# An Electronic Guide Dog for the Blind based on Artificial Neural Networks

S. Lopatin[0000−0001−5031−0267], F. v. Zabiensky[0000−0003−0951−0288], M. Kreutzer[0000−0003−0748−7707], K. Rinn[0000−0002−1776−9751], and D. Bienhaus[0000−0002−9207−3412]

Technische Hochschule Mittelhessen, University of Applied Sciences,
Institute of Technology and Computer Science, Giessen, Germany
(sergej.lopatin | florian.von.zabiensky | michael.kreutzer | klaus.rinn | diethelm.bienhaus)@mni.thm.de

**Abstract.** This paper presents a feasibility study of an electronic assistance system to support blind and visually impaired people in finding their way in the area of public traffic. Optical recognition of walkways is implemented. For this purpose, a neural network for semantic segmentation is trained from scratch. In the practical test, an NVIDIA® Jetson Nano™ is used as the computing unit. A voice output gives the user feedback for orientation on the pavement.

**Keywords:** electronic travel aid, blind sidewalk detection, portable ETA system, electronic travel aid technology, computer vision, convolutional neural network

## 1 Introduction

The goal of this work is to design a deep learning based *Electronic Guide Dog* to support visually impaired people in orientation and navigation tasks, especially in following unmarked side-walks and obstacle avoidance. It uses a camera to detect points of interests and obstacles in the environment and offers the blind user an intuitive human machine interface with audio feedback for safe navigation.

Typically, blind people use a white cane to scan their surroundings for obstacles and orientation marks. The cane covers an area up to about 1.5 m in front of the user from ground to about waist height. Due to this limitation, obstacles such as letterboxes or branches that protrude into the pavement can often not be detected and injuries occur. Because of the short range, moving obstacles are often detected too late, which leads to collisions with oncoming pedestrians. On footpaths that do not have a special guidance system for the blind and distinctive tactile edge markings, it often happens that the blind person strays from the path. Once the blind person is on the road, it is difficult to regain orientation because tactile markings are missing and there is a high risk of accidents with vehicles. [8]

Technical devices to solve these problems are called Electronic Travel Aids (ETAs). Elmannai and Elleithy presented a comparative survey of the wearable and portable sensor-based assistive devices for visually-impaired people in order to show the progress in this field. [4]

Some devices require the expansion of the infrastructure with electronic beacons or special markers. A Radio Frequency Identification Walking Stick (RFIWS) was designed in [7] in order to help blind people navigating on the sidewalk. This system helps detecting and calculating the approximate distance between the sidewalk border and the blind person. The system requires the preparation of pavements with RF-ID tags and is therefore costly and only available in a few locations.

A classic camera image-processing approach to side-walk tracking, is presented in [5]. The wearable system uses image segmentation, edge detection and boundary searching and is limited to recognize special markers (also called blind sidewalk) which are not always available.

A Deep Learning based approach to detect known obstacles or orientation points is shown in [1]. The framework employs transfer learning on a Single-Shot Detection (SSD) mechanism for object detection and classification using the Inception v3 convolutional neural network for object detection. Names of object-classes are presented to the visually impaired person in audio format. The main disadvantage of this approach is, that only learned (known) objects are recognized and localized in the camera-image, while unknown obstacles remain unidentified and therefore dangerous. So the system can only be used to identify orientation marks or other points of interest. Detecting the center of a sidewalk is also not possible in this way.

Core of this work is to use a state of the art segmenting neural network (SNN), which is able to perform a pixel by pixel segmentation of regions like pavement. While the pavement is recognized as a walkable region, unknown obstacles appear as holes in this region and can thus be localized. By analyzing the region map with classical methods of image processing, the position on the pavement can also be determined and safe guidance with voice commands can be realized. An NVIDIA® Jetson Nano™ provides the necessary computing power in mobile operation.

## 2 System Architecture

The Nano™ is a single board computer equipped with a GPU unit designed for deep learning tasks. One advantage is, that the computation of neural networks can be performed on the fly without internet connection. So latency problems can be avoided and data security is provided. Furthermore, the *Electronic Guide Dog* consists of a camera, a chest harness for the user to attach the camera, simple ear phones and a mobile battery. We use the Robot Operating System (ROS) version 2 for the systems software infrastructure. Figure 1 illustrates the image processing steps. Our system currently consists of five nodes as shown in Figure 2.

