Outdoor Navigation for Visually Impaired based on Deep Learning

Saleh Shadi
M.Sc., Chair of Computer Engineering, Chemnitz University of Technology
Chemnitz, Germany D-09111
shadi.saleh@informatik.tu-chemnitz.de

Saleh Hadi
PhD, Associate Professor of National Research University Higher School of Economics,
Associate Professor of Vladimir State University named after Alexander and Nikolay Stoletov,
Moscow, Russia 125319
hsalekh@hse.ru

Mohammad Amin Nazari
M.Sc, National Research University Higher School of Economics
Moscow, Russia 125319
amin.puia@gmail.com

Wolfram Hardt
Prof, Chair of Computer Engineering, Chemnitz University of Technology
Chemnitz, Germany D-09111
hardt@cs.tu-chemnitz.de

Abstract: Visually impaired and blind people frequently have no knowledge of outdoor obstacles. and need guidance in order to avoid colliding risks. The aim of this research is to develop a mobile-based navigation system for helping visually impaired people in outdoor navigation. The proposed system will be able to reduce the obstacle collision risks by enabling users to walk outside smoothly with voice awareness. The current used systems for navigating visually impaired have several drawbacks such as cost, dependency, and usability. The suggested solution includes a mobile-based camera vision system to build an independent application for outdoor navigation. Moreover, the system has high usability to navigate visually impaired people in unfamiliar environments such as a park, roads and so on. In the presented work, the deep learning algorithms are employed for recognizing and detecting different objects and it is implemented as a mobile navigation application. The suggested smartphone-based system is not restricted to the defined outdoor environments and does not depend on any other positioning system. Therefore, the proposed solution is not limited to any specific environment and provides the voice aids about surrounding obstacles for users.

Keywords: Mobile-based system, blind people, navigation application, visually impaired people, machine learning, deep learning, obstacle recognition, object recognition.

1 Introduction

According to the world blind union, there are about 253 million people around the world with serious vision problems and about 47 million of them are blind [1]. Visually impaired people can sense light and shadows only, and not able to see objects in front of them. Furthermore, they move around based on their senses and experiences with the aids of guidance cans to detect and avoid collision with moving and stationary obstacles. Sometimes, the guide canes don't offer they required safety levels because they don't provide perception of the obstacles or objects types and also do not give information about the walking path. Moreover, when an unexpected collision happens, visually impaired people have to react based on their experience. However, their experiences do not provide enough information to predict possible hazards because there are many obstacles in indoor and outdoor which need a quick reaction to avoid a collision.

2 Problem Statement

Navigating through unknown environments is a challenging tricky task for individuals with vision impairments. Two main challenges affect a person's ability to navigate. The first is obstacle avoidance, which addresses the objects and terrain in the person's immediate surroundings environment such as persons, stairs, walls and tables. Many tools have been invented and are being actively employed to help and assist visually impaired people to avoid obstacles such as guide canes and seeing-eye dogs. layouts or terrain mappings is a difficult task. In general, there is not any proficient system to be used as a navigation proposes for blind people. Studies illustrate that the visually impaired feel that they have a special difficulty of learning new routes [2]. In addition, the inability to handle these types of situations themselves has an adverse impact on the feeling of independence of the visually impaired. Visually impaired people have little opportunity to find their way in an unfamiliar place. It is useful to have a person as a guide, however, it is not always a good option because it creates a sense of dependence on other people. Asking for directions is common. However, it has several challenges. Many people might find it tricky to transform directions into meaningful and effective information for the visually impaired individual. The difficulty arises from the fact that most people without visual impairments navigate differently, typically relying on visual landmarks. Moreover, people with visual impairments have to memorize the directions, as it is impossible to note them down. If the visually impaired get lost, the only way is to find someone to help them. Braille signs [3] are a good solution to help visually impaired people, however, the problem with this approach is that it cannot be used as a navigation tool. Nowadays, many public areas such as hospitals, railway stations and educational buildings, doors, elevators and another part of the building are equipped with Braille signs to make navigation easier for visually impaired people. Despite the fact that Braille characters can help visually impaired people to know their location, they do not provide navigation assistance. Moreover, most visually impaired people do not have the knowledge to understand the Braille characters. For example, in the United State of America, only about 10 percent of visually impaired people know Braille [4]. Another widely used system is the Global Positioning System (GPS) [5]. It has been used for outdoor navigation by everyone, including visually impaired people. The GPS is included in smartphones that allow the translation of text into voice guidance. However, the GPS itself is not sufficient for blind and visually impaired people due to its low accuracy and possible interference for outdoor navigation. The discussed methods like (guide canes, GPS and Braille signs) are obviously not feasible techniques for visually impaired people to find their way in an unfamiliar environment or outdoors. Recently, several outdoor navigation systems have been developed to aid blind people in navigation. Unfortunately, most of them are not easy to use due to complicated configurations and dependencies [3,6]. Other solution strategies have various dependencies such as sensors and other devices that are not suitable for visually impaired people to carry around for purposes of navigation [7]. In this context, most studies also showed that blind people do not prefer to be disturbed by other unnecessary devices [5]. On the other hand, costs play a major role in the development of outdoor navigation devices. In conclusion, the navigation system for the visually impaired must be portable, independent, user-friendly, lightweight and cost-effective for the blind to use as a navigation system. Finally, the visually impaired people hands should be free for other operation such as holding their personal properties and guide cane to avoid some sudden collision.

