Article

A CNN-Based Wearable Assistive System for Visually Impaired People Walking Outdoors

I-Hsuan Hsieh ¹, Hsiao-Chu Cheng ¹, Hao-Hsiang Ke ¹, Hsiang-Chieh Chen ² and Wen-June Wang ^{1,3,*}

¹ Department of Electrical Engineering, National Central University, Taoyuan 320317, Taiwan; rrich922@g.ncu.edu.tw (I.-H.H.); a108521126@g.ncu.edu.tw (H.-C.C.); kevinke@ncu.edu.tw (H.-H.K.)

² Department of Electrical Engineering, National United University, Miaoli 360301, Taiwan; chc@nuu.edu.tw

³ Pervasive Artificial Intelligence Research Labs, MOST, Hsing-Chu 30010, Taiwan

* Correspondence: wjwang@ee.ncu.edu.tw

Abstract: In this study, we propose an assistive system for helping visually impaired people walk outdoors. This assistive system contains an embedded system—Jetson AGX Xavier (manufacture by Nvidia in Santa Clara, CA, USA) and a binocular depth camera—ZED 2 (manufacture by Stereolabs in San Francisco, CA, USA). Based on the CNN neural network FAST-SCNN and the depth map obtained by the ZED 2, the image of the environment in front of the visually impaired user is split into seven equal divisions. A walkability confidence value for each division is computed, and a voice prompt is played to guide the user toward the most appropriate direction such that the visually impaired user can navigate a safe path on the sidewalk, avoid any obstacles, or walk on the crosswalk safely. Furthermore, the obstacle in front of the user is identified by the network YOLOv5s proposed by Jocher, G. et al. Finally, we provided the proposed assistive system to a visually impaired person and experimented around an MRT station in Taiwan. The visually impaired person indicated that the proposed system indeed helped him feel safer when walking outdoors. The experiment also verified that the system could effectively guide the visually impaired person walking safely on the sidewalk and crosswalks.

Keywords: wearable device; visually impaired people; deep learning; semantic segmentation; depth map; obstacle avoidance

Citation: Hsieh, I.-H.; Cheng, H.-C.; Ke, H.-H.; Chen, H.-C.; Wang, W.-J. A CNN-Based Wearable Assistive System for Visually Impaired People Walking Outdoors. *Appl. Sci.* **2021**, *11*, 10026. <https://doi.org/10.3390/app112110026>

Academic Editor: Giuseppe Andreoni

Received: 10 August 2021

Accepted: 22 October 2021

Published: 26 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Scholars estimate that the number of visually impaired people worldwide will increase from 38.5 million in 2020 to more than 115 million by 2050 [1]. So increased societal and governmental attention will be needed in the future. Specifically, the visually impaired need assistive tools when they walk outdoors. White canes and guide dogs are currently the most well-known assistive tools for the visually impaired [2]. Although white canes are cheap and easy to use, they cannot provide important visual information, such as obstacle location, type, and proximity. Visual information is indispensable for environmental perception and movement safety during outdoor navigation [3]. Guide dogs can assist the visually impaired to avoid obstacles; however, their life span is about 8 to 12 years [4], and the cost of breeding and training a guide dog is very high. Many researchers have developed various assistive devices for the visually impaired [5–7]. To safely guide the visually impaired while walking, the assistive tools should have the ability to recognize the surrounding environment quickly and accurately.

A wearable device for visually impaired people was developed by [5] in which the device used different types of sounds to inform users whether there were obstacles in front of them or not. It also used different frequencies and decibels to indicate the location of the obstacles. In the paper [6], the authors combined wearable glasses and augmented reality technology and then integrated the traversable direction visual enhancement function to help partially sighted people to walk safely. In addition, they developed three-voice

prompts to guide severely or profoundly visually impaired people in a safe direction. A wearable navigation device [7] with an RGB-D camera, a gyroscope, and a smartphone were proposed to guide the user in bypassing obstacles when the visually impaired travel through indoor and outdoor environments.

