CNN-Based Recognition Algorithm for Four Classes of Roads

Sung-Min Cho¹ and Byung-Jae Choi²

¹Department of Rehabilitation Industry, Daegu University, Gyeongsan, Korea

²Department of Electronic and Electric Engineering, Daegu University, Gyeongsan, Korea

ljfis

Abstract

In recent years, location-based augmented reality games have become popular globally. Consequently, the risk of collisions or accidents while walking with mobile devices has increased. Using smartphones while walking can distract pedestrians and can lead to negative consequences for traffic safety. In addition, a survey of visually impaired people revealed that they found border recognition inconvenient due to the lowered jaws between the driveway and sidewalks. In this study, an accident prevention system is proposed based on a convolutional neural network by segregating the walking environments into four classes (sidewalks, driveways, crosswalks, and braille blocks). A total of 3,200 datasets (3,000 for training and 200 for test) were used in our study. We show that the proposed system has the accuracy of 90% for validation data, and the recognition rate of 90% or above for test data.

Keywords: CNN, Image recognition, Walking environment, Cautionary dispersion

1. Introduction

Mobile devices, including smartphones, have made our life easier. We spend a considerable amount of time using them. However, the use of smart mobile devices has led to a noticeable increment in the number of accidents owing to distraction caused by these devices [1]. Furthermore, in recent years, location-based augmented reality mobile games have become very popular globally. This increases the risk of a person being involved in a crash or an accident while walking with their mobile devices [2]. There are also other problems experienced by pedestrians who are visually impaired. According to them, it is difficult to recognize the boundary between sidewalks and driveways [3].

The purpose of this study is to investigate four different classes of frequently used roads, such as sidewalks, driveways, crosswalks, and braille blocks. A model based on convolutional neural network (CNN) is proposed to prevent distraction-related accidents caused by the use of mobile devices, and to provide better pedestrian infrastructure for people who are visually impaired. CNN is a neural network that excels in a number of applications based primarily on image recognition [4]. A neural network is advantageous to image learning [5].

Section 2 describes the dataset acquisition and simulation environment. Section 3 presents the CNN-based neural network design. Section 4 displays several results of the computer simulation. Finally, Section 5 presents the concluding remarks.

1.1 Related Works

Received: Apr. 4, 2020
 Revised : May 28, 2020
 Accepted: May 28, 2020

Correspondence to: Byung-Jae Choi
 (bjchoi@daegu.ac.kr)
 ©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 1. Dataset samples.

In related works, Phubber [6] introduced safety applications, which exhibited an accuracy of 74.19% with low computing resources of six classes (driveway, sidewalk, blind road, stairs, manhole covers, and car barrier) [7]. The perception of a stop line and crosswalk in the neighboring environment of the road increased the accuracy to 80% [8]. In [9], a support vector machine (SVM)-based application was introduced, in which they used RGBD (red, green, blue, and depth) cameras to detect staircases and crosswalks for the blind. This application displayed an average recognition accuracy of 93.7% [9].

2. Simulation Environment and Dataset

2.1 Simulation Environment

The simulation environments used in this study are as follows: CPU i5-3230M, 8 GB RAM, Windows OS, Python (Anaconda 4.7.11), TensorFlow 1.14.0 framework, and OpenCV 4.1.0.25 imaging processing.

2.2 Image Preprocessing

Datasets were captured through the rear camera of a Galaxy 9S+ smartphone. After filming the video, a total of 3,200 copies of 3,000 training data points (750 per class) and 200 test data points (50 per class) were obtained through frame extraction. The extracted images consist of a colored square image of size (1440) * length. Figure 1 presents the sample images.

2.3 Image Preprocessing

In this study, only a CPU is used for simulation of the experiment. The image size of the dataset is reshaped to $50 \times 50 \times 1$ (grayscale) OpenCV and stored as a varnish file with the NumPy format (.npy).

Figure 2. Design of neural network.

3. CNN-based Design of Artificial Neural Networks

The neural network design consists of five convolutional layers and two fully connected layers. Five convolutional layers perform the convolutional calculation and max-pooling using the activation function ReLU (Rectified Linear Unit). Figure 2 depicts the design of the proposed model based on CNN.

In the fully-connected layer, dropout prevents overfitting and a softmax function is used to separate the four classes in the last layer.

3.1 Conv (convolutional)

Conv (convolutional) is an operator that multiplies two functions by a reversal of the moving values and integrates the section to obtain a new function. For images, it is a two-dimensional plane (height, width) and consists of pixels. When one function moves to another function, the weight of the filter is updated through an operation that calculates the sum of multiplication for each element. Fourier transforms and Laplace transforms are closely related and are often used in signal processing [4].

