International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

Navigation Techniques for the Visually Impaired implemented using Deep Learning : A Survey

[1]Prabakaran S, [2]Samanvya Tripathi, [3]Utkarsh Nagpal

ABSTRACT—*Navigation plays a major role in the life of visually impaired people as they cannot get around easily without the assistance of escorts and hence, they must rely on their sense of touch and hearing in order to tackle the obstacles on their way. The use of guide dogs, sticks have not proven helpful. They are not able to make the visually impaired individuals depend on themselves. In order to solve these problems, there are sev- eral computing techniques explored for real-time navigation for visually impaired such as Artificial Intelligence, Deep Learning, Machine Learning. With the help of deep learning algorithms, it is easy to identify different objects and these algorithms can be implemented in the form of a mobile navigation application. In this paper, we review existing techniques related to Deep Learning and try to enhance the experience for the visually impaired.*

Keywords—*Deep Learning, Image Processing, Artificial Intelligence, Navigation, Visually Impaired, Survey*

I. INTRODUCTION

The process of navigation can be divided into two main tasks that is obstacle avoidance and environment-perception (safe, hostile). Obstacle avoidance can be further subdivided into sub-tasks like Obstacle Detection, Obstacle Distance- Detection and Obstacle approach-speed detection. For a person with normal sight, the tasks like crossing a street or walking on a busy walkway are generally easy and natural to carry out but a blind person can only use his sense of hearing, orientation and memory which makes it very difficult or close to impossible to navigate in a dynamically changing environment.

A. Evolution of Navigation for Visually Impaired (NAVI)

From “Electronic Travel Aids” which utilized Sonar sensors to "Artificial Vision" which uses state-of-the-art Computer Vision and Deep Learning Algorithms, the **Navigation Assis- tance for Visually Impaired (NAVI)** has seen a lot of devel- opment. In the beginning, canes with sensors were developed to help the **Visually Impaired (VI)** but that did not see a huge success because of it’s low accuracy and unpredictability. Then came along a haptic and auditory feedback of the surroundings which was welcome and improved the interaction of the VI with the environment. Along with sensors, GPS became a popular method to provide accurate navigation in an outdoor setting.

[1] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203, Tamil Nadu, India prabakas@srmist.edu.in
[2] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203, Tamil Nadu, India samanvyatripathi_amrendra@srmuniv.edu.in
[3]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203, Tamil Nadu, India utkarshnagpal_pankaj@srmuniv.edu.in

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

B. Deep Learning

Deep learning is a concept that has taken the world by a storm. It is a subdivision of Artficial Intelligence which involves algorithms enlivened by the working structure and functioning of the human brain. It involves teaching the com- puter to screen the inputs through a set of layers to determine the prediction and classification of the information. These deep learning algorithms are very powerful at detecting patterns and sequences sometimes even outperforming humans. The problem of dynamic object detection given an image can be solved using **Region Convolutional Neural Networks (R- CNN)** architecture.

It is a type of CNN that is used to identify an object in an image and also get information about its location within the image.

An RCNN model takes an input image, extracts the re- gion proposal and then only performs convolutions on those extracted regions of interest to help predict the label of the object in that region. The model also returns a set of "offset values" to help the bounding box fit more accurately and tightly. Other architectures like Fast-RCNN and Faster-RCNN are improvements made to this RCNN model.

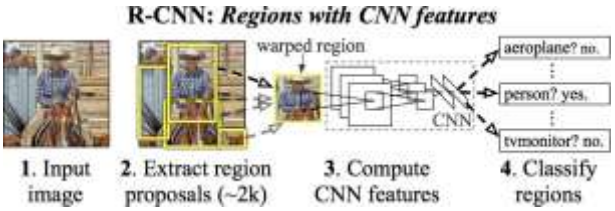

Figure 1: Overview of R-CNN

The YOLO model architecture is much different from that of the previously mentioned RCNN models. A single CNN predicts the bounding boxes and class probabilities for these boxes, which is done by splitting the image into an SxS Grid and the bounding boxes with a certain threshold are only used to detect the object in the image.