When visually impaired people walk outdoors, obstacle avoidance, safe intersection transition, and accurate route navigation are of great importance. Many assistive frameworks have been developed to achieve these goals, which rely on smartphones, sensors, or computer vision. The authors of the paper [8] reviewed several related articles and analyzed the advantages and disadvantages of the different techniques. In the paper [9], the authors used the images captured by RGBD camera to train a semantic segmentation network and a classifier such that the visually impaired can recognize the road environment and walk toward their correct direction. However, there is no guidance if it is a dead-end in front of the user. Yang et al. [10] used smart glasses and a laptop with a real-time semantic segmentation network to lead the user. The system proposed by [11], composed of a pair of smart glasses and a portable processor, can recognize guide bricks and crosswalks for visually impaired users in outdoor environments. Paper [12] provided a segmentation network to recognize crossroads and used an image processing method of color space to identify traffic lights and help the visually impaired use the crosswalk safely. The study of paper [13] used a smartphone to help the visually impaired avoid aerial obstacles. For the visually impaired, a set of glasses with integrated infrared (IR) transceiver sensors and a white cane with accelerometer and gyroscope sensors [14] are used to avoid obstacles and detect whether the user falls. In the paper [15], the authors proposed a sensor-based wearable device for the walking assistance of visually impaired people. An algorithm based on image depth information and fuzzy logic was proposed in [16] to detect multiple objects in real-time, and it achieved high accuracy on obstacle avoidance. The inertial measurement unit (IMU) and the smartphone camera were utilized in [17] to design a navigation system for assisting the visually impaired during indoor and outdoor activities. Aladrén et al. [18] used floor-segmentation and road-segmentation to identify safe indoor walking paths for visually impaired people. Paper [19] developed a navigation assistant for the visually impaired to effectively avoid obstacles when walking outdoors using the CNN-based object detection, tracking, and recognition methods. However, it does not guide the visually impaired user to walk on either the sidewalk or the crosswalks. In the paper [20], a visual localizer, composed of a ConvNet descriptor and a global optimizer, was used to achieve robust visual localization to assist visually impaired people's navigation. To guide the visually impaired through the intersection, [21] introduced a wearable system with a CNN-based object detection model to recognize crosswalks. In addition, it uses iAPS (intelligent, accessible pedestrian signal) mechanisms to receive the traffic status.

So far, developing an improved assistive system to help the visually impaired walk safely is still an active topic for current society. This article presents a wearable assistive system that can safely help visually impaired people walk on sidewalks and crosswalks.

The organization of this article is as follows. Section 2 introduces the materials and methods, including the main objective, system configuration, and the main functions required by the visually impaired in the proposed assistive system. This section has two neural network-based models to be trained and used in the walking guide strategy for the visually impaired. Section 3 presents the practical experiment by a visually impaired user. Section 4, there are some discussions about the experiments. At last, the conclusions and future works are drawn in the final section.

2. Materials and Methods

2.1. The Main Method

This study designs and implements an assistive system for the visually impaired to walk on the sidewalk and crosswalk without hitting obstacles safely. In more detail, this assistive system can help the visually impaired recognize the surrounding environment,

such as sidewalk, crosswalk, and obstacles, and then guides the visually impaired user to walk in a safe and correct direction.

2.1.1. The Hardware System Configuration

The whole assistive system is wearable and contains an embedded device—Jetson AGX Xavier launched by Nvidia [22], a binocular camera named ZED 2. Additionally, a voice prompt manager is installed in the AGX. This work connects a sound card between the AGX and the headphone to produce voice prompts. The AGX featuring a Tegra System on Chip (Xavier SoC) comprises 8 NVIDIA Carmel CPU Cores (architecture compatible with ARMv8.2) and an integrated GPU based on the NVIDIA Volta architecture with 512 CUDA cores. In addition, it also adds a Tensor Core, an NVIDIA deep learning accelerator, a VLIW (Very long instruction word) video processor, and an image signal processor (ISP) to enhance AI computing performance. The ZED 2 is a stereoscopic camera designed for tracking vehicles and people with a vast field of view. It offers a more excellent depth detection with a 120-degree wide-angle field of view. These functions provide a great sense of safety to visually impaired people. The aim here is to combine depth-of-field detection (up to 20 m) with object detection. The whole assistive system is shown in Figure 1, powered by a lithium battery with a capacity of 5200 mAh (Figure 1a). The battery with fully charged (16.8V) can supply the system operation time of about 2.5 h. The battery and AGX are installed on the backboard of the user. The ZED 2 is fixed on the brim of the user's cap (see Figure 1b). Moreover, the headset is connected to the sound card, and the voice instruction will be output to the user.

Figure 1. The hardware of the whole assistive system, (a) AGX, battery, and sound card on the backboard; (b) The ZED 2 on the cap.

2.1.2. The Software Structure of the System

The software structure of the system is shown in Figure 2, in which there are two deep learning models, Fast-SCNN and YOLOv5s, performing the environment recognition task on the AGX. The result of FAST-SCNN with the depth map provided by ZED 2 can identify the walkable area in front of the user and suggest a walking direction. The YOLOv5s with the depth map can detect the obstacle ahead and calculate the obstacle distance away from the user. Moreover, the voice prompt manager produces audio instruction about obstacles and walkability for the visually impaired person.

Figure 2. The software structure of the assistive system.

2.2. The Functions Required by the Visually Impaired People

In this section, we will present the main configurations of the system in detail and explain how the assistive system helps visually impaired people walk outdoors. Since the most common areas for visually impaired people walking are sidewalks or crosswalks. Recognition of sidewalks and crosswalks should be done first. Second, we should provide the user with the front environment information, either a completely open area or an area with objects. The third task is to select the correct direction for visually impaired people walking forward on the sidewalk or crosswalk. At last, we recognize the types of obstacles to visually impaired people such that the user can know what obstacle is in front of them. Therefore, the following subsections will describe the above three tasks respectively in a more detailed way.

