

# Gait Classification of Common Pedestrians and Smartphone Zombies Using Micro-Doppler Radar

Kazuki Yasuda<sup>1</sup>, Teppei Tsuyuhara<sup>1</sup>, Masao Masugi<sup>1</sup> and Kenshi Saho<sup>1,2</sup>

<sup>1</sup> Ritsumeikan University, 1-1-1 Noji-Higashi, Kusatsu, Shiga, Japan

<sup>2</sup> Toyama Prefectural University, 5180 Kurokawa, Imizu, Toyama, Japan

## Abstract

In this study, a micro-Doppler radar-based method for classifying gaits with and without smartphone texting while walking (i.e., the classification of normal pedestrians and smartphone zombies) is proposed. The motion features in the gait of the participants texting on a smartphone while walking were collected using the micro-Doppler radar as time-frequency distributions (spectrograms), which showed the velocity variation of body parts with respect to time. We adopted a method of inputting spectrogram images obtained from a short-time Fourier transform of the received signals of the Doppler radar to a convolutional neural network. Experimental results showed that the proposed method achieved accurate classification with over 80 % accuracy.

## Keywords

Micro-Doppler radar, gait measurement, gait recognition, texting while walking

## 1. Introduction

With the increasing use of smartphones, the number of people texting on smartphones while walking (called "smartphone zombie") has increased, which has led to many accidents, such as collisions and falling. Various approaches have been studied to detect the action of people using smartphones while walking to avoid such accidents and to develop warning and collision avoidance systems. For example, a method that uses built-in smartphone sensors has been proposed [1]. The remote detection of smartphone zombies is a promising approach, and methods for remote detection of smartphone zombies using cameras and image processing techniques have been proposed [2, 3]. Another approach using lidar has also been proposed to recognize pedestrian types [4]. However, the detection accuracies of these optical sensor-based techniques depend on the lighting conditions and subjects' clothes. In addition, the use of a camera has privacy issues.

In contrast, radar techniques can measure motion information without these problems. In particular, the micro-Doppler radar technique is a promising candidate for accurate pedestrian sensing because of its capability to remotely measure the time variation in the velocities of human body parts [5, 6]. The micro-Doppler radar achieves various types of accurate human motion recognition, such as gesture recognition [7], human identification [8], and fall detection [9]. However, micro-Doppler radar applications for detecting the action of texting while walking have not been reported.

In this paper, we present a method for classifying the gaits of common pedestrians and smartphone zombies using micro-Doppler radar data. The micro-Doppler radar data of the gaits of participants walking normally and participants on a smartphone while walking were collected and analyzed. Then, an accuracy of more than 80 % was demonstrated in classifications by inputting a convolutional neural network (CNN) to time-frequency distribution (spectrogram) images calculated from radar-received signals. We investigated the classification accuracies using representative architectures of CNNs, such as VGGNet, ResNet, and AlexNet, and revealed an efficient CNN model for micro-Doppler radar-based detection of smartphone zombies.

The 4th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2022), August 24-26, 2022, Ibaraki, Japan

EMAIL: ri0095xv@ed.ritsumei.ac.jp (K. Yasuda); masugi@fc.ritsumei.ac.jp (M. Masugi); saho@pu-toyama.ac.jp (K. Saho)

ORCID: 0000-0003-2088-1231 (K. Saho)

