Mobile System for Guiding Blind Person Using CNN

[a]RanaJawadGhali , [b]Dr.Karim Q. Hussein

[a]Computer Science Dept. / College of Science / Mustansiriyah University , Baghdad, Iraq.
Ranajawad995@gmail.com.
[b]Assist Prof. / Computer Science Dept. / College of Science / Mustansiriyah University , Baghdad, Iraq.
Karimzzm@yahoo.com , karim.q.h@uomustansiriyah..edu.iq.

Abstract: The advancement of mobile technology has resulted in significant improvements in giving help to the blind person.The Glasses project assists blind people in detecting and recognizing things in their environment, which they see through a tiny camera, fixed on their glasses.This technique helps the blind person to be connected with the objects around himthrough a voice message sent to an earphone placed on the blind ear.The goal is to develop an intelligent system that can mimic the human eye.We utilize a tiny gadget called the "Raspberry Pi" for this purpose, which operates similarly to the human brain;with the help of the camera. Using deep learning algorithms; is known as Convolution Neural Network algorithm to recognize the objects.Moment the image's features are recognized finally, each object's sound is transmitted to tell the blind to the objects in front of him or her.Python was used to create this project. On the COCO dataset, the Blind CNN classifier attained a precision of 100 percent, according to the results.
Keywords: Deep learning, Convolutional Neural Network, Blind Person, Common Objects in Context.

1. Introduction

There are large numbers of visually impaired people, which led us to develop such system in order to help them to avoid obstructions. Smart technology has helped blind people in many different life aspects, such as ascending stairs, reading e-mails and using computers and mobile phones. According to World Health Organization (WHO) official statistics in 2009, there are 314 million visually impaired persons in the world in which 45 million of them are blind and 87% of them are from developing countries. **(Manoufali et al., 2011)**

Could you imagine how the life of a person who is blind could be? Many of them can't even walk without the help of others. Their life always depends upon their caregivers and can be quite difficult for them alone. The increasing number of people with disabilities in world attracts the concern of researchers to invent various technologies, hoping that these technologies can assist the disabled people in carrying out their tasks in everyday life like normal people. So we want to make something for them that would help them become independent. An open source smart glasses project is what we want to create. **(Abd & Ibrahim, n.d.)**

The aims of project of Blind assistance is promoting a widely challenge in computer vision such as recognition of objects of the surrounding objects practiced by the blind on a daily basis. The camera placed on blind person's glasses, MS COCO is large-scale object detection, segmentation, are employed to provide the necessary information about the surrounding environment. A dataset of objects gathered from daily scenes is created to apply the required recognition. Objects detection is used to find objects in the real world from an image of the world such as bicycles, chairs, doors, or tables that are common in the scenes of a blind. Based on their locations, and the camera is used to detect any objects. **(AL-Najjar et al., 2018)**

The basic concept of facial and object detection and identification system is a very commonly known factor. Not only for the aiding of visually impaired people, this notion is in implementation in many sectors such as security and industrial manufacturing. Different software system models are designed as such that it firstly takes the input images fetching from the database and implement the deep learning process to classify and then specifically identify the required result such as objects, facial identity or expression, forgery etc. for the real-time circumferences. **(Rahman et al., 2019)**

2. Related work

Here is some related work using different techniques for blind persons in the past few years:

(Iswahyudi et al., 2020), development of a Visual Aid Tool for Blind People Based on Faster R-CNN. This article focused on developing blind visual aid using a Faster R-CNN identifies an object. The Faster R-CNN is instilled on the raspberry pi equipped with a camera. The output is an audio signal about the object and its position refer to the users.

(Patil, 2017), Real time objects detection on a Raspberry Pi. This thesis explores the suitability of running object detection on the Raspberry Pi 3, a popular embedded computer board. Two controlled experiments are conducted where two state of the art object detection models SSD and YOLO are tested in how they perform in accuracy and speed. The accuracy for 4 meters in YOLO is 67% and the SSD is 93%.

(Kumar et al., 2019), An Object Detection Technique For Blind People in Real-Time Using Deep Neural Network. This paper proposed an object detection technique for blind people in real time to detect objects on any device running this model. Used single shot multi-box detector algorithm along with the help of architecture of faster region convolutional neural network. Uses standard VOC and COCO datasets. The accuracy of the model is more than 75%.