2.1 Solutions Strategies

The investigation of an efficient and accurate navigation system for visually impaired people is an interesting area of active scientific research. Recently various approaches and methods have been proposed, such as Electronic Travel Aids (ETAs) [8], Electronic Orientation Aids (EOAs) [9], Position Locator Devices (PLDs) [10] and Microsoft Seeing AI [11]. Each method has some pros and cons. Table 1 illustrates the comparison and analysis of existing systems based on cost, supported voice capability and ability to operate on the mobile device, hardware/sensors used, real-time capability and level of accuracy.

Table 1: Comparison table of analyzing existent system.

	Cost	Voice awareness	Implemented on mobile	Hardware dependency	Implemented technology	Real-time awareness	Accuracy
Electronic Travel Aids (ETAs)	Very high	Not Fully	Yes	Very High	RFID, Multi Sensors	Yes	Good as indoor
Electronic Orientation Aids (EOAs)	Very high	Yes	Not	Very High	Camera and multi sensors	Yes	Good
Position Locator Device (PLDs)	High	Yes	Yes	Not much	GPS and GIS	In some condition	Not Good
Microsoft Seeing AI	Cheap	Yes	Yes	Not	CNN	Not	Good but Not for Navigation propose

As shown in Table 1, it was concluded that there is a gap between all existing systems and outdoor navigation systems for visually impaired people. Microsoft Seeing AI [11] is one of the best options for converting text to voice, it can help blind people learn more about the product and price, it is a useful approach to read the scanned page of the book or to know about the objects and people. However, it is not suitable to navigate blind or visually impaired people in order to estimate the safety of a way in an outdoor environment. RFID chips (Radio Frequency Identification) are another solution [12] utilized for the navigation system, however, it is not an optimal solution due to the high price in addition to the risk of damage by rain and sun. Furthermore, most of the current system depends on some extensive peripheral devices, and the usage of these devices is generally not easy.

3 Research Objective

Most studies and existing systems about an outdoor navigation system for blind or visually impaired people focus on the distance calculation and positioning system. Therefore, the implementation of such a system is limited to get a good understanding of the environment that is critical for our situation. This study proposes an efficient novel solution based on the smartphone camera and deep learning algorithms to detect and recognize different objects and obstacles. Where segmentation-based object recognition is implemented to detect and recognize objects based on pixels with specific segmentation color and class probabilities. The segmented predication color for an object and class probabilities in a single neural network directly is produced from the full image in one evaluation. Because the detection pipeline from a single neural network, it is possible to optimize them directly as end-to-end detection performance. The images of the environment are continuously captured in front of the user, then object detection and image processing are performed to recognize the objects with the estimated distance, in order to deliver voice comments to visually impaired people about the objects or obstacles in front of them. As the result showed, visually impaired people will be more alerted to their surrounding environment and the proposed system will assist them not only in navigating in an unfamiliar environment but also in getting more information about the various obstacles they may face them with the estimated distance.

3.1 Object Detection using Deep Learning

Recent studies have shown that none of the current object recognition algorithms based on vision sensors has achieved the high accuracy as human eyes and it is hard to replace them by human eyes [13]. Nowadays, in the industrial and commercial environment, most of the implemented technologies have integrated the neural network and object recognition. However, there are many serious limitations in terms of high accuracy, large training data, computing resources, lack of appropriate data analysis techniques, object differentiation, speed of moving objects, and lack of good data for the training model. Therefore, it is inevitable to review and evaluate different object recognition techniques in order to understand the existing limitation.

Object recognition and tracking problem have been a very hot topic over the last two decades and it is an interesting area to research until now. Several studies consider the change of the dynamic scene on the objects and update the appearances of the objects [14]. Other approaches focus on the fusion of multi-sensor systems for object recognition and tracking. Convolutional Neural Network (CNN) techniques are commonly applied in object recognition, and despite its high accuracy, CNN has some workout difficulties due to the overfitting problem caused by noisy labeled data, especially the few training examples [15]. An additional research attempted to address the issue of object recognition from the moving stream in video through adaptive learning tools combined with efficient classification of description (using Visual Vocabulary Model and Bag of Words) ,then the specific object of interest in subsequent video images is tracked by training the proposed model with Support Vector Machine (SVM) [16]. By considering the training of the Online Convolutional

Neural Network (CNN) [17] as an individual-based leaner for each output feature map, the sequential training of the convolutional network was introduced. In order to avoid overfitting and reduce correlation, each individual network training uses a different loss criterion, to decide on the learning methods, all are sequentially sampled into the ensemble. the international VOT 2016 Tracking Contest [18] contributed by the Continuous Convolution Operator Tracker (C-COT) [19]. The C-COT CNN Tracker learns about discriminatory convolution operators in the continuous spatial domain. The system enables the integration of feature maps with multiple resolutions and the subpixel objects localization.