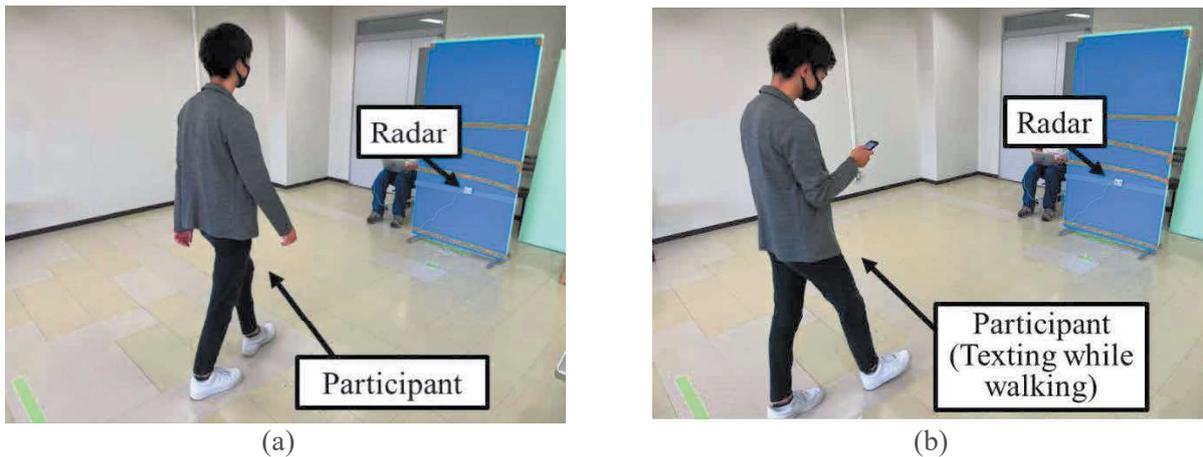


© 2022 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

## 2. Methods

### 2.1. Experimental protocol

Figure 1 shows the experimental conditions of the radar. A 24 GHz continuous-wave (CW) radar (BSS-110, ILT Office, Toyama, Japan) was placed in front of the participant. The radar was a monostatic micro-Doppler radar installed at a height of 60 cm, and the -3 dB beamwidths in the vertical and horizontal planes of the antenna were  $\pm 14^\circ$  and  $\pm 35^\circ$ , respectively. The participants were instructed to walk towards the radar along an 8-m straight walkway. The radar signals received at a sampling frequency of 600 Hz were obtained after demodulation using a quadrature detector. We used the acquired time series of in-phase and quadrature components as complex signals. The participants were 18 adults, 14 men and 4 women with ages and heights of  $21.5 \pm 0.8$  years and  $168.9 \pm 11.5$  cm, respectively. Seventeen participants performed the following two types of walking motions 15 times each and one participant performed them 3 times each: (a) Common walking (Figure 1 (a)), where participants walked normally at a self-selected comfortable pace, and (b) texting on a smartphone while walking (smartphone zombie) (Figure 1 (b)), where the participants walked following the instruction "please walk while reading texts on your smartphone." The aim of this study is to classify the radar data of walking motion types (a) and (b).



**Figure 1:** Experimental situations. (a) common walking and (b) texting while walking (smartphone zombie).

### 2.2. Dataset generation

We then generated images of the time-frequency distribution (spectrogram images) of the received signals using a short-time Fourier transform (STFT), similar to our previous study [8]. Before the STFT process, we used a Butterworth high-pass filter on the received signals to remove the zero-frequency components corresponding to the static targets. We then calculated the STFT of the received signal as

$$S(t, f) = \int s(\tau)w(\tau - t)e^{-j2\pi f\tau}d\tau, \quad (1)$$

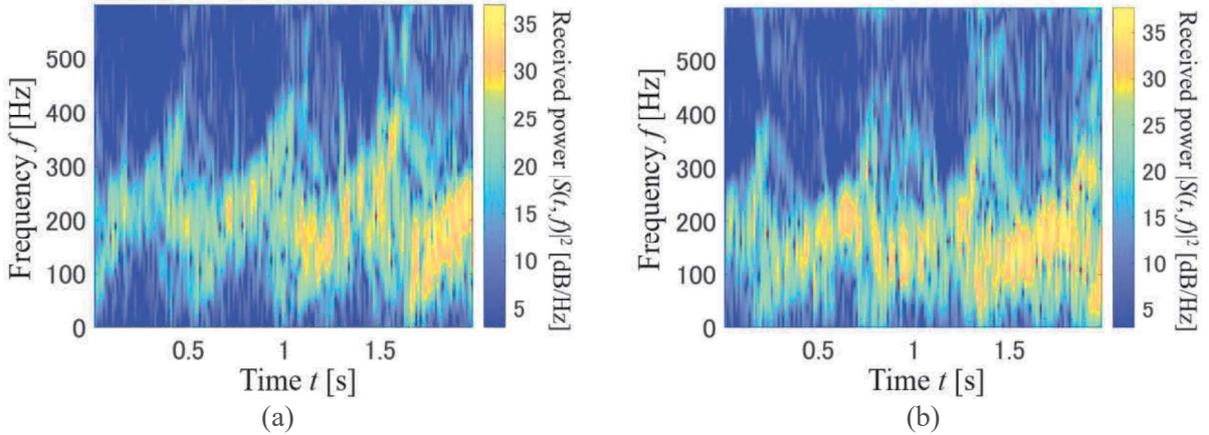
where  $s(t)$  is the high-pass filtered received signal,  $w(t)$  is a window function, and  $f$  is the Doppler frequency corresponding to target velocity  $v$ . The relationship between  $f$  and  $v$  is

$$f = \frac{2}{\lambda}v, \quad (2)$$

where  $\lambda$  denotes the wavelength of the transmitted signal. Because this study used a 24 GHz continuous wave,  $\lambda$  was 12.5 mm. For  $w(t)$ , we empirically used the commonly used Hamming window function