2.2.1. Recognition of Outdoor Environment and the Training of FAST-SCNN

To lead the visually impaired user to walk safely on the sidewalk and crosswalk, we designed the assistive system to identify the sidewalk or crosswalk in front of the visually impaired user. Fast Segmentation Convolutional Neural Network (Fast-SCNN) [23] is a real-time semantic segmentation model on high-resolution image data and offers an efficient computation on embedded devices with limited memory. In addition, Fast-SCNN has fewer parameters than those in other well-known segmentation networks [24]. Therefore, we employed the Fast-SCNN as the segmentation model to recognize the outdoor environment, including sidewalks, crosswalks, stairs, and asphalt roads. Here we suppose that only sidewalks and crosswalks are walkable for visually impaired people.

To train the Fast-SCNN, we collected significant photos captured from the perspective of pedestrians on the street around the test field to form the training dataset. The most common outdoor environments for pedestrians, including sidewalks, crosswalks, stairs, and roads, were involved in our dataset. To be precise, we collected the training data (available via the link <https://github.com/kev72806/TW-NCU-ICIP-Lab-dataset> (accessed on 24 October 2021)) on sunny days during daylight hours and in four different seasons. The training data set includes 22,798 photos, in which there are 3673 original captures and other augmented data. Since the images were collected by ZED 2 video continuously, we randomly adopted one-tenth of the total images as the validation set and screened out some similar images in the set. Furthermore, we used MATLAB Image Labeler to label our data at the pixel level. The total data distribution is shown in Table 1. Use the following methods to do the data augmentation: 1. Randomly rotate the image anti-clockwise or clockwise within 5 degrees and crop 80% of the original size from the center, then enlarge it back to its original size; 2. Change the image's brightness by an arbitrarily finite degree; 3. Flip the image horizontally. Then let the Fast-SCNN network be trained using those training data. Table 2 shows the data or parameters used in the network training.

After training, the Fast-SCNN outputs a segmentation result for the image in front of the user to identify sidewalks, crosswalks, stairs, and asphalt roads in their walking path. Figure 3 shows two examples of semantic segmentation results, where the left and right parts are the original capture and its segmentation result, respectively. The green, blue, grey, and red colors represent crosswalks, sidewalks, asphalt roads, and stairs, respectively. Table 3 lists the performance of Fast-SCNN with pixel accuracy and Intersection over Union (IoU).

Table 1. The data distribution.

Data Type		Pieces
Training data	Sidewalk	3355
	Crosswalk	1676
	Stairs	316
	Road	3263
Total data		3673
Augmentation data		22,798

Table 2. The parameters used in network training.

Item	Value
Resolution	512×288
Epoch	600
Batch size	64
Learning rate	0.01
Momentum	0.9
Weight decay	0.0001
FLOPS	483.412768 M
Memory usage	844 MB

Figure 3. Street view image from ZED 2 and the result of Fast-SCNN.

Table 3. The performance of FAST-SCNN.

Item	IoU
Sidewalk	84.53%
Crosswalk	84.08%
Stairs	76.15%
Road	82.50%
Mean	81.81%
Pixel Accuracy	95.10%

2.2.2. Depth Map and the Openness Values

Although the FAST-SCNN can recognize the above six street environments in front of visually impaired users, it cannot give the environment distance ahead of the user. Therefore, ZED 2 provides the depth map for the image ahead to reinforce this deficiency. Let the color image captured from ZED 2 be on the left of Figure 4. The corresponding depth map is the middle one which was the output from ZED SDK. Take Figure 4 as an example, in which the images (a) and (b) represent the original color image and the depth map corresponding to (a), respectively. Suppose $M_d(i, j)$ is the value corresponding to the pixel (i, j) , where (i, j) is the position of the pixel in the depth map and is defined as

$$M_d(i, j) = \begin{cases} -\infty, & \text{if } d < 1.6 \text{ m} \\ d, & \text{if } 1.6 \leq d \leq 20 \\ 20, & \text{if } d > 20 \end{cases} \quad (1)$$

where d is the distance of the pixel (i, j) in the depth map away from the user, since the ZED 2 is mounted on the cap of the user, the front ground within 1.6 m is invisible, but if an object located within 1.6 m and is high enough to be seen by ZED 2, therefore, the pixel (i, j) on the object with distance $d < 1.6$ m will have $M_d(i, j) = -\infty$. If the pixel on the image with distance $1.6 \leq d \leq 20$, then let the value of $M_d(i, j)$ be the distance; if the pixel on the image with a distance of more than 20 m, then the value of $M_d(i, j)$ is set 20.

Figure 4. (a) The original image; (b) the real depth map corresponding to the image in (a); (c) the depth map of a completely open area.

Next, we consider an entirely open area as the image Figure 4c, in which there is no object except the ground within the image. Let us define $M_o(i, j)$ be the value calculated from (1) in the entirely open area (such as image (c)). The darker the image's color (c) means, the closer it is to the user. Hence, the bottom pixels of the image are close to the user, and the white color of the upper part denotes the pixels are more than 20 m away from the user. Suppose we define the openness value of pixel (i, j) as (2).

$$Open(i, j) = \frac{M_d(i, j)}{M_o(i, j)} \quad (2)$$

It is seen that $\sum_i \sum_j Open(i, j) = -\infty$ or $\in (0, N]$ for any outdoor image, where N is the total number of pixels in the image. In other words, if $\sum_i \sum_j Open(i, j) = -\infty$ which means there is a close obstacle (within 1.6 m) in front of the visually impaired user, but

$\sum_i \sum_j Open(i, j) \in (0, N]$ means the depth map image is not entirely open. In other words, there are some objects except the ground in the image. If $\sum_i \sum_j Open(i, j) = N$ which means the depth map is an entirely open area.