It can be represented by Σ . Typically, pixel images (i, j) expressed in matrices are calculated by the convolution of the original image f and filter g . The formula for this is as follows Eq. (1).

$$(f * g)(i, j) = \sum_{x=0}^{h-1} \sum_{y=0}^{w-1} f(x, y)g(i - x, j - y). \quad (1)$$

However, the Conv layer uses cross-correlation instead of convolution because, for operation, the filter must be inverted prior to application. As Conv aims to learn the values of filters, there is no difference when cross-correlation is applied. However, filters should be constant in the learning and reasoning phases. For this reason, the TensorFlow and other deep learning frameworks are implemented in cross-correlations rather than

synthetic ones [4]. The formula is as follows Eq. (2).

$$(f * g)(i, j) = \sum_{x=0}^{h-1} \sum_{y=0}^{w-1} f(x, y)g(i + x, j + y). \quad (2)$$

3.2 ReLU

In the deep learning network, the input values for the node are not directly passed to the next layer. Instead, they pass through nonlinear functions. The function used at this time is called the activation function. In this study, the ReLU function was used. ReLU is a function that shows the output as it is if the input exceeds 0, but displays 0 if it is less than 0. It is widely used because it solves the problem of vanishing gradient in the existing sigmoid function [4]. The formula is as follows Eq. (3).

$$y = \begin{cases} x & (x > 0), \\ 0 & (x \leq 0), \end{cases} \quad \max(0, x). \quad (3)$$

Although it may vary depending on the architecture of existing neural networks, faster learning has been observed in neural networks using ReLU than neural networks with extreme activation values [5].

3.3 Filter

In the Conv function, the receptive field is called a filter or kernel. The filter corresponds to the weight parameter in Conv, and it is trained to find an appropriate filter in the learning phase. By applying the input data filter, it outputs the feature map that emphasizes the area of the image similar to the filter and delivers it to the next layer [4].

3.4 Max Pooling

There are two types of pooling: maximum value pooling and average value pooling. However, in the field of image recognition, the maximum value pooling is primarily used. The result of the pooling does not change well because it shows that the same value is obtained as output even when the input data shift to the right by one space [4].

3.5 Dropout

Dropout is a method in which learning occurs by the arbitrary deletion of neurons. During the learning process, neurons in the hidden layer are randomly selected and deleted to prevent overfitting [4].

3.6 Softmax

Softmax normalizes all the entered values that fall between 0 and 1, and the sum of the output values is always a function of “1”. The class with the largest value is printed to classify the class.

In Eq. (4), $\exp(x)$ is an exponential function that means e^x , n stands for the number of neurons in the output layer and y_k represents the k -th output of these neurons. The molecule of the function consists of an exponential function of the input signal a_k , and the denominator consists of the sum of the exponential function of all input signals.

$$y_k = \frac{\exp(a_k)}{\sum_{i=1}^n \exp(a_i)}. \quad (4)$$

4. Simulation Results

The learning results are as follows: training steps, 2,350; epoch, 50; learning rate, 1e-3; learning time, 7 minutes 31 seconds; recognition accuracy, 98.25%; and recognition accuracy (validation), 88%.

4.1 Learning Graph

In Figure 3, the horizontal axis represents the training steps (0-2,400) and the vertical axis represents the learning recognition rate (0-1). In the 800 training-step updates, the recognition rate is higher than 90% and a slight increment and decrement has been observed in the data since then.

In Figure 4, the recognition rate of the validation data was higher than 90% at the 600th training-step update. This value was attained faster than the learning progress of training data. However, it has been unstable since then.

Figure 3. Learning recognition accuracy graph by training step (training data).

Figure 4. Learning recognition accuracy graph by training step (validation data).

Table 1. Recognition rate

	Recognition rate (%)
Total	90
Sidewalk	90
Driveway	96
Crossway	96
Braille block	78

4.2 Recognition Result

The total recognition rate of the test data is higher than 90%. The detailed results are presented in Table 1.

The recognition rate for both driveways and crosswalks is 96%, which is very high. In addition, the recognition rates for sidewalks and braille blocks are 90% and 78%, respectively. Braille blocks have a somewhat lower recognition rate than other environments. This seems to be because they are composed of more complicated patterns than other environments.

After processing by the neural network, the test data are classified with the labels (walk, drive, cross, and block), which are placed on top of the image. Figure 5 presents 12 images of the test data that were recognized by the proposed system.