(Dharanidharan et al., 2019), Convolutional Neural Network for Object Detection System for Blind People. This work proposed a smart object detection system based on Convolutional Neural Network (CNN) to provide a smart and safe living for visually impaired people using Raspberry Pi. The results conclude that SSD has the highest mean Average Precision 73.7 compared to Faster R-CNN 73.2 and YOLO 63.4.

(AL-Najjar et al., 2018), Real Time Object Detection and Recognition for Blind People. The Smart Glasses project helps blind and visually impaired people to detection and recognition the office tools around them, which they see through a small camera, the camera is fixed on the glasses. By implement Neural Network on Raspberry Pi 3.

3. Pre-processing Images:

In this thesis used Microsoft COCO Common Objects in Context, Objects are labelled using per-instance segmentations to aid in precise object localization. Our dataset contains photos of 91 objects types that would be easily recognizable by a 4 year old. With a total of 2.5 million labelled instances in 328k images, **(Lin et al., 2014)**The data that used for training and test purpose is obtained by dividing the extracted corrections into 30% test data and 70% training data:

3.1 Conversion color to gray scale image:

Converting color image to gray-scale image is achieved for keeping the maximum information associated to original colored images. Extra information is needed regarding color images in the case when being converted into gray-scale. Each color in the pixel of an image is represented via 3 colors which are Red, Green, and Blue (RGB). Converting color image into gray-scale image means converting RGB value (24bits) to gray-scale value (8bits). Weighted method or Luminosity method could be used for converting to gray-scale using equation (2.1).

$$GY = 00.56G + 00.33R + 00.11B \quad (2.1)$$

According to the above-mentioned equation, Green color contributes 56%, which is more compared to the other colors and succeed Red color that contributes in (33%) and finally Blue color that contributes in 11%. **(Dhanoon & Ali, 2013)**

Figure.1showing the converting of colored image to gray scale.

3.2 Image Enhancement by histogram equalization:

In Digital Signal Processing (DSP), we store the number of pixels (frequencies) of same intensity values into a histogram array, which is commonly called "bin". For an 8-bit grayscale image, the size of histogram bin is 256, because the range of the intensity of 8-bit image is from 0 to 255. **(M et al., 2015)**

For digital image A with size of M ×N, we have

$$P(X) = \frac{\sum_{i=1}^{MXN} \delta(X,Xi)}{MXN}$$

Where $\delta(x, y) = \begin{cases} 1, x = y, \\ 0, x \neq y, \end{cases} \quad and \quad x \in [x_0, x_L],$

X representing the intensity. **(Cheng & Shi, 2004)**

Figure.2 converting the gray scale image to histogram equalization

3.3 Resize images based on bilinear interpolation:

Interpolated point is filled with four closest pixel's weighted average. In this method we performed two linear interpolations, in horizontal direction and then linear interpolation in vertical direction. We need to calculate four interpolation functions for grid point in Bilinear Interpolation. The interpolation kernel for linear interpolation is:

$$U(x) = \begin{matrix} \{0 & |x| > 1 \\ \{1 - |x| & |x| < 1 \end{matrix}$$

Where x = distance between interpolated point and grid point. **(Prajapati et al., 2012)**

The figure showed the image resized to 20x20 pixels.

Figure.3 20x20 resized image.

3.4 Features extraction PCA:

A technique called Two-dimensional PCA (2DPCA) (Yang and Zhang, 2004) (also referred to as image PCA (IMPCA) in a previous paper by Yang and Yang (2002)) was proposed to cut the computational cost of the standard PCA.

Unlike PCA that treats images as vectors, 2DPCA views an image as a matrix. With a proper criterion, 2DPCA results in an eigenvalue problem, but has a much lower dimensionality than PCA.**(Wang et al., 2005)**

$$J(x) = x^T \left[\sum_{i=1}^{n} A(i)^T A(i)\right] x.$$

4. Deep Learning

4.1 Convolution neural network CNN:

CNNs are hierarchical neural networks whose convolutional layers alternate with subsampling layers, reminiscent of simple and complex cells in the primary visual cortex [Wiesel and Hubel, 1959]. CNNs vary in how convolutional and subsampling layers are realized and how the nets are trained. **(Cires et al., 2003)**