2.2.3. The Selection of Walking Direction

In this study, the semantic segmentation defines the sidewalk and crosswalk with blue and green colors, respectively, as walkable areas. Since the ZED 2 is mounted on the user's cap and suppose the user's height is H (unit: m). Then the bottom width of the captured image field of view is about $3.52 \times H$ meters (when the user looks forward). We split the image evenly into seven divisions from left to right (see Figure 5) and number them from 0 to 6. For example, when the H is 1.7 m, each division has a width of about $3.52 \times H/7 = 0.86$ m which is wide enough for a person to go through. We plan to select the most suitable division from these seven divisions to be the suggested direction for guiding the visually impaired user to walk in.

Figure 5. Seven divisions with $Conf_s$.

The Fast-SCNN identifies the walkable areas, and the ZED 2 provides the depth map for the user. Hence, based on the seven divisions, the next task is to guide the visually impaired user to walk in a safe direction. We define two variables $Conf_s$ and $Conf_d$, to determine the walkability confidence index for seven divisions. $Conf_s$ is the confidence from the segmentation result, and $Conf_d$ is the confidence from the depth map. Now let us introduce the calculation of $Conf_s$ first. Taking Figure 5 as an instance, we binarize the pixel value in the walkable area to 1 and others to 0. The closer the position is to the bottom of the image, the closer it is to the user in the actual field. We define that the pixels closer to the bottom have a higher weight value, and the weight value from the bottom to the top of the image decreases linearly from 2 to 0. Therefore, the pixel values will be weighted by the spatial distance. Since the image is divided into seven equal divisions, as shown in Figure 5, the average weighted pixel value of all pixels in each division will be within $[0,1]$, and that will be regarded as the confidence degree $Conf_s$ of that division, where $Conf_s$ is computed from (3).

$$Conf_s = \frac{1}{N} \sum_i \sum_j w(i, j) * p(i, j) \quad (3)$$

where $w(i, j)$ is the weight and $p(i, j)$ is the pixel value of the position (i, j) in a division. N is the total number of pixels in the division. For instance, the third division from the right in Figure 5 contains large green and blue areas, so its confidence degree is 0.79, the largest in the seven divisions. The rightmost division has a tiny green area, so its $Conf_s = 0.04$ is very low. In other words, the direction shown in the third division from the right is most walkable. Next, let us consider the calculation of $Conf_d$. According to the openness value $Open(i, j)$ in (2), we can calculate a confidence degree based on the depth map. In each

division, the average weighted openness value is computed from (4) and will be regarded as the confidence degree $Conf_d$ of the depth map.

$$Conf_d = \frac{1}{N} \sum_i \sum_j Open(i, j) \quad (4)$$

If there is a pixel with $M_d(i, j) = -\infty$, the $Conf_d$ must be $-\infty$. For instance, in Figure 6, there is $Conf_d$ value for each division, and $Conf_d = -\infty$ in the middle division because there is a tree very close to the user. After we have values of $Conf_s$ and $Conf_d$ for each division, the walkability confidence degree $Conf$ in each division is computed from (5).

$$Conf = \text{minimum}(Conf_s, Conf_d) \quad (5)$$

Figure 6. Seven divisions with $Conf_d$.

Based on (5), we have three confidence degrees on the bottom of each division, as shown in Figure 7, where the first, second, and third rows denote the values of $Conf_s$, $Conf_d$, and $Conf$, respectively. The $Conf$ value will be the critical value to determine the walkability confidence index of each division. For instance, in image (a) of Figure 7, there is no sidewalk or crosswalk, almost all divisions with $Conf_s = 0$, and there are close balusters in front of the user, then all divisions with $Conf_d = -\infty$. In the second division from the left of image (d), there is a small area with blue color (it denotes the sidewalk), then $Conf_s = 0.35$ in that division, and there is a tree close to the user, then the division with $Conf_d = -\infty$.

Figure 7. Confidence values for the seven divisions, (a) the front view without sidewalks and crosswalks; (b) the front view with a sidewalk; (c) the front view of the user who is going to the crosswalk from the sidewalk; (d) the front view with a tree obstacle on the sidewalk; (e) the front view of the user who is crossing the road through crosswalk but looks to the front left; (f) the front view of the user who is walking on the crosswalk.

2.2.4. Walking Guide Strategy

After the confidences for seven divisions are obtained, we give a walking guide strategy for the visually impaired user as follows.

Case 1: If the $Conf > 0.5$ in the middle division, then the middle division is chosen, and the voice prompt gives the instruction “go straight.” Otherwise, check $Conf$ in all divisions and go to Case 2.

Case 2: Find the division with the highest $Conf$, which is larger than 0.2, in all divisions, then guide the user to walk to that division and the voice prompt gives instructions such as “slightly right/left,” “right/left,” or “go straight.”