However, the training process of the neural network is difficult and complex because it requires a large resource with an important test phase process. Therefore, depending on the speed for all the above methods of tracking, the specified range should be from 0.8 FPS to 10 FPS, while the highest result of performing algorithms based on evaluated test runs at 1 FPS on GPU [20].

3.2 Analysis Performance of Existing Approaches

The proposed approach should be able to recognize a large number of objects and classify them into specific object categories based on the layers for each image, with real-time capabilities and minimal resource consumption. Moreover, the task of specific objects within images generally involves providing bounding boxes and labels for individual objects. This task differentiates itself from the classification and localization tasks by applying the classification and localization for many objects and not only for a single dominant object.

In order to cope with this, researchers have proposed different architectures and frameworks. Modern object detectors are based on the following network architectures such as DeepLab [21] Fast R-CNN [22], YOLO [23], Multibox [24], SSD [25] and R-FCN [26]. Nowadays, these approaches are sufficient to be employed in consumer products, and some have proven to be fast enough to be used on mobile devices. However, it can be challenging for practitioners to determine which architecture is most appropriate for their application. Standard-accuracy metrics, such as mean average accuracy (mAP), do not provide full coverage of the problem, as run-time and memory consumption are essential for the actual use of image processing systems in the real-world deployment .Overall, the DeepLab approach with the Pascal VOC and City Scopes [21] datasets provide mean Average Precision (mAP) as 81.3 out of 100, while the running frames are about 44 frames per second. Compared to other approaches such as SSD and Fast YOLO, which achieved the best speed of real-time frames per second. However, the size of the model is large, which reduces the efficiency of its usage for mobile object detection. The comparisons of the different techniques used to train the object recognition model are shown in Table 2. From this comparison, we can understand the important factors that are necessary for implementing the Deep Learning application on the mobile device.

The most important factors for the deployment of an outdoor navigation system are the size of the model, the accuracy and the implement ability on the mobile device. In summary, our main contributions are as follows: DeepLabV3+ is the most appropriate option to use as a model for implementing the outdoor navigation system.

Table 2: Comparison and analysis table of the performance of different existing classifier models

Detection Framework	mAP	FPS	Test Time per image	Size of Model (MB)	Number of objects	Implementation on mobile device
R-CNN	44.3	0.2	47 Sec	480	10400	No
Fast R-CNN	66.9	0.5	2 Sec	360	9100	No
Faster RCNN	73.2	7	0.2 Sec	350	8000	No
YOLO	63.4	45	0.5	188	10000	Yes
Fast YOLO(V2)	77.8	59	0.1	130	9800	Yes
SSD	76.8	46	0.2	174	7300	Yes
DeepLabV3+	**81.3**	44	**0.1**	**87**	8500	Yes

4 Problem Solution and Methodology

In this study, the proposed solution should be able to run in real-time on the mobile phone with high accuracy and with minimal size and resource usage. The mobile application uses a smartphone camera to capture and deliver sequence input images to the deep learning model, which perform object recognition with distance estimation and provide voice comments that help the visually impaired understand the objects and obstacles in the outdoor environment. Furthermore, visually impaired people will receive more information (such as type of the different obstacle objects) about their environment and the mobile application will help them not only navigate in an unfamiliar environment, as well it will provide more information about the various obstacles with the estimated distance.

Normally, visually impaired people have different behaviors when they walk in some new areas. Therefore, our application provides two different scenario modes. The first mode is called stable mode, in which the walking human is in a completely unknown environment at first. The walking speeds of the human are usually reduced, and the human needs a navigation system with high accuracy of obstacle detection to provide safe navigation with especial voice comments to describe this environment. Visually

impaired people will therefore be able to become familiar with the environment very quickly. In the second scenario, it is outdoors in crowded environments where many obstacles and objects are encountered such as people, cars, fences, and buildings, in this case, the probability of users colliding with these obstacles is high. Thus, the navigation system requires high detection speed of the obstacles in the real system in order to inform the user about the different detected object. Therefore, this mode is referred to as Fast Mode according to the previous specifications of a navigation system. The main objective in this mode is to report a high number of detected objects in real time. The other major challenge is that most of the existing models and frameworks listed in Table 2 are not suitable for blind navigation because they do not contain all the necessary objects needed to ensure a very high perception for a blind navigation environment. For instance, visually impaired people have to understand obstacles such as walls, sidewalks, streets, buildings, fences, poles, and so on. However, none of these objects exists in the existing datasets and models. Our proposed approach provides an efficient approach to solve these issues, as it is designed to meet the requirements and the needs of visually impaired people. In addition, our model is trained on the pre-defined dataset that meets the blind person's needs in order to identify the set of objects needed for blind navigation with very high accuracy and recognition speed for individual walking situations.