Figure 6 presents the unrecognized images. It seems that recognition failure is due to external environmental factors, such as excessive angle, sunlight, shadow, etc.

5. Concluding Remarks

With the development of advanced mobile technology, the use of smart devices has increased significantly while walking, thereby increasing the risk of accidents. In this study, we presented the results of a few methods that recognize four types of walking environments: roads, sidewalks, pedestrian crossings, and braille blocks. It is very useful to prevent accidents and distractions

Figure 5. Recognized images.

Figure 6. Unrecognized images.

caused by carelessness of smart device users or visually impaired persons while walking. A CNN-based neural network technique was used for recognizing the four types of walking environments. In previous similar studies, the recognition rates for the walking environment were “73.19%” and “80%”. However, the proposed method showed a high recognition rate of “over 90%” using a test of 200 datasets. Increasing the dataset will further improve the recognition rate, which is left as a future work. Development of the proposed recognition system can provide pedestrians with information, such as display, warning sounds, and vibrations, thereby exhibiting potential to be a competitive product in the market.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This research was supported by the KIAT (Korea Institute for Advancement of Technology) grant funded by the Korea Government (MOTIE: Ministry of Trade Industry and Energy). (No. N0001792, HRD program for Rehabilitation Industry).

References

- [1] S. C. Kang, S. W. Lee, and J. I. Sim, "A study on patterns and distraction of smart devices usages while walking," *Journal of Transport Research*, vol. 23, no. 2, pp. 27-39, 2016.
- [2] S. Yoshiki, H. Tatsumi, K. Tsutsumi, T. Miyazaki, and T. Fujiki, "Effects of smartphone use on behavior while walking," *Urban and Regional Planning Review*, vol. 4, pp. 138-150, 2017. <https://doi.org/10.14398/urpr.4.138>
- [3] Korea Ministry of Land, Infrastructure, and Transport, "Transportation and traffic facilities of the person with disability survey research," *Final Report*, 2018.
- [4] S. Goki, *Deep Learning from Scratch*. Seoul, Korea: Hanbit Media, 2017.
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing System*, vol. 25, pp. 1106-1114, 2012.
- [6] NAVER, "Phubber," Available <http://endic.naver.com/userEntry.nhn?sLn=kr&entryId=6e8e89153f258364b56861e379676d72>
- [7] C. Sun, J. Su, Z. Shi, and Y. Guan, "P-Minder: a CNN based sidewalk segmentation approach for phubber safety applications," in *Proceedings of 2019 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, 2019, pp. 4160-4164. <https://doi.org/10.1109/ICIP.2019.8803417>
- [8] J. H. Lee and H. Yoon, "Design and implementation of the stop line and crosswalk recognition algorithm for autonomous UGV," *Journal of Korean Institute of Intelligent Systems*, vol. 24, no. 3, pp. 271-278, 2014. <https://doi.org/10.5391/JKIIS.2014.24.3.271>
- [9] S. Wang and Y. Tian, "Detecting stairs and pedestrian crosswalks for the blind by RGBD camera," in *Proceedings of 2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops*, Philadelphia, PA, 2012, pp. 732-739. <https://doi.org/10.1109/BIBMW.2012.6470227>
- [10] N. Chauhan and B. J. Choi, "Performance analysis of classification techniques of human brain MRI images," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 19, no. 4, pp. 315-322, 2019. <https://doi.org/10.5391/IJFIS.2019.19.4.315>
- [11] W. S. Jeon and S. Y. Rhee, "Plant leaf recognition using a convolution neural network," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 17, no. 1, pp. 26-34, 2017. <https://doi.org/10.5391/IJFIS.2017.17.1.26>

Sung-Min Cho received his Bachelor of Business Administration degree at Kyungil University, Korea in 2017. He is a graduate student at Daegu University majoring in the Rehabilitation industry. His research interests include rehabilitation,

neural network, and Image processing.

E-mail: getbusy@daegu.ac.kr

Byung-Jae Choi received his B.S. degree in Electronic Engineering, in 1987 from Kyungpook National University, Korea. He received his M.S. and Ph.D. degrees in Electrical and Electronic Engineering, in 1989 and 1998, respectively, at Korea

Advanced Institute of Science and Technology in Korea. He is a Professor of School of Electronic and Electrical Engineering, Daegu University, Korea, since 1999. His current research interests include intelligent control and its applications.

E-mail: bjchoi@daegu.ac.kr