Case 3: If all $Conf$ values are smaller than 0.2, it means there is a dead-end ahead, and the user must turn around to find another way. In addition, it also occurs in the condition that the user looks up or down. Then the voice prompt plays the instruction “dead.”

Remark 1: If the selected division is the middle division continuously in Case 1 or Case 2, the voice prompt will not repeatedly play “go straight” to avoid interference with the user. For instance, all images in Figure 7 with the walking directions selections are shown in Table 4, and the corresponding voice prompts are given too.

The above walking guide strategy is summarized as Algorithm 1 and Figure 8 below.

At last, we have to add one more remark about the voice prompts. Although both crosswalks and sidewalks are walkable to the visually impaired user, they are different environments that the visually impaired users should know. When the visually impaired user walks on the crosswalk, they should pay more attention to the surrounding environment. In the walking guide strategy, one more function is added. When the middle division is chosen, and the user walks from the sidewalk (or crosswalk) to the crosswalk (or sidewalk), the voice prompt will alert the user that the environment is changed.

Table 4. Walking guides and voice prompts in different cases.

Figure 7	Selected Division (Voice Prompt)
Figure 7a	dead-end, (dead)
Figure 7b	the third division from the left, (slightly left)
Figure 7c	the leftmost division, (left)
Figure 7d	the third division from the right (slightly right)
Figure 7e	the second division from the right, (right)
Figure 7f	The middle division (go straight)

Algorithm 1 Walking guide strategy

Inputs:

The confidence score based on segmentation, $Conf_s \in R^{1 \times 7}$.

The confidence score based on the depth map obtained from ZED 2, $Conf_d \in R^{1 \times 7}$.

Outputs:

The instruction of walking guide, $select \in \{\text{“go straight”, “slightly left”, “slightly right”, “left”, “right”}\}$.

Algorithm:

while the system non-stop **do**

$Conf = \text{minimum}(Conf_s, Conf_d)$

if $Conf[3] > 0.5$, $select = \text{“go straight”}$

else if $Conf[ALL] > 0.2$, Find a value I satisfy $\text{maximum}(Conf) \text{ equal } Conf[I]$

if ($I \text{ equal } 3$), $select = \text{“go straight”}$

else if ($I \text{ equal } 2$), $select = \text{“slightly left”}$

else if ($I \text{ equal } 4$), $select = \text{“slightly right”}$

else if ($I \text{ equal } 0$ or $I \text{ equal } 1$), $select = \text{“left”}$

else if ($I \text{ equal } 5$ or $I \text{ equal } 6$), $select = \text{“right”}$

else, $select = \text{“dead”}$

end while

Figure 8. The flowchart of the walking guide strategy.

2.2.5. Obstacle Detection and the Training of YOLOv5

However, obstacles still may appear on the walking way when the visually impaired user is under the guidance of our assistive system. It should be detected further to prevent the user from danger. Compared with the white cane, the proposed system has the capability of identifying the obstacles. In this work, the frequently-used object detection, namely YOLOv5 [25], is adopted to perform obstacle detection and recognition. Since YOLOv5 has the advantages of lightweight and real-time computation. Especially the size of the YOLOv5s model is 27 MB, which is much smaller than the 245 MB of YOLOv4. Here, considering the most common objects on the road, we pre-determine six categories of objects, including the motorcycle, bicycle, person, car, truck, and bus, that must be detected. The training data for the YOLOv5s comes from a part of the COCO dataset [26] and the street-view captures near Taoyuan High-Speed Rail station, Taiwan. The distribution of our dataset is shown in Table 5. We used 10,000 images from a part of the COCO dataset as training data, and the other 3122 images were collected by ourselves, in which 2534 images were added into the training set, and the other 588 images were used for validation. The last column denotes the label count for each specific object. For instance, 38,095 “person” labels mean that some pictures have multiple objects labeled “person,” so that there are 38,095 “person” labels in the entirety of the 12,534 images. There was no data augmentation during the training stage because the number of samples in the training data was sufficient.

Table 5. The data distribution of our data set is combined with the part of the MS COCO data set.

Data	Image Count	Type	Label Count
Training data	12534	person	38,095
		car	12,506
		motorcycle	2445
		bicycle	1195
		truck	2438
		bus	1079
Validation data	588	person	207
		car	1507
		motorcycle	259
		bicycle	27
		truck	224
		bus	38

Table 6 shows the data and parameters used in network training. Table 7 shows the object detection results by YOLOv5s, where $AP_{50:95}$ in the first column takes values from 0.5 to 0.95, and the best results occur when the value is 0.5. The output of YOLOv5s is a bounding box in which the object can be identified. For instance, the bounding boxes shown in Figure 9 are obtained from YOLOv5s, and the number at the top of each bounding box is the confidence value for the object identification. Further, the distance between the user and the object is obtained from the ZED 2. Let the bounding box obtained from YOLOv5s be cropped to 64% of the original box size by shrinking each side of the box by 10% (see Figure 10). We took the median of all those pixel depths in the cropped box as the object's distance ahead of the user.