II. TECHNIQUES

As discussed earlier, the evolution of technological devices for the aid of VI has gone from sound and distance sensors to complex computer vision techniques. The advent of Deep Learning has paved way for simulating how a human brain interprets it's environment visually thus allowing researchers to come up with a holistic yet easy way to comprehend the

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

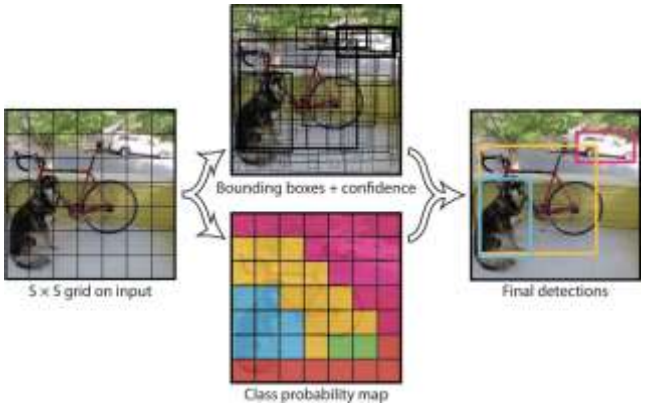

Figure 2: Overview of YOLO

surroundings. Object detection has been improved significantly with the help of algorithms like Faster-RCNN and YOLO. Wecan see the evolution of travel aids for the VI in Figure 3. The cost of first generation travel aids was higher because of its hardware dependencies which included the various sensors. Then came along the reduction in price of processors, increase of computing power at the edge and Artificial Intelligence which meant that anyone with a smartphone can run a state- of-the-art object detection algorithm on the device as well as utilise its GPS to navigate around. Although the accuracy and interaction with the apps can be improved, the path forward is promising.

Shadi Saleh et al [1] created an iOS application that runs on iPhone 6s and is able to perform image segmentation to classify the environment. Footpath, road, trees, poles, cars and buildings were segmented with good accuracy. The authors implement DeepLabV3+ model architecture with an encoder which extracts useful information about surrounding objects and then the decoder is used to reconstruct the original scene with the proper segmentation and colours. The use of DeepLabV3+ is explained in Figure 4. It clearly beats the other algorithms on "Test Time per Image" or Inference time and also the size of the model is drastically small compared to YOLOv2 which is also an added benefit making it easier to run on edge devices.

Alexy Bhowmick et al [3] made use of Microsoft Kinect's depth sensors to extract depth data along with RGB data which is then passed through SURF and BOVW model thereby making it a machine learning implementation. The labels predicted by an SVM Classfier is passed through TTS Engine and then sent to the VI through an earpiece. This paper presented a good way to use the depth data and combine it with object detection but lacked in the portability and speed aspect (Figure 6).

Baljit Kaur et al [4] represented a cost-effective scene perception system aimed towards visually impaired individual. It uses a laser to detect the distance and high quality webcam paired with a multicolumn CNN that is fine tuned to detect certain objects with a higher accuracy than others.

Xiaofei Fu et al [5] built a mobile assistant app for visually

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

impaired people. The app runs completely offline on the mobile device, it includes a deep convolutional neuron network for gender detection, and have a way to represent picture content with sound.

Ashwani Kumar et al [6] presented a Raspberry Pi so- lution for object detection by training a CNN model using the CIFAR-10 dataset and achieving 90% accuracy on the validation dataset. The predicted output labels were played back through the audio/haptic module.

Chucai Yi et al [7] focused their efforts on creating a very specialized object detection algorithm by training their algorithm on objects found inside homes like Refrigerator, Sink, Glasses, keys. According to them, a VI usually spends most of his/her time in a familiar environment and in-order to help them out with their daily tasks, the object detection algorithm was fine-tuned to detect thees daily objects by using SURF and SIFT, which finds objects based on pattern- recognition.

Rui Forest Jiang et al [8] implemented a visual-auditory experience for the VI by introducing 3-D bianural sound of the objects, which means that the further a person is from the detected object, the fainter the sound is and the closer he/she is, the louder. It uses the depth sensors from a Microsoft Kinect and implements Object Detection. The detected label along with spatial data is passed to a Unity Game Engine called 3Dception which renders the data using HRTF and finally produces and transmits this 3-D sound back to the user in about 15 seconds worst case. A dataflow pipeline can be seen in Figure 7 .