1. The user's environment is captured frame by frame by a Camera Node.
2. The CNN Segmentation Node composes our trained neural network model to detect the environment's objects, at this stage merely sidewalk. The reason we use the segmentation task is to determine the shape of sidewalks. We train a convolutional neural network (CNN) model of BiSeNetV2 [9] with our handcrafted dataset labeled with sidewalks.
3. The Environment Node extracts some features of the sidewalk, for example, the computed width, edges, centroid and the intersection of the edges. These features provides a basis to compute the user's deviation from the center line of the sidewalk.
4. The Sidewalk Node of the system is used to compute the navigation feedback w.r.t the deviation.
5. The user receives the navigation feedback via voice output realized by the systems audio interface.

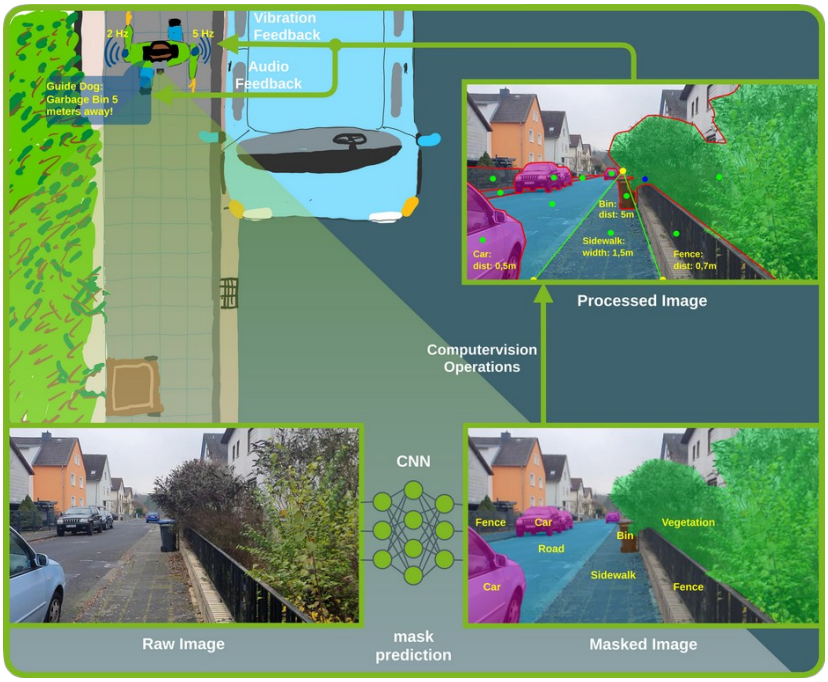


**Fig. 1.** Functional Scheme of the *Electronic Guide Dog*

## 3 CNN Evaluation

The DeepLabV3 [2] network architecture was developed by Google in 2017. DeepLabV3 achieves a mean intersection over union (mIoU) score of 85.7%

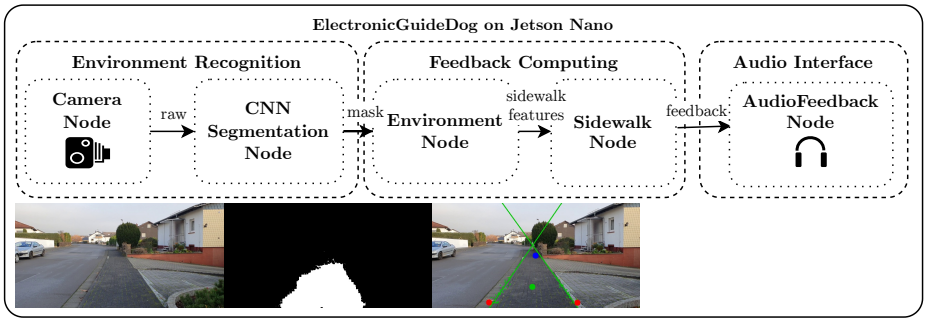


**Fig. 2.** System architecture of the *Electronic Guide Dog* . Doted border shapes represent ROS nodes. Dashed bordered shapes define a subsystem. Arrows describe the message-oriented data flow. The computation results of the single steps are shown in the bottom images.

and 81.3% on PASCAL VOC 2012 and Cityscapes, respectively. The averaged inference performance is 8 FPS on an NVidia Titan X GPU and 0.5 FPS on the CPU. The internal PyTorch implementation of DeepLabV3 with the backbone ResNet50 uses a model size of 321 MB. BiSeNetV2 is the newer architecture, published by Yu et al. [9]. A mIoU score of 72.6% was achieved and an inference performance of 156 FPS. Our trained model of CoinCheung's implementation of BiSeNetV2 has a model size of 58 MB [3]. BiSeNetV2 provides the better trade-off between speed and precision. The inference speed of 156 FPS is much higher than 8 FPS for the inference of DeepLabV3 and the specified precision of 72.6% mIoU is a sufficiently good value. Figure 3 shows the comparison between the architectures in terms of speed and precision. The smaller model size of BiSeNetV2 compared to DeepLabV3 is suited better for mobile devices.