Table 6. The parameters used in YOLOv5s training.

Item	Value
resolution	640 × 640
epoch	300
batch size	32
learning rate (lr0)	0.01
momentum	0.937
weight decay	0.0005
FLOPS	8446.739600 M
memory usage	56.236 MB

Table 7. The performance of YOLOv5s.

Six Categories	$AP_{50:95}$	AP50	AP75
person	45.9%	83.2%	47.9%
car	60.2%	93.0%	68.9%
motorcycle	48.9%	86.8%	53.9%
bicycle	39.5%	70.5%	35.8%
truck	57.1%	89.9%	65.9%
bus	64.2%	87.4%	78.5%
mean average precisions (mAP)	52.6%	85.1%	58.5%

Figure 9. Object detection result by YOLOv5s.

Figure 10. The cropped box of object detection.

We also took the vertical line going through the center of the bounding box as the obstacle's position relative to the user. Here we have to combine the walking guide strategy in Section 2.2.4 and the obstacle detection result to cover obstacle identification and avoidance. If the obstacle's position appears in the middle division of the image in front of the user and is close enough, the walking guide strategy will guide the user to avoid the obstacle. At the same time, the voice prompt will inform the type of obstacle they are approaching. Here, the voice prompt will play if the obstacle is located in front of the user within 3 m. For instance, a car located in the third and fourth divisions from the right in Figure 11 is 6.43 m away from the user. When the user walks forward until the distance between the car and the user is within 3 m, the voice prompts the user "car" to remind the visually impaired user to avoid it. However, when the user encounters an obstacle out of the six categories or is hung at the height of the upper body, the system still can avoid the obstacle by the method described in Section 2.2.4 even YOLOv5 cannot recognize it.

Figure 11. The car appears in the third and fourth divisions from the right.

3. Experiments

We experimented with the proposed assistive system to guide a visually impaired person walking on sidewalks and crosswalks. Figure 12 shows that a visually impaired person is joining together to experiment. Notably, the prototype of our proposed system was carried on his back. The experiment was performed outdoors around the MRT station nearby the Taoyuan High-Speed Rail station in Taiwan. The visually impaired user only needs to set the destination, and Google Maps will plan the route from his departure point to the destination. In this experiment, the walking path was planned and shown by the small blue dots in Figure 13, in which the route length is about 280 m. The experimental results show that if we ignore the time consumption of the voice prompts, the processing speed of the system is about 6 FPS (Frames Per Second) which includes 0.09 s for segmentation, 0.06 s for object detection and depth measurement, and 0.0145 s for confidence calculation and division selection. This meets the real-time requirement of use.

Figure 12. The visually impaired person was experimenting, (a) the front view; (b) the back view.

Figure 13. The route that the visually impaired person walked.

We took 3–4 h for one experiment with a visually impaired person. It may have at least 7 or 8 rounds in an experiment. The walking path for an experiment was about 300–1200 m and must include sidewalks and crosswalks. We performed the experiments with three different paths, but those paths were all around the MRT station. Moreover, even though we used the same path, we might put different obstacles in front of the user in different positions. Here, one of the results is shown at the link: <https://www.youtube.com/watch?v=G5QkhQY7h5M> (accessed on 24 October 2021) in which the voice prompts are “right,” “left,” “slightly left,” “slightly right,” “go straight,” and “poly.” The prompt “poly” is played when the user is arriving at a specific point in the planned route, and the prompt “poly” always appears at the curved section (see Figure 13). In the middle of the video, the visually impaired person traversed a street by walking on a crosswalk without a traffic light. In Taiwan, a company called International Integrated Systems, Inc. has a traffic light platform, “Invignal,” [27] to provide the traffic light status and timing for each

intersection in northern Taiwan. However, the experiment using the platform “Invignal” is still in progress. We will publish the study result if we have a more stable experiment result in the future.

4. Discussions

Having performed many rounds of experiments, the visually impaired user gave us some feedback as follows:

1. This system is helpful to the user, and the user appreciates that we care about the needs of visually impaired people;
2. The user likes to be informed by the system if there are objects around the user, but the prompts cannot play too often to bother the user;
3. The walking guide strategy provided by the system gives the user a great sense of security and reduces the number and time of his orientation training;
4. The user would prefer the system to be lighter and smaller just in case the user still needs the white cane;
5. The user needs to take some time to become familiar with this system. The user hopes the system can be friendlier without long time training. We appreciate that the user gave us some positive feedback and suggestions to improve our development in future work.

In Section 2.2.1, we have mentioned that the training data were collected on sunny days during daylight hours to train the Fast CNN. However, we also have done some experiments that used the trained Fast-SCNN on rainy days or at night. The experiment results are stated below. The system still works well in cloudy weather with sufficient brightness. However, it almost fails on rainy days when there is stagnant water affecting the segmentation result of the sidewalk. Furthermore, the system is unstable to recognize the sidewalk and crosswalk without stagnant water. Moreover, it did not work well at night since the environment is too dark, or streetlights of different brightness may confuse the segmentation result. Therefore, for the sake of safety, we suggest the visually impaired uses the proposed system on daytimes and not rainy days. In order to improve the system to be also feasible on rainy days and at night, it needs to collect a lot of extra training data on rainy days and at night. We need more time and more effort to study this issue in the future.