Nadia Kanwal et al [9] created a prototype using Kinect's RGB-D camera which provides general but important auditory information like "Stop", "Left", "Right" thereby providing the VI with a safe path. The corner detection and depth sensors together provide a robust model which can accurately predict objects and how far they are.

Ariadna Quattoni et al [10] talk about the importance of recognizing indoor scenes and that the object detection algorithms perform poorly when indoor because they are unable to use both global and local spatial information. In this paper they created a model that can do exactly that, ie. combining global and spatial information and use that to accurately describe an indoor environment. The innovative model was trained on the largest available dataset on indoor scenes with 67 different labels/categories, and it out-performed a state-of-the-art classifier.

There are also existing mobile applications in the market like TapTapSee [11] and Seeing AI [12] which implement tasks like Object Detection, Object Description and Anno- tation. Seeing AI goes a step further and adds a lot of useful modules within the app like Currecy Identification, Document-Orientation, Person/Friend including it's emotions, Color Descriptions and a Bar Code scanner.

Products like VISION-800, which is a wearable device (Figure 8) especially designed for the VI runs the software vOICe. vOICe sensory substitution technology for the blind is a state of the art software that converts Visual Data into an auditory video representation of the same input. This technique has a slight learning curve to it and may take some getting

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

Figure 3: Table Courtesy: [1]

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

Figure 4: Table Courtesy: [1]

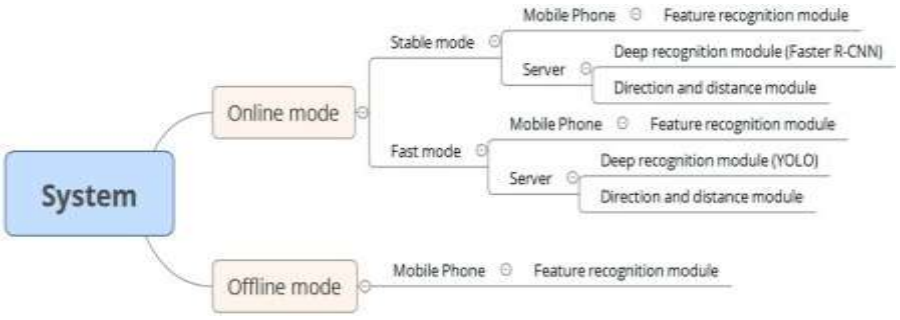

Figure 5: Relationships between various modules and modes, courtesy [2]

Figure 6: IntelliNavi [3] User-Description

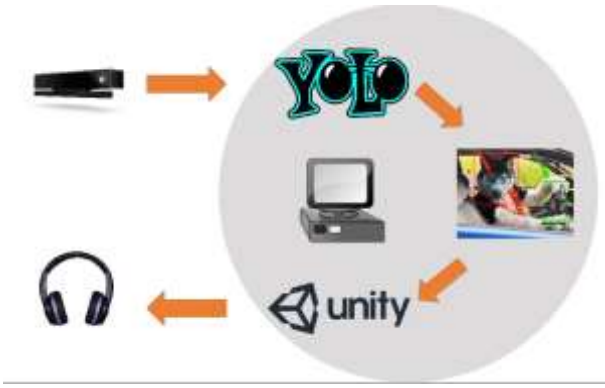

Figure 7: Dataflow pipeline of [8]

used to, but has showed great promise and gained reputation in the market for its effectiveness. [13]

III. NOMENCLATURE YOLO: You only Look Once

CNN: Convolution Neural Networks

RCNN: Recurrent Convolution Neural Networks FPS: Frames Per Second

RGB-D: Red Green Blue-Depth VI: Visually Impaired

NAVI: Navigation for Visually Impaired BOVW: Bag of Visual Words

SURF: Speeded Up Robust Features SIFT: Scale-Invariant Feature Transform HRTF: Head Related Transfer Function