4.1 Preparing Dataset

The identification of the appropriate dataset for the training and testing of an algorithm is an essential step in the development of a deep learning model. Using a sufficient dataset helps us to avoid or notice errors in our algorithm and to improve the results of our application. Although creating a custom dataset is time-consuming. However, it is possible to create a custom dataset, since the accuracy of the system depends on a good dataset. For this reason, our dataset is collected and defined to meet the requirements of our application. Our dataset contains 15 different categories that are necessary for my system. The model for the outdoor navigation system is trained from the scratch on our pre-defined dataset, as the objects required for the navigation of visually impaired people are not included in the pre-trained models. Three main steps have been performed to deliver a dataset as follows:

- **Data annotation**: In this step, the images are collected and labeled with the ground truth for semantic segmentation to differentiate all objects with a specific color pixel. These images are color-indexed. Each color index represents a unique class (with unique color). In addition, several segmentation classes are provided, as shown in the figure.
- **Index creation**: The objective of this step is to split the dataset into three parts (Train, Validation, and Test dataset).
- **TF Record Data Generation**: In this step, all images in TFRecord format are changed in order to read all images from TFRecord as binary data. This enables to accelerate training seep.

Figure 1 – Shows several color classes for object segmentation with specific objects

4.2 Training Model

The training model is done with DeepLabV3+ [27]. DeepLabV3+ is considered as one of the most powerful techniques for semantic image segmentation with Deep Learning which is implemented in the TensorFlow library published by Google researchers recently [29]. Semantic segmentation involves understanding the input image at the pixel level and then assigning a label to each pixel in the input image. Therefore, pixels with the similar label share certain properties. DeepLabV3+ offers a different approach to semantic segmentation. It demonstrates an architecture for manipulating signal decimation and learning multi-scale contextual features. In order to implement the DeepLabV3+ model in the system, it is necessary to focus on the following three components image segmentation, Atrous windings and Atrous Spatial Pyramid Pooling (ASPP) [28]. The ASPP module applies the same concept as the ASPP module of DeepLabv2, However, in a different way that leads to better performance. DeepLabV3+ employs ImageNet's pre-trained Resnet architecture with atrous convolutions as the major feature extractor. In the modified ResNet model, the last ResNet block utilizes atrous convolutions with varying dilatation rates. It applies Atrous Spatial Pyramid Pooling and bilinear upsampling to the decoder model on the modified ResNet block. DeepLab V3+ employs Aligned Xception as the major feature extractor with the three modifications:

- The maximal pooling operations are swapped by a depthwise separable convolution with striding.
- Additional batch normalization and ReLU activation function are provided after every 3 x 3 depthwise convolution.
- The depth of the model is extended without varying the structure of the entry flow network.

The figure2 shows the DeepLabv3+ model, which consists of two steps:

Encoding phase: The goal of this phase is to obtain and extract important information from the image. This is achieved by using a pre-trained Convoutional Neural Network.

Decoding phase: The information obtained in the encoding phase is utilized here to reconstruct the output of adequate dimensions.

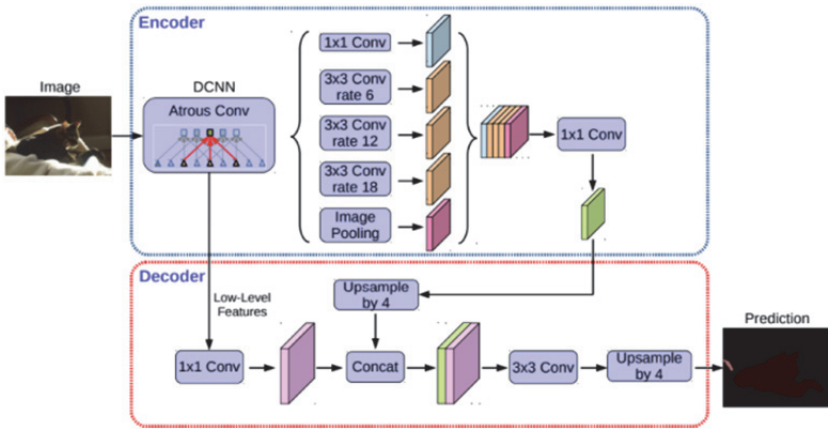

Figure 2 – Shows DeepLabV3+ Model Architecture

This model trained on the 2760 images in 15 different classes, after 230 steps in 26 hours of training on medium GPU specification which achieved about 0.78 accuracy with 1.4 losses. It is possible to minimize loss and improve the accuracy of the trained model by using a bigger dataset which required more training time. Figure 3 and figure 4 illustrate all the important factors of the training model, which are generated with TensorBoard during the training model. The trained graphic shown in figure 3 describes the specific values of a scalar-tensor that changes over time and iterations.