We have to admit that we did not invite enough visually impaired persons to do the experiments since it is hard to find a large number of visually impaired people to test in a short time in Taiwan. However, we have to supplement that before we invited the natural visually impaired person to experiment one time, we would request several (about 5–6) students in my lab wear blindfolds to do many experiments in advance (see Figure 14). This paper just provides a prototype of the assistive system for visually impaired people and proposes a preliminary result for helping the visually impaired. We will continue finding more visually impaired users to experiment with verifying the effectiveness of the proposed system.

Figure 14. Experiments by students with blindfolds.

5. Conclusions

This study has proposed a wearable assistive system to help the visually impaired walk safely on the sidewalk or crosswalk without hitting obstacles. The semantic segmentation model Fast-SCNN was trained to recognize the user's surrounding environment, and the depth map created by the ZED 2 was used to measure the object's distance in front of the user. Combining the above two environmental information, we have developed a walking guide strategy for the visually impaired. Moreover, the object detection model YOLOv5s was trained to detect and identify obstacles. With the aid of the proposed assistive system, the experiment has shown that a visually impaired user can walk on the sidewalk and crosswalk safely without hitting any obstacles. However, we have to admit that if an object appears suddenly in the camera's blind spot within 0.2–1 m in front of the user, they may not have time to avoid it. In this situation, the white cane is still helpful for the visually impaired person. Therefore, we suggest that the visually impaired person still uses the white cane to maximize safety when using the assistive system. In addition, the study on visually impaired people traversing an intersection with traffic lights is still in progress. The traffic light platform "Invignal" can provide the traffic light status and timing for each intersection in northern Taiwan. This platform will be used in our consequent experiment. We will publish the study result if we have a more stable experiment result in the future. Furthermore, we will continue experimenting with more visually impaired users to verify the effectiveness of our proposed system. For the limited testers, we have completed different experiments many times, and the performance is almost consistent.

Author Contributions: Conceptualization, W.-J.W.; Formal analysis, W.-J.W.; Funding acquisition, W.-J.W.; Investigation, W.-J.W.; Methodology, I.-H.H., W.-J.W., and H.-C.C. (Hsiao-Chu Cheng); Resources, H.-C.C. (Hsiang-Chieh Chen); Software, I.-H.H. and H.-C.C. (Hsiao-Chu Cheng); Validation, W.-J.W. and H.-H.K.; Visualization, I.-H.H., H.-C.C. (Hsiao-Chu Cheng), and H.-H.K.; Writing—original draft, I.-H.H., H.-C.C. (Hsiao-Chu Cheng), and H.-H.K.; Writing—review and editing, I.-H.H., W.-J.W., H.-C.C. (Hsiao-Chu Cheng), H.-H.K., and H.-C.C. (Hsiang-Chieh Chen). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science and Technology of Taiwan with the grant number 110-2634-F-008-005.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available in <https://github.com/kev72806/TW-NCU-ICIP-Lab-dataset> (accessed on 24 October 2021).

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bourne, R.R.; Flaxman, S.R.; Braithwaite, T.; Cicinelli, M.V.; Das, A.; Jonas, J.B.; Keeffe, J.; Kempen, J.H.; Leasher, J.; Limburg, H.; et al. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis. *Lancet Glob. Health* **2017**, *5*, e888–e897. [\[CrossRef\]](#)
2. Wei, Y.; Lee, M. A guide-dog robot system research for the visually impaired. In Proceedings of the 2014 IEEE International Conference on Industrial Technology (ICIT), Busan, Korea, 26 February–1 March 2014; pp. 800–805.
3. Blasch, B.B.; Wiener, W.R.; Welsh, R.L. *Foundations of Orientation and Mobility*, 2nd ed.; AFB Press: New York, NY, USA, 1997.
4. Bai, J.; Liu, D.; Su, G.; Fu, Z. A cloud and vision-based navigation system used for blind people. In Proceedings of the International Conference on Artificial Intelligence, Automation and Control Technologies, Wuhan, China, 7–9 April 2017; pp. 1–6.
5. Bai, J.; Lian, S.; Liu, Z.; Wang, K.; Liu, D. Smart guiding glasses for visually impaired people in indoor environment. *IEEE Trans. Consum. Electron.* **2017**, *63*, 258–266. [\[CrossRef\]](#)
6. Bai, J.; Lian, S.; Liu, Z.; Wang, K.; Liu, D. Virtual-blind-road following-based wearable navigation device for blind people. *IEEE Trans. Consum. Electron.* **2018**, *64*, 136–143. [\[CrossRef\]](#)
7. Bai, J.; Liu, Z.; Lin, Y.; Li, Y.; Lian, S.; Liu, D. Wearable travel aid for environment perception and navigation of visually impaired people. *Electronics* **2019**, *8*, 697. [\[CrossRef\]](#)