Figure 8: VISION-800 Glasses

IV. CONCLUSION

In this paper, we reviewed existing techniques and the arrangement depends on the usage of a smartphone camera and the use of deep learning algorithms to detect various obstacles and objects with the assessed separation along with providing extra data to help the visually impaired to comprehend their condition. At present, this framework has been created to nav- igate the visually impaired individuals with voice commands and direction and the precision of the detected object with separation estimation was acceptable and constrained by a particular list of objects required for this framework. This methodology can utilize walk voice direction to make users aware of the obstacles before them for a guarded outdoor nav- igation. A smartphone camera is utilized to procure nonstop depictions of the general condition before a user and perform image processing and object acknowledgment to tell the user of the acknowledgment results. For the future, the framework is ought to be extended to incorporate a bigger number of objects with a bigger dataset for the acknowledgment of outside and indoor objects too. The framework can illuminate the visually impaired individuals about distinctive sort of objects. In this manner, visually impaired people comprehend the objects which are near and are ready to discover the objects that they need in indoor just as outside. The determined separation ought to be improved, and the mistake to be limited.

REFERENCES

1. S. Saleh, W. Hardt, and M. Nazari, “Outdoor Navigation for Visually Impaired based on Deep Learning,” 2019. [Online]. Available: http://ceur-ws.org/Vol-2514/paper102.pdf
2. B.-S. Lin, C.-C. Lee, and P.-Y. Chiang, “Simple Smartphone-Based Guiding System for Visually Impaired People,” *Sensors*, vol. 17, no. 6, pp. 1371–1371, 2017. [Online]. Available: 10.3390/s17061371;https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5492085/
3. A. Bhowmick, Prakash, Saurabh, Bhagat, Rukmani, V. Prasad, Hazarika, Shyamanta *et al.*, “IntelliNavi : Navigation for Blind Based on Kinect and Machine Learning,” pp. 172–183, 2014. [Online]. Available: 10.1007/978-3-319-13365-2_16;https:
4. //www.researchgate.net/publication/271328979_IntelliNavi_Navigation_ for_Blind_Based_on_Kinect_and_Machine_Learning
5. B. Kaur and J. Bhattacharya, “A scene perception system for visually impaired based on object detection and classification using multi-modal DCNN,” 2019. [Online]. Available: 10.1117/1.jei.28.1.013031

6. X. Fu, “Mobile assistant app for visually impaired people, with face detection, gender classification and sound representation of image.”
7. A. Kumar and A. Chourasia, “Blind Navigation System Using Artificial Intelligence,” *International Research Journal of Engineering and Technology*, 2018. [Online]. Available: https://www.irjet.net/archives/ V5/i3/IRJET-V5I3134.pdf
8. C. Yi, R. W. Flores, R. Chincha, and Y. Tian, “Finding objects for assisting blind people,” *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 2, pp. 71–79, 2013. [Online]. Available: 10.1007/s13721-013-0026-x;https://link.springer.com/article/ 10.1007/s13721-013-0026-x
9. . Rui, Q. F. Jiang, S. Lin, and Qu, “Let Blind People See : Real-Time Visual Recognition with Results Converted to 3D Audio,” 2016. [Online]. Available: https:
10. //www.researchgate.net/publication/312593672_Let_Blind_People_See_ Real-Time_Visual_Recognition_with_Results_Converted_to_3D_Audio
11. N. Kanwal, E. Bostanci, K. Currie, and A. F. Clark, “A Navigation System for the Visually Impaired : A Fusion of Vision and Depth Sensor,” 2015. [Online]. Available: 10.1155/2015/479857;https:
12. //www.mdpi.com/1424-8220/17/6/1371/htm
13. A. Quattoni and A. Torralba, “Recognizing indoor scenes,” 2009. [Online]. Available: 10.1109/CVPR.2009.5206537;https://people.csail. mit.edu/torralba/publications/indoor.pdf
14. “TapTapSee Mobile Application.” [Online]. Available: https://taptapseeapp.com
15. “Seeing AI Mobile Application.” [Online]. Available: https://www. microsoft.com/en-us/ai/seeing-ai
16. “vOICe : Seeing with Sound.” [Online]. Available: https://www. seeingwithsound.com