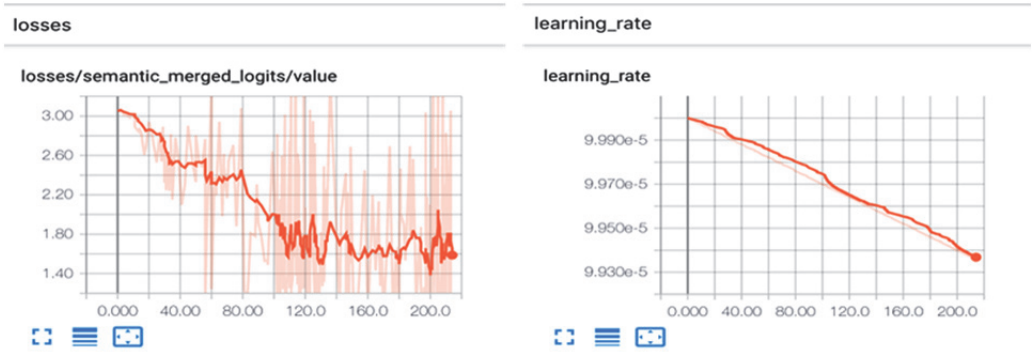

Figure 3 – shows the training model scalars graphs for learning and losses

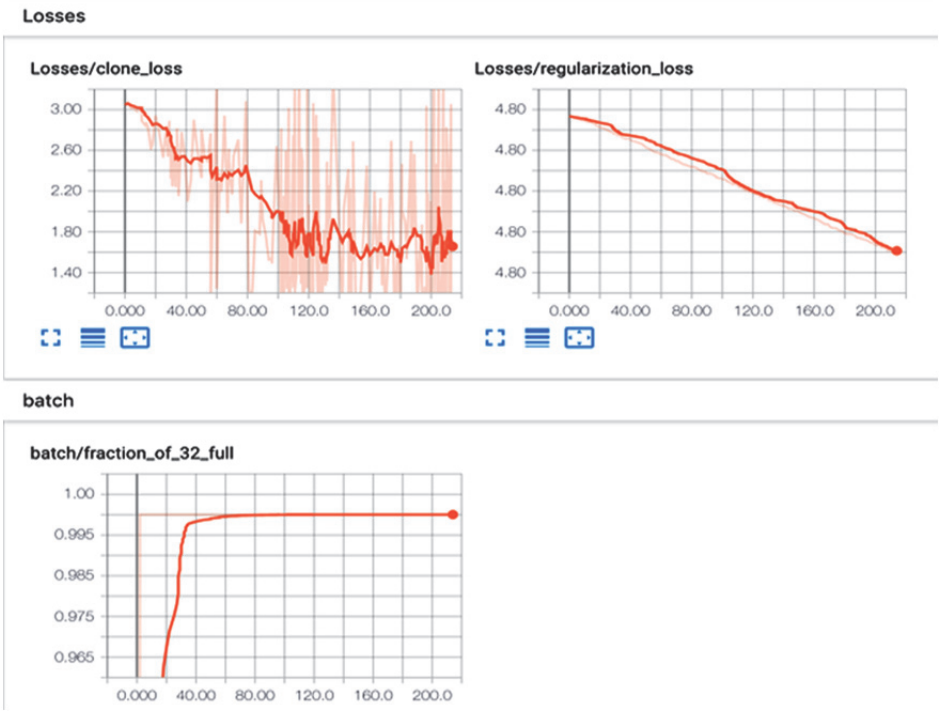

Figure 4 – model training scalars graph for losses, regularization loss, and batch.

The graph in figure 4 displays the changes in loss function and classification accuracy. The x-axis and y-axis represent the 200 steps and the corresponding values (random values from a standard normal distortion) of the variables. The graph describes the total training and loss after each training step of the trained model. The line in the figure 4 indicates the total training and loss after each training step of the trained model which was obtained by training the model on my own dataset, due to the limited number of trained images and the middle GPU specification, the result was good, however for a limited training scope it is an acceptable model with trained accuracy ~80% , validation accuracy ~79% and mean Intersection over Union (mIoU): ~66%.