with a length of 21.7 ms. We calculated the spectrogram as  $|S(t, f)|^2$ . Consequently, we removed the components whose received powers were smaller than 3 dB/Hz as noise in  $|S(t, f)|^2$ , trimmed the spectrogram corresponding to steady-state walking, and converted the trimmed spectrograms into RGB-colored PNG images of size  $164 \times 218$ . We used the generated  $(17 \text{ (participants)} \times 15 \text{ (times of walking)} + 1 \text{ (participant)} \times 3 \text{ (times of walking)}) \times 2 \text{ (types of motion)} = 516$  PNG images as the dataset for this study.

Figure 2 shows examples of the generated spectrogram images of walking motion types (a) and (b). The frequencies of the spectrogram correspond to the target velocity, components in the spectrogram images with larger received power correspond to the motion of the torso, and accompanying relatively larger frequency variations correspond to the forward motion of the legs. Although motion type (b) (smartphone zombie) tended to have a slightly lower frequency compared with motion type (a) (common walking), clear differences between the two types could not be confirmed in the spectrogram images. This is because the essential differences between these two types of walking motion were less visible in the spectrograms. The simplest difference between these types was the presence of an arm-swinging motion. However, the arm motions were not clearly confirmed in the spectrograms because the received powers of the echoes from the arms were quite small compared to those from the torso and legs [10]. Additionally, the velocities of the arms were similar to those of the legs. Thus, the components corresponding to the motions of the arms were mixed with background noise and components corresponding to leg motions. Another difference between the two types of walking motion was the position of the head. However, acquisition of this head (or neck) position difference was difficult for the Doppler radar because it measured only velocity information, and position-related information was slightly included in the spectrograms (as very slight differences in the received powers corresponding to the head and neck). Thus, the spectrogram images of common pedestrians and smartphone zombies were not clear.



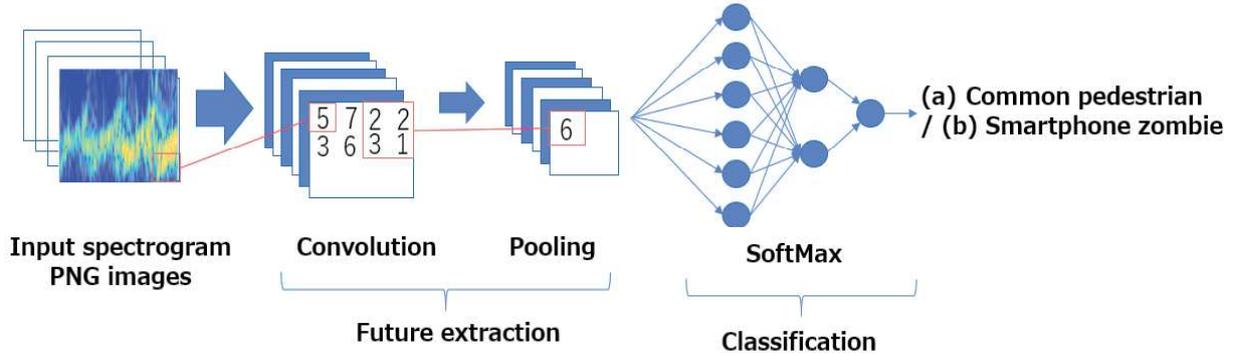
**Figure 2:** Examples of spectrogram images. (a) common walking and (b) smartphone zombie.

### 2.3. Gait classification using CNN and accuracy evaluation

To extract and use the slight differences between the spectrogram images of the two groups, we used a CNN for feature extraction and gait classification. Figure 3 shows an outline of the proposed classification method. The CNN was used as the deep-learning method for classification, similar to previous studies on Doppler radar-based personal identification [5, 6]. The generated spectrogram images were used as input data for the CNN. We investigated the classification performance of CNNs with representative architectures of LeNet [11], AlexNet [12], ResNet-18 [13], and VGGNet [14] because they are effective for radar-based human motion recognition problems. In addition, ResNet-50 [15] was considered to investigate the effectiveness of the deeper network. The hyperparameters of each network were empirically optimized. The examples of the hyperparameters for AlexNet are as follows: the loss function was a cross-entropy function and the stochastic gradient descent with

momentum optimization algorithm was used for the modeling. We trained for 50 epochs and used a batch size of 64. The learning rate was 0.01.