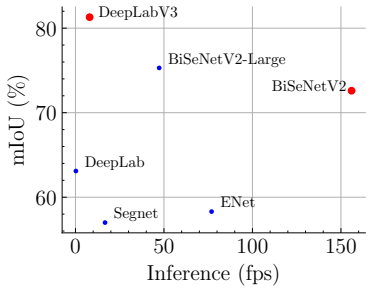


**Fig. 3.** Comparison of speed and precision of the architectures. Red dots indicate the architectures that are available for selection. Blue dots are other notable architectures.

We develop a dataset with focus on sidewalks by taking pictures under different weather conditions. The images also contain different perspectives of the walkway to train a robust CNN that can be used in realistic conditions. As well known, the labeling task consumes a considerable amount of time. Hence, we first labeled only sidewalks. The dataset includes 850 verified images and is divided into 70% training, 15% testing and 15% validation subsets.

%subsectionTraining Results

New models of DeepLabV3 and BiSeNetV2 were trained with the sidewalk dataset from scratch. The loss of the DeeplabV3 converges quickly over 20 epochs to an approximate value of 0.1 and reaches an mIoU score of 78% on our testset. We train CoinCheung's implementation of the BiSeNetV2 for more than 1500 epochs [9], [3]. The model achieves a mIoU score of 58%. Further, we measure the inference speed on the GPU of the Nano$^{TM}$ using PyTorch. The DeeplabV3 and BiSeNetV2 models reach an inference speed of 0.3 FPS and 4.2 FPS. Figure 4 shows one test sample for each model.

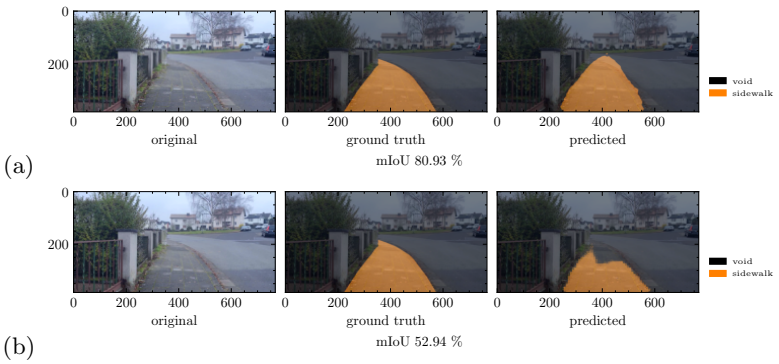


**Fig. 4.** The test result of a random sample for Deeplabv3 (a) and BiSeNetV2 (b) can be seen in the top and bottom row respectively.

The resulting prediction quality of 58% mIoU is 20% worse and the inference speed of 4.2 fps is 14 times higher compared to the Deeplabv3 model. To achieve usable results in field tests, an inference speed of 1 FPS needs to be surpassed. Hence, the trained Deeplabv3 model doesn't fit this requirement for the current system setup and we decide to use BiSeNetV2 as the appropriate segmentation model for our system.

## 4 User Interface

In order to use the system as an assistance system, it must be able to inform the user about the sidewalk course. For this purpose, the *Electronic Guide Dog* starts with the idea of a virtual guide system for the blind. Grooves in the ground, that can be felt with a white cane signal the direction of a sidewalk. The ETA user

should get feedback of the ideal path on the sidewalk in a similar way, especially if there is no guide system for the blind installed in the floor.

For this purpose, the *Electronic Guide Dog* uses results of the image processing and utilizes the walkway found in the image. As shown in Figure 5, the edge of the sidewalk is detected and the preferred walking direction is calculated based on an image region close to the user. Using only a nearby image region is based on the idea that the sidewalk in 10m distance is irrelevant for planning the next step. Due to the perspective in the image, the trajectories from the left and right edges of the sidewalk intersect. The horizontal difference of this intersection point to the center of the image represents the deviation from the center line of the walkway if the user plans to walk straight on. Under real conditions, the direction of the user's movement in relation to the image depends on the position and orientation of the camera. To simplify the problem, we assumed that the column of the center of the image corresponds to the direction of the user's gaze and movement.

The implemented user interface is separated from the underlying algorithm by means of using only the deviation from the walkway center line. Based on this value, audio commands are generated to guide the user to the center line. Text to speech is used to generate orientation commands. In case of deviations from the center instructions to go to the opposite direction are given. In addition, reaching and crossing the edges of the walkway is communicated.