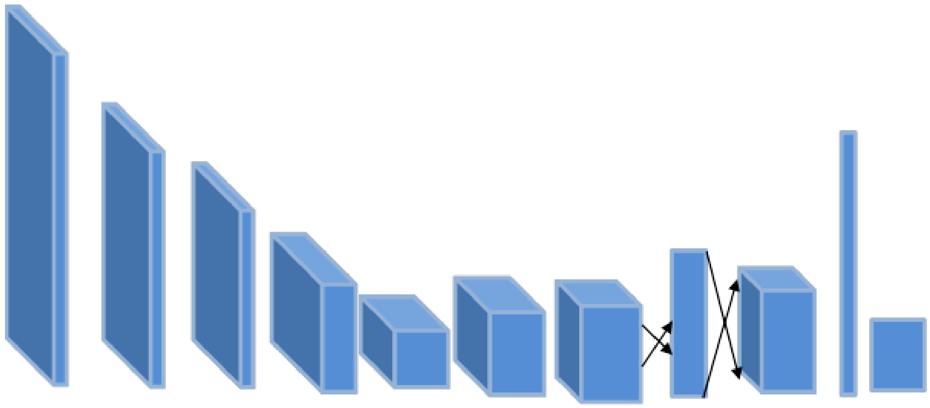


Figure.4 showing the CNN model layers.

Table.1 showing every layer values in the CNN model.

Fi=16	Fi =32	Fi=64	Fi =128	Fi =256	Fi =512	Fi=1024	Fi=1024	Fi=125	Flatten	Dense
Conv1D S=1 K=3 Maxpool S=1 Poolsize= 1	Conv1D S=1 K=3 Maxpool S=1 Poolsize= 1	Conv1D S=1 K=3 Maxpool S=1 Poolsize= 1	Conv1D S=1 K=3 Maxpool S=1 Poolsize= 1	Conv1D S=1 K=3 Maxpool S=1 Poolsize= 1	Conv1D S=1 K=3 Maxpool S=1 Poolsize= 1	Conv1D S=1 K=3 Maxpool S=1 Poolsize= 1	ConvlD S=1 K=3	Conv1D S=1 K=3		softmax

5. Results

On the same dataset COCO tried working with machine learning SGD and DT and the results from testing and training with deep learning show in the figure:

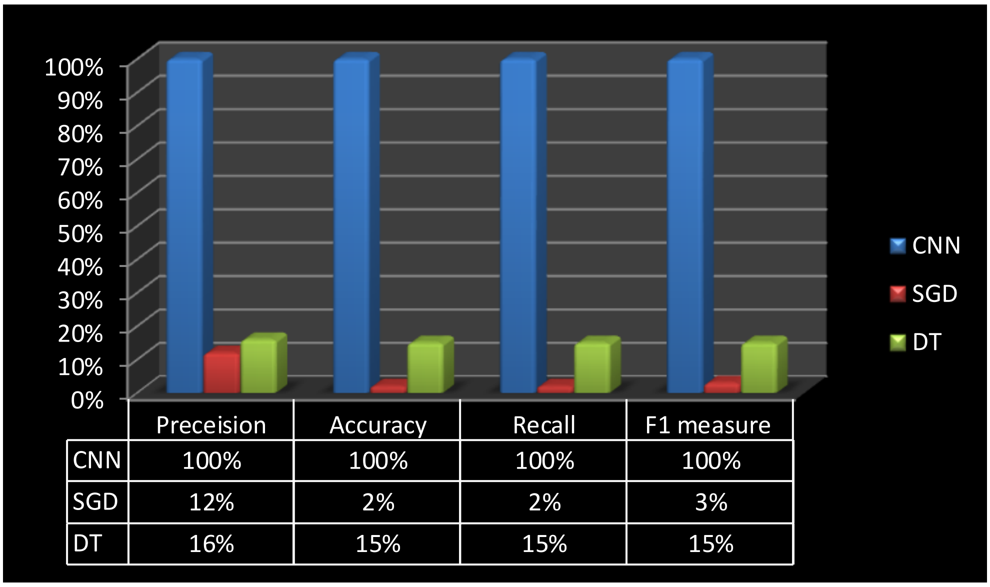
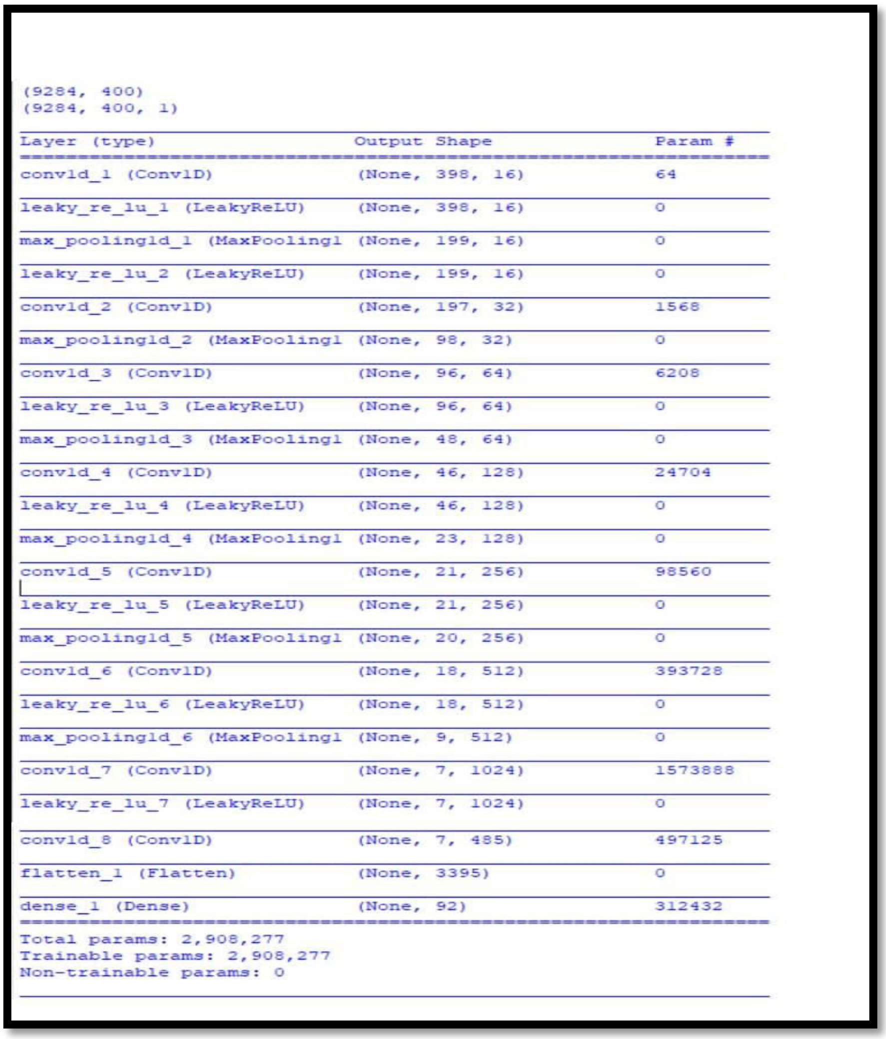


Figure.5 showing the results of objects detection after using different techniques.

Where show that the results from using CNN are high (precision, accuracy, recall and F1 measure = 100%) And SGD precision 12%, accuracy 2%, Recall 2% and F1 measure 3%.
While the DT precision 16%, accuracy 15%, Recall 15% and F1 measure 15%.



```
(9284, 400)
(9284, 400, 1)
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_1 (Conv1D)            (None, 398, 16)           64
leaky_re_lu_1 (LeakyReLU)    (None, 398, 16)           0
max_pooling1d_1 (MaxPooling1 (None, 199, 16)           0
leaky_re_lu_2 (LeakyReLU)    (None, 199, 16)           0
conv1d_2 (Conv1D)            (None, 197, 32)           1568
max_pooling1d_2 (MaxPooling1 (None, 98, 32)            0
conv1d_3 (Conv1D)            (None, 96, 64)            6208
leaky_re_lu_3 (LeakyReLU)    (None, 96, 64)            0
max_pooling1d_3 (MaxPooling1 (None, 48, 64)            0
conv1d_4 (Conv1D)            (None, 46, 128)           24704
leaky_re_lu_4 (LeakyReLU)    (None, 46, 128)           0
max_pooling1d_4 (MaxPooling1 (None, 23, 128)           0
conv1d_5 (Conv1D)            (None, 21, 256)           98560
leaky_re_lu_5 (LeakyReLU)    (None, 21, 256)           0
max_pooling1d_5 (MaxPooling1 (None, 20, 256)           0
conv1d_6 (Conv1D)            (None, 18, 512)           393728
leaky_re_lu_6 (LeakyReLU)    (None, 18, 512)           0
max_pooling1d_6 (MaxPooling1 (None, 9, 512)            0
conv1d_7 (Conv1D)            (None, 7, 1024)           1573888
leaky_re_lu_7 (LeakyReLU)    (None, 7, 1024)           0
conv1d_8 (Conv1D)            (None, 7, 485)            497125
flatten_1 (Flatten)          (None, 3395)              0
dense_1 (Dense)              (None, 92)                312432
=================================================================
Total params: 2,908,277
Trainable params: 2,908,277
Non-trainable params: 0
```

Figure.6 showing the outputs of each layer.