To evaluate the classification accuracy, we performed a hold-out validation. In each architecture, the CNN was trained using 80 % of the generated images, and the remaining 20 % were used as test data. Ten hold-out validation trials were conducted by randomly varying the training and test data. The mean and standard deviation of the classification accuracies across all trials were calculated.



**Figure 3:** Outline of the classification method using CNN. Input: spectrogram images. Output: classified walking type.

### 3. Results and Discussion

Table 1 presents the test results for the classification of walking motion types (a) and (b) using various CNN architectures. LeNet, AlexNet, and ResNet-50 achieved accurate classification with over 80 % accuracy. The highest accuracy of 85.7 % was achieved using AlexNet. The accuracy of ResNet-50 was relatively stable, with a smaller standard deviation. These results indicate that the CNN with spectrogram images accurately classified the gaits of the common pedestrian and smartphone zombie groups, even though their differences in the images were minor. The results for LeNet and AlexNet show that a relatively simple and lower-layered network can sufficiently capture the gait features of smartphone zombies. However, relatively deeper ResNet-50 results were stable, which implies that such a network may be effective when the data of a larger number of participants and training samples are used; this should be an important future direction of this study.

The effectiveness of the proposed method using AlexNet, which achieved the highest accuracy, is discussed using the confusion matrix and convergence curve. Table 2 presents the confusion matrix for the results using AlexNet and indicates that there are some cases in which the use of the smartphone did not change the speed of the torso and leg motions; therefore, these data were not correctly classified. However, a sensitivity of 84 % and a specificity of 86 % were achieved. Thus, the proposed method can accurately screen for smartphone zombies. Figure 4 shows an example of the convergence curves of the proposed method using AlexNet and ResNet-50. For both networks, the accuracy converged to approximately 50 epochs during the test process, and overfitting was not confirmed. Thus, the proposed method effectively trained the classification model. Although AlexNet achieved a better accuracy result, the steady state accuracy of ResNet-50 was stable compared to that of AlexNet, and these results indicate a tendency similar to that in Table 1, which indicates the possibility of the effectiveness of a deeper network.

Finally, we discuss the mechanism of accurate gait classification in our results. As shown in Section 2.2, the differences between the images of the two walking types were not clear because the motion differences in the arms and head were not sufficiently measured. However, our results indicated that a classification with an accuracy of more than 80 % was achieved using the spectrogram images that mainly reflected the motions of the torso and legs. It appears that the proposed method extracted features corresponding to slight differences in such motions using the CNN. A biomechanical study clarified

slight reductions in gait speed and changes in the tibialis anterior and gastrocnemius when texting on a smartphone while walking for young adults [16]. These apparently affect the motion of the legs and torso. Thus, it can be considered that the features corresponding to these differences could be efficiently learned from the spectrogram images using the CNN in the proposed method.

**Table 1**

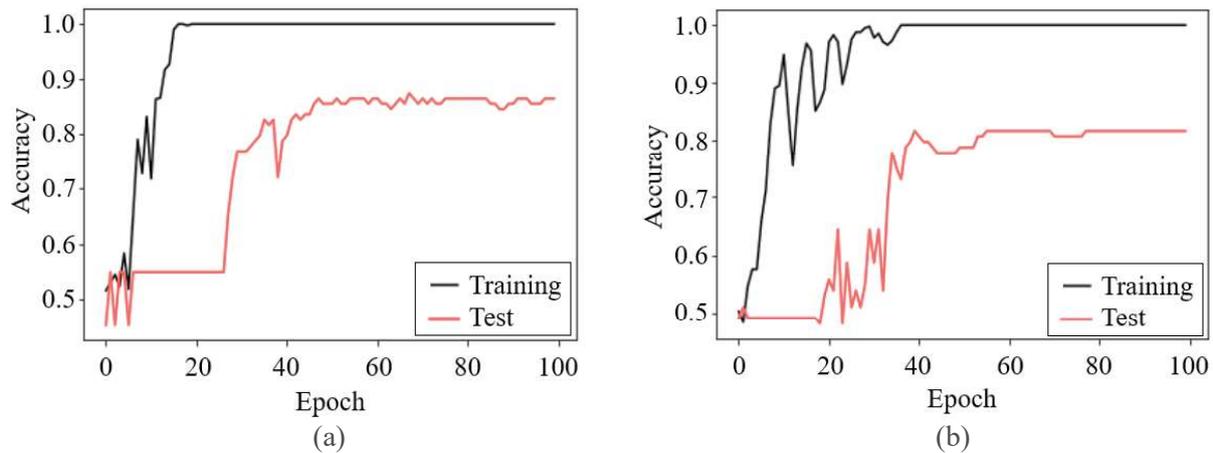
Classification results.