4.3 Distance Estimation

The prediction of the distance from single images is managed as a regression problem [30]. The depth of detected obstacle can be defined as the distance from the camera pose to the obstacle. Several strategies can be applied to estimate the distance between the target and the camera position, such as Temporal method which calculates the distance based on the temporal sequence of the object copies in the images (e.g. visual odometry). Another approach is the Per-Frame, in which, the calculated distance in the actual frame is completely unrelated to previous frames (e.g: triangulation techniques or depth estimation based on CNN techniques). In this study, the distance estimation based on the per-frame triangulation method is implemented with some optimizations to reduce the errors in the distance calculation. In order to compute the distance between the detected object and the camera position, several parameters should be specified, such as the actual focal length of the camera lens F (mm), the actual height of the object Rh (mm), the camera frame height Fh (pixel)), the image height Ih (pixel) and the sensor height Is(mm). The formula for obtaining the distance of the detected object from the camera position is as follows

$$Distance = \frac{F * Rf * Fh}{Ih * Sh}$$

There is a geometric correlation between the focal length of a camera lens (F), the distance from the lens to the target object (O), and the distance between the lens and the projected image (I). The ratio between the distances shown in Figure 5 can be described as follows:

$$\frac{1}{F} = \frac{1}{I} + \frac{1}{O}$$

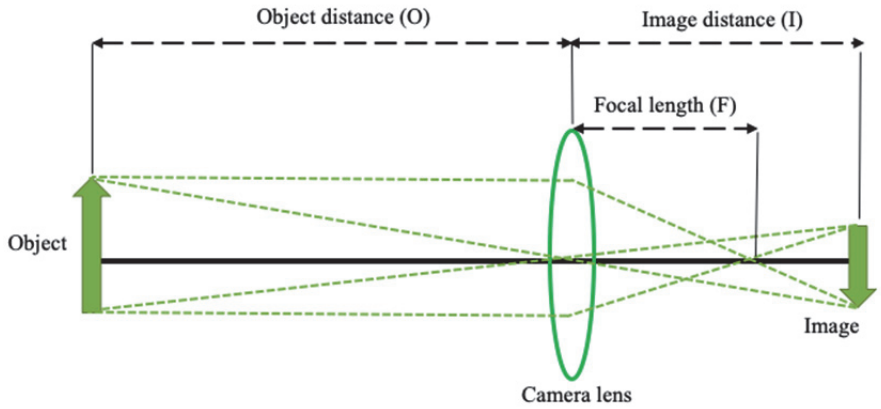

Figure 5 – shows how the image formation is used to estimate the distance

5 System Testing and Evaluation

The outdoor navigation IOS application was implemented and tested on the iPhone 6s with IOS 12.3.1 under the requirements for objects recognition, objects segmentation and distance estimation in real-time capabilities which runs at (23 FPS - 30 FPS). As a result, we have had several successful segmentation models with object recognition in a variety of environments. Figure 6 shows inference results for a set of images that are not included in the training set the model has implemented on the smartphone (iPhone 6S with IOS 12.3.1).

