APS OF FOUR CNN MODELS ON OUR TEST SET FOR WOT ACTION.

Models	Pretrain	AP
ResNet18	ImageNet	0.98
ResNet50	ImageNet	0.96
DenseNet121	ImageNet	0.95
Simple CNN	None	0.98

As shown in Table IV, we also tested four models on WoT action that was not used during training. The detection performance was evaluated in both frontal and backside views (see Fig.1). For WoT action, only one camera was in front of a treadmill and the other two cameras captured walking actions from the backside. From the results, it has been confirmed that all models can detect gait events

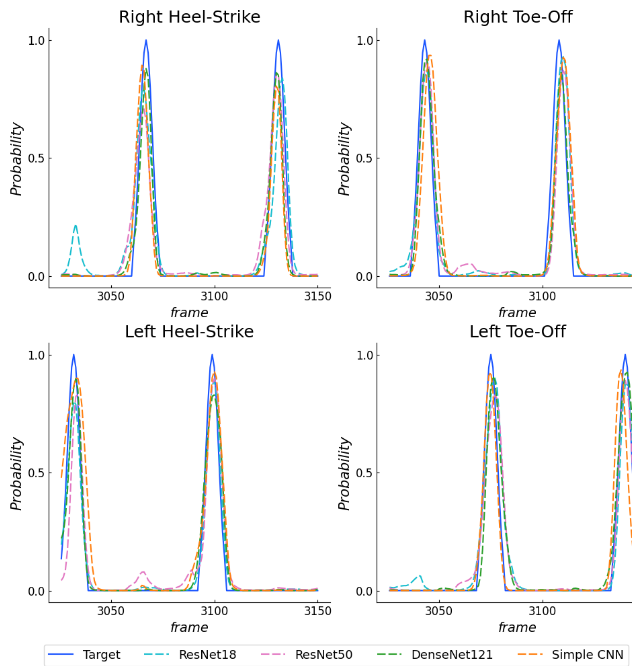

Fig. 5. The probability distributions of gait events: WS action

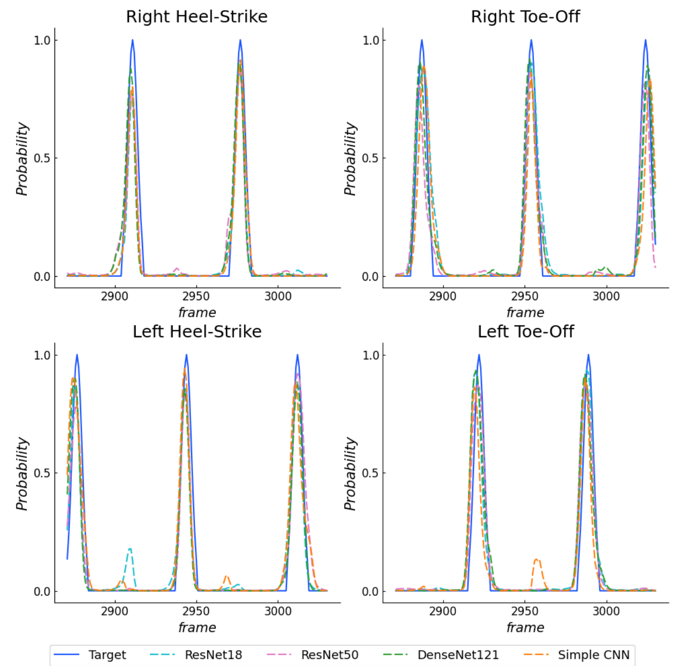

Fig. 6. The probability distributions of gait events: WA action

accurately from the backside as well as frontal and lateral views.

Several previous studies focused on the frontal views data to detect HS and TO events. Nieto-Hidalgo *et al* designed a vision-based gait analysis system that enabled the detection of HS and TO events using only frontal view gait silhouettes. Because in the silhouettes of frontal views, toes are always visible, but heels are occluded. Therefore, the system found the locations of the left and right foot toe by finding the minimum component of each half of the silhouette. Based on both toe locations, it detected HS and TO events, and achieved the detection accuracy of 89.1% on their own dataset for normal gaits [19]. On the other hand, Xu *et al* evaluated the accuracy of the Microsoft Kinect for the measurement of gait parameters during a treadmill walking in frontal view. By comparing positions of human lower limb joints obtained with the Kinect sensor, HS and TO events were detected and showed similar results to [19] also for normal gaits [24].