8. Islam, M.M.; Sadi, M.S.; Zamli, K.Z.; Ahmed, M.M. Developing walking assistants for visually impaired people: A review. *IEEE Sens. J.* **2019**, *19*, 2814–2828. [[CrossRef](#)]
9. Lin, Y.; Wang, K.; Yi, W.; Lian, S. Deep Learning Based Wearable Assistive System for Visually Impaired People. In Proceedings of the International Conference on Computer Vision Workshop, Seoul, Korea, 27 October–2 November 2019.
10. Yang, K.; Wang, K.; Bergasa, L.M.; Romera, E.; Hu, W.; Sun, D.; Sun, J.; Cheng, R.; Chen, T.; López, E. Unifying Terrain Awareness for the Visually Impaired through Real-Time Semantic Segmentation. *Sensors* **2018**, *18*, 1506. [[CrossRef](#)] [[PubMed](#)]
11. Cao, Z.; Xu, X.; Hu, B.; Zhou, M. Rapid Detection of Blind Roads and Crosswalks by Using a Lightweight Semantic Segmentation Network. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 6188–6197. [[CrossRef](#)]
12. Yang, K.; Cheng, R.; Bergasa, L.M.; Romera, E.; Wang, K.; Long, N. Intersection perception through real-time semantic segmentation to assist navigation of visually impaired pedestrians. In Proceedings of the International Conference on Robotics and Biomimetics, Kuala Lumpur, Malaysia, 12–15 December 2018.
13. Sáez, J.M.; Escolano, F.; Lozano, M.A. Aerial obstacle detection with 3-D mobile devices. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 74–80. [[CrossRef](#)] [[PubMed](#)]
14. Chang, W.; Chen, L.; Chen, M.; Su, J.; Sie, C.; Yang, C. Design and Implementation of an Intelligent Assistive System for Visually Impaired People for Aerial Obstacle Avoidance and Fall Detection. *IEEE Sens. J.* **2020**, *20*, 10199–10210. [[CrossRef](#)]
15. Ramadhan, A.J. Wearable smart system for visually impaired people. *Sensors* **2018**, *18*, 843. [[CrossRef](#)] [[PubMed](#)]
16. Elmannai, W.M.; Elleithy, K.M. A highly accurate and reliable data fusion framework for guiding the visually impaired. *IEEE Access* **2018**, *6*, 33029–33054. [[CrossRef](#)]
17. Croce, D.; Giarre', L.; Pascucci, F.; Tinnirello, I.; Galioto, G.E.; Garlisi, D.; Valvo, A.L. An Indoor and Outdoor Navigation System for Visually Impaired People. *IEEE Access* **2019**, *7*, 170406–170418. [[CrossRef](#)]
18. Aladren, A.; Lopez-Nicolas, G.; Puig, L.; Guerrero, J.J. Navigation Assistance for the Visually Impaired Using RGB-D Sensor with Range Expansion. *IEEE Syst. J.* **2016**, *10*, 922–932. [[CrossRef](#)]
19. Tapu, R.; Mocanu, B.; Zaharia, T. DEEP-SEE: Joint object detection, tracking and recognition with application to visually impaired navigational assistance. *Sensors* **2017**, *17*, 2473. [[CrossRef](#)] [[PubMed](#)]
20. Lin, S.; Cheng, R.; Wang, K.; Yang, K. Visual localizer: Outdoor localization based on convnet descriptor and global optimization for visually impaired pedestrians. *Sensors* **2018**, *18*, 2476. [[CrossRef](#)] [[PubMed](#)]
21. Chang, W.J.; Chen, L.B.; Sie, C.Y.; Yang, C.H. An artificial intelligence edge computing-based assistive system for visually impaired pedestrian safety at zebra crossings. *IEEE Trans. Consum. Electron.* **2020**, *67*, 3–11. [[CrossRef](#)]
22. NVIDIA NVIDIA Jetson AGX Xavier. Available online: <https://www.nvidia.com/it-it/autonomous-machines/embeddedsystems/jetson-agx-xavier/> (accessed on 3 August 2021).
23. Poudel, R.P.; Liwicki, S.; Cipolla, R. Fast-SCNN: Fast semantic segmentation network. *arXiv* **2019**, arXiv:1902.04502.
24. Jahromi, S.A.F.; Khani, A.A.; Shahreza, H.O.; Baghshah, M.S.; Behrooz, H. A Deep Learning Framework for Viable Tumor Burden Estimation. In Proceedings of the 2020 6th Iranian Conference on Signal Processing and Intelligent Systems, Shahrood, Iran, 23–24 December 2020; pp. 1–7.
25. Jocher, G.; Stoken, A.; Borovec, J.; NanoCode012; ChristopherSTAN; Changyu, L.; Laughing; tkianai; yxNONG; Hogan, A.; et al. Ultralytics/yolov5: v4.0—nn.SiLU() Activations, Weights & Biases Logging, PyTorch Hub Integration. 2021; Available online: <https://doi.org/10.5281/zenodo.4418161> (accessed on 3 August 2021). [[CrossRef](#)]
26. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
27. International Integrated Systems, Inc. Available online: <http://www.invignal.com/> (accessed on 24 October 2021).