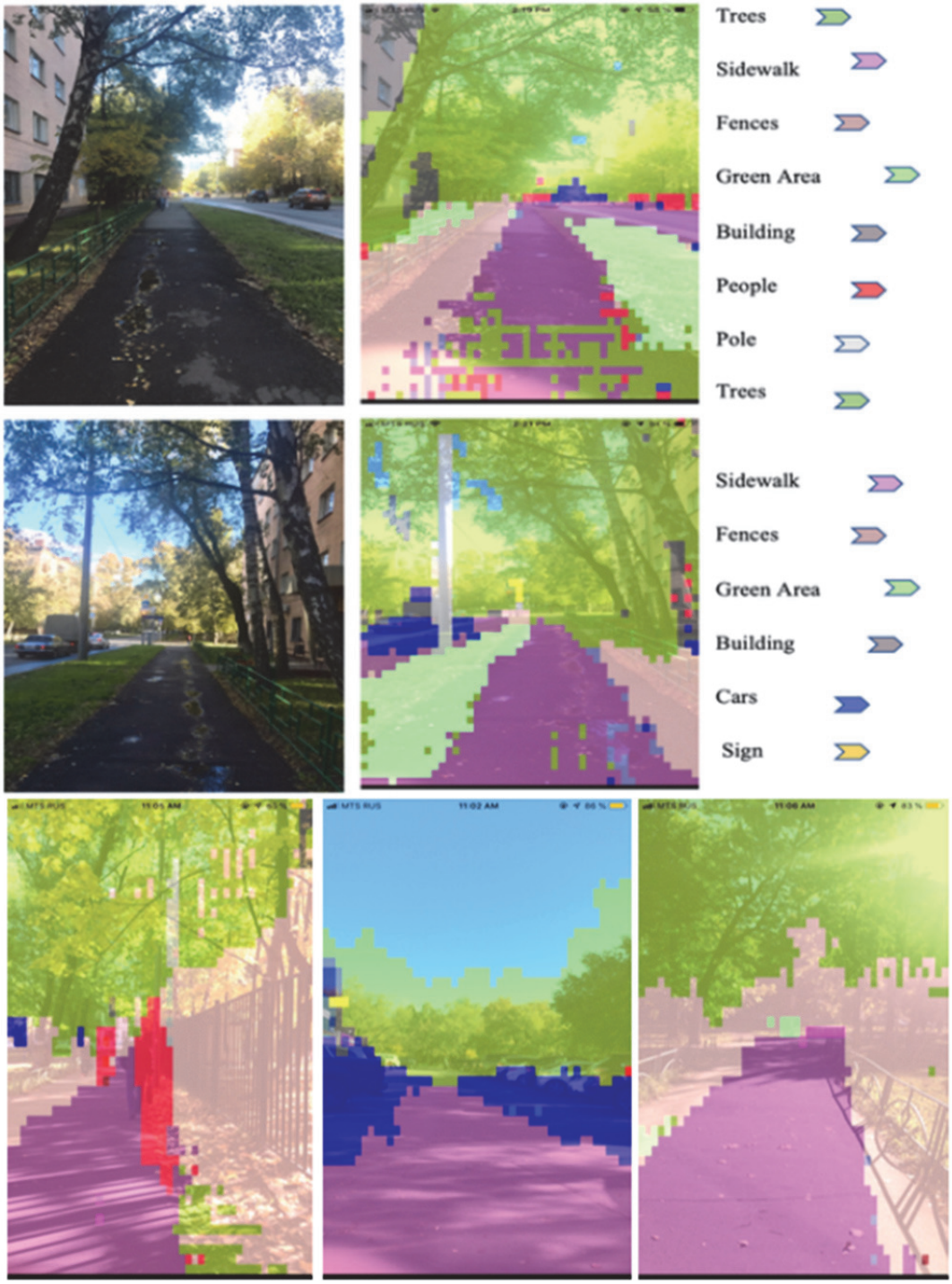

Figure 6 – the output of the outdoor navigation for visually impaired people displayed on the IOS smartphone, where different colors are assigned to different regions and recognized objects.

As shown in Figure 6, there are some side effects, for example the same pixels of color which bod sidewalk and road has it detected as the same color for both. The main reason behind that is due to a small dataset which is used to train our model. Therefore, the requirement to train our model with heterogeneous and large datasets will help us achieve optimal accuracy for mIoU and reduce the prediction error.

6 Conclusion

In this study, a novel outdoor navigation solution for visually impaired people is proposed which overcomes the limitations of other systems. The suggested solution is based on the utilization of a smartphone camera and deep learning algorithms to recognize different obstacles and objects with the estimated distance as well as to provide additional information to help the visually impaired to understand their environment. This approach can use walk voice guidance to alert users to the obstacles

in front of them for safe outdoor navigation. In this work, a new approach of object recognition is based on matching a set of pixels and adjusting these collections of pixels with the same color as output to different class probabilities. A single neural network predicts a set of pixels and class probabilities directly from full frames in one evaluation. Since the entire recognition pipeline is a single network, it allows end-to-end optimization directly on recognition performance. A smartphone camera is used to acquire continuous snapshots of the surrounding environment in front of a user and to perform image processing and object recognition to notify the user of the recognition outcomes. These results allow the user to gain a more comprehensive understanding of the surrounding environment. This system enables visually impaired people not only to know the rough direction and distance to an obstacle, as well as what the obstacle is.

6.1 Future Work

Currently, this system has been developed to navigate the visually impaired people with voice commentary and guidance, and the accuracy of the captured object with distance estimation was good and limited by a specific list of objects necessary for this system. For the future, the system should be expanded to include a larger number of objects with a larger dataset for the recognition of outdoor and indoor objects as well. The system is able to inform the visually impaired persons about different type of object. Therefore, impaired people understand what the objects are around and able to find the objects that they need in indoor as well as outdoor. The calculated distance should be improved, and the error minimized. Therefore, the depth information should be estimated in the meantime with objects recognition using a single monocular camera and a light neural architecture to predict pixel-wise depth map.

References

[1] A. Scheinwald, “Who Could Possibly Be Against A Treaty For The Blind?” , 22 Fordham Intell. Prop. Media & Ent.L.J. 445, 2012.

[2] Dan Jacobson, Rob Kitchin, Tommy Gärling, Reg Golledge, and Mark Blades, Learning A Complex Urban Route Without Sight: Comparing Naturalistic versus Laboratory Measures, University College Dublin, Ireland August 17-19, 1998.

[3] Oi-Mean Foong and Nurul Safwanah Bt Mohd Razali, Signage Recognition Framework for Visually Impaired People, 2011 International Conference on Computer Communication and Management Proc. of CSIT vol.5 (2011).

[4] A Report to the Nation by the National Federation of the Blind Jernigan Institute, The Braille Literacy Crisis in America, 2009.

[5] Kumar, G.A.; Ramakrishna, P.; Srinivasulu, M. Naveye A Guiding System for Blinds. IJIT 2015, 3, 593–598.