In comparison to these studies, we achieved outstanding accuracy on detecting HS and TO events from frontal views, which also can detect gait events accurately from lateral views.

We expect that our study can serve effective, practical, and feasible gait event detection method, and also for the researches on the relationships between an individual's gait and intrinsic, physical, psychological, and pathological characteristics.

This study has a limitation related to generalizability. The participants in this study were young and healthy, and most of them had normal gaits. It may affect detection performances when models detect gait events from a person who has an abnormal gait. To overcome this limitation, we plan to increase the sample size, types of subjects of our dataset and improve our approach to detect the gait events from various subjects such as the elderly, patients who have abnormal gait parameters.

V. CONCLUSIONS

In this study, we presented the accurate, multi-view approach for fully automatic gait events detection using deep convolutional neural networks without requiring a special camera and feature engineering process. The deep CNN models trained with only few sequential video frames enabled reliable detection of gait events from both frontal and lateral views. Our approach has the potential to be applied for gait-based health monitoring both at home and in a clinical setting. Future studies with various subject-groups such as the elderly and patients with abnormal gaits will be conducted to generalize the findings of this study.

TABLE V

APS OF FOUR CNN MODELS AT DIFFERENT MOTIONS.

Models	**Pretrain**	**AP**	
		Straight Walking Stage	***Turning Back Stage***
ResNet18	ImageNet	0.97	0.60
ResNet50	ImageNet	0.96	0.58
DenseNet121	ImageNet	0.96	0.66
Simple CNN	None	0.93	0.51

TABLE VI
COMPARISON OF DETECTION RESULTS ON TO AND HS EVENTS FOR WS ACTION

Models	AP			
	RTO	***RHS***	***LTO***	***LHS***
ResNet18	0.97	0.90	0.92	0.89
ResNet50	0.96	0.89	0.91	0.90
DenseNet121	0.95	0.89	0.90	0.89
Simple CNN	0.93	0.85	0.89	0.83

TABLE VII
COMPARISON OF DETECTION RESULTS ON TO AND HS EVENTS FOR WA ACTION

Models	AP			
	RTO	***RHS***	***LTO***	***LHS***
ResNet18	0.98	0.96	0.99	0.97
ResNet50	0.97	0.94	0.99	0.95
DenseNet121	0.97	0.95	0.98	0.96
Simple CNN	0.97	0.94	0.97	0.95

ACKNOWLEDGMENT

This research was supported in part by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project No. 1711139131) and in part by the Korea Institute of Science and Technology Institutional Program (Project No. 2E31051).

REFERENCES

[1] K.-R. Mun, G. Song, S. Chun, and J. Kim, "Gait Estimation from Anatomical Foot Parameters Measured by a Foot Feature Measurement System using a Deep Neural Network Model," *Scientific Reports*, vol. 8, no. 1, p. 9879, 2018. [Online]. Available: https://doi.org/10.1038/s41598-018-28222-2

[2] D. Jung, M. D. Nguyen, J. Han, M. Park, K. Lee, S. Yoo, J. Kim, and K. R. Mun, "Deep Neural Network-Based Gait Classification Using Wearable Inertial Sensor Data," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 3624–3628.

[3] J. Jellish, J. Abbas, T. Ingalls, P. Mahant, J. Samanta, M. Ospina, and N. Krishnamurthi, "A System for Real-Time Feedback to Improve Gait and Posture in Parkinson's Disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, p. 1, aug 2015.

[4] P. Aqueveque, E. Germany Morrison, R. Osorio, and F. Pastene, "Gait Segmentation Method Using a Plantar Pressure Measurement System with Custom-Made Capacitive Sensors," *Sensors*, vol. 20, p. 656, jan 2020.

[5] N. Zahradka, K. Verma, A. Behboodi, B. Bodt, H. Wright, and S. Lee, "An Evaluation of Three Kinematic Methods for Gait Event Detection Compared to the Kinetic-Based 'Gold Standard'," *Sensors*, vol. 20, p. 5272, sep 2020.