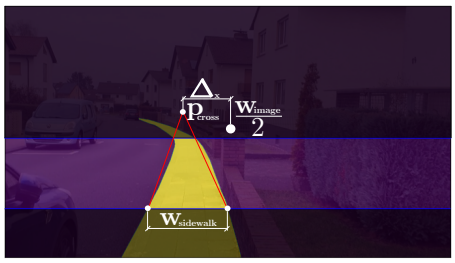


**Fig. 5.** Parameters for calculating the deviation from the optimal path

## 5 Experiments and Results

We have tested the system in good weath

er conditions. The test location is comparable to the dataset recordings. Figure 6 shows the experimental prototype.

During the test we record the processing steps as image files for later analysis. The Figure 7 shows image frames overlayed with test results.

(a) (b) (c) (d)

**Fig. 6. Experimental prototype**. (a) Assembly (b) Camera fixed on breast harness (c) The computing unit covered in a bag (d) Earphones.

**Fig. 7. Segmentation examples of field test.** Some good segmentation results.

The output of the navigation instructions was responsive enough. On wide pavements the system navigated with the appropriate direction in about two-third of all cases.

The system has weaknesses in the classification of narrow sidewalks and if the terrain is unknown. This is caused by the small data base of only 850 samples. A generalization of the sidewalk segmentation is not yet achieved. Our dataset contains samples of sidewalk records with asphalt pavement, which is also a characteristic of roads. As a result, our system has problems distinguishing pavements from roads. Occasionally, the full width of a pavement is not segmented. Divergence of the camera position related to the user's path are caused by the user's walking movement which may cause a wrong audio instruction. Sometimes the system generates wrong instructions due to shaky recordings. Therefore, image stabilization algorithms must be implemented.

## 6 Conclusion

This work illustrates that convolutional neural networks (CNN) are appropriate for semantic segmentation of camera-captured environment images for ETAs to guide visually impaired users on the center line of a sidewalk.

In our approach a CNN performs a semantic segmentation of the camera-captured environment. The resulting masked images are processed using computer vision algorithms to detect an obstacle free and safe walkway. Finally, the user receives information about orientation on the walkway via voice output. Before deployment, the network was trained with 850 labeled samples from the pedestrian view. The implementation provides a classification at a frame rate of

approximately 4 fps on a NVIDIA® Jetson Nano™ using PyTorch for inference. A short demonstration of the *Electronic Guide Dog* is provided as video in [6].

Initial field tests show that the system delivers good results in the trained environment. The user is reliably kept in the middle of the walkway and unknown obstacles are mostly detected. In new environments, however, the error rate increases, which can be attributed to the currently too small base of training data.

Additional annotations of roads and other classes in the data set would lead to the network differentiating the objects of the environment better. A larger data set is required to train a network model to the point of generalization. An inertial measurement unit (IMU) or image correction by homography can be used for image stabilization. The use of a depth camera can improve obstacle detection by delivering distance information.

For the recognition of traffic signs or lights object detection CNN architectures can be used. In comparison to segmentation CNN architectures this enables an easier labeling process as well as faster processing of the image. A combination of different CNN architectures is promising to gain better results by simultaneous object recognition and semantically segmentation of the environment.

## References

1. Bhole, S., Dhok, A.: Deep learning based object detection and recognition framework for the visually-impaired. In: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). pp. 725–728. IEEE (2020)
2. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs http://arxiv.org/abs/1606.00915
3. CoinCheung: CoinCheung/BiSeNet, https://github.com/CoinCheung/BiSeNet, original-date: 2018-11-29T04:27:51Z
4. Elmannai, W., Elleithy, K.: Sensor-Based Assistive Devices for Visually-Impaired People: Current Status, Challenges, and Future Directions. Sensors **17**(3), 565 (Mar 2017). https://doi.org/10.3390/s17030565, http://www.mdpi.com/1424-8220/17/3/565
5. Jie, X., Xiaochi, W., Zhigang, F.: Research and implementation of blind sidewalk detection in portable eta system. In: 2010 International Forum on Information Technology and Applications. vol. 2, pp. 431–434 (2010). https://doi.org/10.1109/IFITA.2010.187
6. Lopatin, S.: Electronic Guide Dog - video of some tests, https://youtu.be/B0bz_o08990
7. Saaid, M.F., Ismail, I., Noor, M.Z.H.: Radio frequency identification walking stick (rfiws): A device for the blind. In: 2009 5th International Colloquium on Signal Processing & Its Applications. pp. 250–253. IEEE (2009)
8. Sheth, R., Rajandekar, S., Laddha, S., Chaudhari, R.: Smart white cane–an elegant and economic walking aid. American Journal of Engineering Research **3**(10), 84–89 (2014)

9. Yu, C., Gao, C., Wang, J., Yu, G., Shen, C., Sang, N.: BiSeNet v2: Bilateral network with guided aggregation for real-time semantic segmentation http://arxiv.org/abs/2004.02147