CNN architecture	Mean classification accuracy among 10 tests [%]	Standard deviation of classification accuracy among 10 tests [%]
LeNet	80.11	4.67
AlexNet	85.72	5.64
ResNet-18	79.11	5.88
ResNet-50	80.42	2.98
VGG-16	74.71	4.22

**Table 2**

Confusion matrix for the classification using the AlexNet.

Predicted\True	Walking type (a) (Common walking)	Walking type (b) (Texting while walking)
Walking type (a)	85.7 %	15.8 %
Walking type (b)	14.3 %	84.2 %



**Figure 4:** Examples of convergence curves of the proposed method. (a) AlexNet, (b) ResNet-50.

## 4. Conclusion

In this study, the gaits of common pedestrians and smartphone zombies were measured using the micro-Doppler radar, and images of spectrograms (time-frequency distributions) of the radar received signals were used for classification. By applying a CNN using the image spectrograms as the input data, a classification accuracy of 86% was achieved using the experimental data. Thus, the effectiveness of the proposed method for remotely detecting the action of texting on a smartphone while walking was verified.

In our experiments, all participants were approximately 20 years old and were measured in the same experimental environment. Therefore, the future direction of our study is to measure various types of

participants of different ages and in different environments. In addition, experiments on multiple pedestrian situations are important to demonstrate the practicality of the proposed method.

## Acknowledgements

We appreciate all the participants for their engagement in this study.

## Reference

- [1] T. Wada and A. Shikishima, "Real-time detection system for smartphone zombie based on machine learning." *IEICE Commun. Express*, 9.7(2020): 268-273.
- [2] H. Hanaizumi, H. Misono, An OpenPose based method to detect texting while walking, in: *Proc. ICISIP2019*, Howard International House, Taipei, Taiwan, 2019, pp. 130-134.
- [3] A. Rangesh and M.M Trivedi. "When vehicles see pedestrians with phones: A multicue framework for recognizing phone-based activities of pedestrians." *IEEE Transactions on Intelligent Vehicles* 3.2(2018): 218-227.
- [4] J. Wu et al. "Smartphone zombie detection from lidar point cloud for mobile robot safety." *IEEE Robot. Automat. Lett.* 5.2(2020): 2256-2263.
- [5] H. Arab, I. Ghaffari, L. Chioukh, S.O. Tatu, and S. Dufour. "A Convolutional neural network for human motion recognition and classification using a millimeter-wave Doppler radar." *IEEE Sens. J.* 22.5 (2022): 449404502.
- [6] S. Z. Gurbuz and M. G. Amin. "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring." *IEEE Signal Process. Mag.* 36.4(2019): 16-28.
- [7] Z. Yu, D. Zhang, Z. Wang, Q. Han, B. Guo, and Q. Wang. "SoDar: Multitarget gesture recognition based on SIMO Doppler radar." *IEEE Trans. Hum. Mach.* 52.2(2022): 276-289.
- [8] K. Saho, K. Shioiri, and K. Inuzuka. "Accurate person identification based on combined sit-to-stand and stand-to-sit movements measured using Doppler radars." *IEEE Sens. J.* 21.4(2020): 4563-4570.
- [9] H. Sadreazami, M. Bolic, and S. Rajan. "Contactless fall detection using time-frequency analysis and convolutional neural networks." *IEEE Trans. Indust. Inform.* 17.10(2021): 6842-6851.
- [10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." *Proc. of the IEEE* 86.11(1998): 2278-2324.
- [11] K. Saho, T. Sakamoto, T. Sato, K. Inoue, and T. Fukuda, Experimental study of real-time human imaging using UWB Doppler radar interferometry, In: *2012 6th European Conference on Antennas and Propagation (EUCAP)*, 2012, pp. 3495-3499.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks", in: *Proceeding of the 25th International Conference on Neural Information Processing Systems*, 2012, pp.1097-1105.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, pp.770-778.
- [14] S. Karen and A. Zisserman. "Very deep convolutional networks for large-scale image recognition.", *arXiv preprint arXiv : 1409.1556* (2014)
- [15] R. U. Khan, X. Zhang, R. Kumar, and E. O. Aboagye, Evaluating the performance of ResNet model based on image recognition, in: *Proc. Int. Conf. Comput. Artif. Intell. (ICCAI)*, 2018, pp. 86-90.
- [16] V. Agostini, F. Lo Fermo, G. Massazza, and M. Knaflitz. "Does texting while walking really affect gait in young adults?" *J. Neuroeng. Rehabil.* 12.1(2015): 1-10.