[7] M. Lempereur, F. Rousseau, O. Rémy-Néris, C. Pons, L. Houx, G. Quellec, and S. Brochard, "A new deep learning-based method for the detection of gait events in children with gait disorders: Proof-of-concept and concurrent validity," *Journal of*

[6] J. C. Pérez-Ibarra, A. A. G. Siqueira, and H. I. Krebs, "Identification of Gait Events in Healthy and Parkinson's Disease Subjects Using Inertial Sensors: A Supervised Learning Approach," *IEEE Sensors Journal*, vol. 20, no. 24, pp. 14 984–14 993, 2020.
Biomechanics, vol. 98, p. 109490, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021929019307407

[8] S. Sahoo, M. Saboo, D. K. Pratihar, and S. Mukhopadhyay, "Real-Time Detection of Actual and Early Gait Events During Level-Ground and Ramp Walking," *IEEE Sensors Journal*, vol. 20, no. 14, pp. 8128–8136, 2020.

[9] L. C. Benson, C. A. Clermont, R. Watari, T. Exley, and R. Ferber, "Automated Accelerometer-Based Gait Event Detection During Multiple Running Conditions," 2019.

[10] T. Miyake, Y. Kobayashi, M. Fujie, and S. Sugano, "Gait event detection based on inter-joint coordination using only angular information," *Advanced Robotics*, vol. 34, pp. 1–11, aug 2020.

[11] M. Nieto-Hidalgo, F. J. Ferrández-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual, and J. M. García-Chamizo, "A vision based proposal for classification of normal and abnormal gait using RGB camera," *Journal of Biomedical Informatics*, vol. 63, pp. 82–89, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046416300806

[12] T. Verlekar, P. Correia, and L. Soares, "View-Invariant Gait Recognition System Using a Gait Energy Image Decomposition Method," *IET Biometrics*, vol. 6, mar 2017.

[13] R. Liao, C. Cao, E. Garcia, S. Yu, and Y. Huang, *Pose-Based Temporal-Spatial Network (PTSN) for Gait Recognition with Carrying and Clothing Variations*, oct 2017.

[14] H. Chao, Y. He, J. Zhang, and J. Feng, "GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 8126–8133, jul 2019.

[15] J. P. Singh, S. Jain, S. Arora, and U. P. Singh, "Vision-Based Gait Recognition: A Survey," *IEEE Access*, vol. 6, pp. 70 497–70 527, 2018.

[16] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "GEINet: View-invariant gait recognition using a convolutional neural network," in *2016 International Conference on Biometrics (ICB)*, 2016, pp. 1–8.

[17] Y. Tang, Z. Li, H. Tian, J. Ding, and B. Lin, "Detecting Toe-Off Events Utilizing a Vision-Based Method," *Entropy*, vol. 21, apr 2019.

[18] M. Nieto-Hidalgo, J. Ferrandez, R. Valdivieso-Sarabia, J. Mora-Pascual, and J. García-Chamizo, *Vision Based Extraction of Dynamic Gait Features Focused on Feet Movement Using RGB Camera*, dec 2015.

[19] ——, *Vision Based Gait Analysis for Frontal View Gait Sequences Using RGB Camera*, nov 2016.

[20] C. Yang, U. Ugbolue, A. Kerr, V. Stankovic, L. Stankovic, B. Carse, K. Kaliarntas, and P. Rowe, "Autonomous Gait Event Detection with Portable Single-Camera Gait Kinematics Analysis System," *Journal of Sensors*, vol. 2016, pp. 1–8, jan 2016.

[21] T. T. Verlekar, H. De Vroey, K. Claeys, H. Hallez, L. D. Soares, and P. L. Correia, "Estimation and validation of temporal gait features using a markerless 2D video system," *Computer Methods and Programs in Biomedicine*, vol. 175, pp. 45–51, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016926071831040X

[22] A. P. Rocha, H. M. P. Choupina, M. d. C. Vilas-Boas, J. M. Fernandes, and J. P. S. Cunha, "System for automatic gait analysis based on a single RGB-D camera," *PLOS ONE*, vol. 13, no. 8, p. e0201728, aug 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0201728

[23] R. Voeikov, N. Falaleev, and R. Baikulov, *TTNet: Real-time temporal and spatial video analysis of table tennis*, apr 2020.

[24] X. Xu, R. McGorry, L.-S. Chou, J.-H. Lin, and C.-C. Chang, "Accuracy of the Microsoft KinectTM for measuring gait parameters during treadmill walking," *Gait & Posture*, vol. 34, may 2015.