[6] Joseph, S.L., Xiao, J., Zhang, X., Subramaniam, L.V. Being Aware of the World: Toward Using Social Media to Support the Blind with Navigation. IEEE Trans. Hum. Mach. Syst. 2015, 45, 399–405.

[7] Huang, H.-C., Hsieh, C.-T., Yeh, C.-H. An Indoor Obstacle Detection System Using Depth Information and Region Growth. Sensors 2015, 15, 27116–27141.

[8] Kun (Linda) Li / PhD Candidate, EECS, UC Berkeley, Electronic Travel Aids for Blind Guidance, An Industry Landscape Study IND ENG 290 Project Report, 2015.

[9] Dimitrios Dakopoulos and Nikolaos G. Bourbakis, Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey, IEEE Transactions on Systems, Man, And Cybernetics—Part C: Applications and Reviews, Vol. 40, No. 1, January 2010.

[10] Bahadur, A.K., Tripathi, N. Design of Smart Voice Guiding and Location Indicator System for Visually Impaired and Disabled Person: The Artificial Vision System, GSM, GPRS, GPS, Cloud Computing. IJCTER 2016, 2, 29–35.

[11] Cecily Morrison, Kevin Doherty, Edward Cutrell, Anja Thieme, Imagining Artificial Intelligence Applications with People with Visual Disabilities using Tactile Ideation, ASSETS, Baltimore, MD, USA, 2017.

[12] Tao YU, Yusuke KUKI, Gento MATSUSHITA, Daiki MAEHARA, Design and implementation of lighting control system using battery-less wireless human detection sensor networks, arXiv:1704.04802v1 [cs.SY] 16 Apr 2017.

[13] Fares Jalled, Ilia Voronkov, Object Detection Using Image Processing, arXiv:1611.07791v1 [cs.CV] 23 Nov 2016.

[14] A. Mittal, A. Monnet, and N. Paragios, “Scene modeling and change detection in dynamic scenes: A subspace approach,” Comput. Vis. Image Underst., vol. 113, no. 1, pp. 63–79, Jan. 2009.

[15] Tianyi Liu, Shuangsang Fang, Yuehui Zhao, Peng Wang, Jun Zhang, Implementation of Training Convolutional Neural Networks, University of Chinese Academy of Sciences, Beijing, China, 2015.

[16] K. S. Ray, A. Chakraborty, and S. Dutta, “Detection, Recognition and Tracking of Moving Objects from Real-time Video via Visual Vocabulary Model and Species Inspired PSO,” Jun. 2017.

[17] Chris Hettinger, Tanner Christensen, Ben Ehlert, Jeffrey Humpherys, Tyler Jarvis, and Sean Wade, Forward Thinking: Building and Training Neural Networks One Layer at a Time, arXiv:1706.02480v1 [stat.ML] Jun 2017.

[18] Kristan M, Leonardis A, Jiri Matas, et al. The visual object tracking vot2016 challenge results. Proceedings of the European Conference on Computer Vision Workshops, 2016: 1-45.

[19] Zheng Zhu, Wei Wu, Wei Zou, Junjie Yan, End-to-end Flow Correlation Tracking with Spatial-temporal Attention, arXiv:1711.01124v4 [cs.CV] 27 Feb 2018.

[20] Vincent Angladon, Simone Gasparini, Vincent Charvillat, Tomislav Pribanić, Tomislav Petković, Matea Đonlić, Benjamin Ahsan, Frédéric Bruel, An evaluation of real-time-D visual odometry algorithms on mobile devices, HAL Id: hal-01654706, 2017.
[21] L.-C. Chen, G. Papandreou, S. Member, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs." arXiv:1606.00915v2, 2017.
[22] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages91–99, 2015.
[23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You Only Look Once: Unified, real-time object detection. arXiv:1506.02640, 2015.
[24] C. Szegedy, S. Reed, D. Erhan, and D. Anguelov. Scalable, high-quality object detection. arXiv:1412.1441, 2014.
[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. SSD: Single shot multibox detector. In European Conference on Computer Vision, pages 21–37. Springer, 2016.
[26] J. Dai, Y. Li, K. He, and J. Sun. r-fcn: Object detection via region-based fully convolutional networks. arXiv preprint arXiv:1605.06409, 2016.
[27] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," Feb. 2018.
[28] Manik Goyal, Param Rajpura, Hristo Bojinov and Ravi Hegde, Dataset Augmentation with Synthetic Images Improves Semantic Segmentation, arXiv:1709.00849v3, Jun 2018.
[29] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," Mar. 2016.
[30] Quoc V Le, Jiquan Ngiam, Adam Coates, Abhik Lahiri, Bobby Prochnow, and Andrew Y Ng. On optimization methods for deep learning. In Proceedings of the 28th International Conference on International Conference on Machine Learning, pages 265–272. Omnipress, 2011.