6. Conclusions and Future Studies:

In this thesis the researcher attempt to implement a system for a blind person in the indoor environment to make them self-reliant and can easily move inside the home day-to-day activity with the help of the system and a little work has been done in this field. During this project, developed a CNN model for classify objects image. Model was trained with greatly dataset. The accuracy of the model reached 100%.

This project will be updated by developing the model to be capable to work in dynamic environment with high accurate and provide the blind person with more specifications.

7.Acknowledgment:

The authors would like to expose their thanks and acknowledgments to Al Mustansiriyah University – Baghdad IARQ, for supporting and encouragement. Thank you.

References

Abd, Y., & Ibrahim, E. A. (n.d.). *Graduation Project F inal R eport*.

AL-Najjar, M., Suliman, I., & Al-Hanini, G. (2018). *Real Time Object Detection and Recognition for Blind People*.

Cheng, H. D., & Shi, X. J. (2004). A simple and effective histogram equalization approach to image enhancement. *Digital Signal Processing: A Review Journal*, *14*(2), 158–170. https://doi.org/10.1016/j.dsp.2003.07.002

Cires, D. C., Meier, U., Masci, J., & Gambardella, L. M. (2003). Flexible, High Performance Convolutional Neural Networks for Image Classificatio. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence Flexible*, 1237–1242. https://www.aaai.org/ocs/index.php/IJCAI/IJCAI11/paper/viewFile/3098/3425

Dhanoon, N., & Ali, H. H. (2013). Handwritten Hindi numerals recognition. *International Journal of Innovation and Applied Studies*, *3*(1), 310–317.

Dharanidharan, J., Puviarasi, R., & Boselin Prabhu, S. R. (2019). Object detection system for blind people. *International Journal of Recent Technology and Engineering*, *8*(2 Special Issue 8), 1675–1676. https://doi.org/10.35940/ijrte.B1129.0882S819

Iswahyudi, Anam, K., & Sujanarko, B. (2020). Development of a visual aid tool for blind people based on faster R-CNN. *AIP Conference Proceedings*, *2278*(October). https://doi.org/10.1063/5.0014538

Kumar, A., Reddy, S. S. S. S., & Kulkarni, V. (2019). An Object Detection Technique For Blind People in Real-Time Using Deep Neural Network. *Proceedings of the IEEE International Conference Image Information Processing*, *2019-November*, 292–297. https://doi.org/10.1109/ICIIP47207.2019.8985965

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8693 LNCS*(PART 5), 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

M, J. R., Rajendran, S., Dorothy, R., Joany, R., Joseph Rathish, R., Santhana Prabha, S., & Rajendran, S. (2015). Image enhancement by Histogram equalization. *Int. J. Nano. Corr. Sci. Engg*, *2*(4), 21–30. https://www.researchgate.net/publication/283727396

Manoufali, M., Aladwani, A., Alseraidy, S., & Alabdouli, A. (2011). Smart guide for blind people. *Proceedings of the 2011 International Conference and Workshop on the Current Trends in Information Technology, CTIT'11*, *January*, 61–63. https://doi.org/10.1109/CTIT.2011.6107935

Patil, S. (2017). Real time object detection on a Raspberry Pi. *Electronicwings*, *V*(Vi), 103–106. https://www.electronicwings.com/raspberry-pi/raspberry-pi-gpio-access

Prajapati, A., Naik, S., & Mehta, S. (2012). Evaluation of Different Image Interpolation Algorithms. *International Journal of Computer Applications*, *58*(12), 6–12. https://doi.org/10.5120/9332-3638

Rahman, F., Ritun, I. J., Farhin, N., & Uddin, J. (2019). AnAssistive model for visually impaired people using YOLO and MTCNN. *ACM International Conference Proceeding Series*, 225–230. https://doi.org/10.1145/3309074.3309114

Wang, L., Wang, X., Zhang, X., & Feng, J. (2005). The equivalence of two-dimensional PCA to line-based PCA. *Pattern Recognition Letters*, *26*(1), 57–60. https://doi.org/10.1016/j.patrec.2004.08.016