

# AUDIO ASSISTED ELECTRONIC GLASSES FOR BLIND & VISUALLY IMPAIRED PEOPLE USING DEEP LEARNING

<sup>1</sup>Khilan Pandya and <sup>2</sup>Dr. Bhargav C. Goradiya

<sup>1</sup>Research Analyst at Numerator

khilanpandya@gmail.com

<sup>2</sup>Professor – BVM Engineering College, V.V.Nagar

bhargav.goradiya@bvmengineering.ac.in

## Abstract

*Encroachment of technology has replaced individuals in almost every field with machines. By introducing machines, automation system has reduced human workload, with specially focused on visually impaired person or completely blind person find difficulties in reading a printing or handwritten text from real world or distinguishing people in front of them. So, they cannot perform particular tasks without someone's assistance or help. Hence looking at this challenge, it is necessary to make the reading device for blind people which is capable of recognizing text, human objects and, differentiate the bank currencies across the world. Obviously, Output will be in the form of audio, which will help the blind user to read the texture like a normal person. To overcome this difficulty for the visually weakened group, this paper presents a device which will help blind people for giving them guidance efficiently and safely to read book, recognize faces of their near and dear ones as well recognize different bank currencies. In this paper, we propose a deep neural network algorithm like CNN integrating with OCR model achieves a recognition accuracy of 95% for human detection and 90% for disgusting bank currencies.*

**Keywords:** Deep Learning, CNN, Electronic Glasses, Visually Impaired, OCR.

## I. Introduction

Blindness has been a malady which has been prevalent for centuries, there are a plethora of people suffering from this condition. We cannot even imagine the problems they are going through. They face many challenges in their life as the gift of eyesight is denuded from them. Following are the causes for blindness: 1) Glaucoma 2) Diabetic Retinopathy 3) Cataract 4) Age-Related Macular Degeneration 5) Refractive Errors 6) Amblyopia and 7) Strabismus etc.

With modern advancement we as engineers strive to facilitate them in any mean possible. These are some facts released by WHO (World Health Organization): 1) An estimated 253 million people have vision impairment: 36 million are blind and 217 million have moderate to severe vision impairment. Out of these over 15 million blind people are in India. 2) 81% of people who have blindness or have moderate or severe vision impairment are aged 50 years and above.

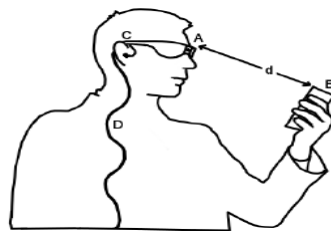
Furthermore, visually weakened people face many encounters in their every-day life, and a number of these could be relieved to some degree via the use of an aiding device based on artificial vision systems [1]. One particular problem is such that determining the denomination of a bank currencies,

difficulties in reading a printing or handwritten text from real environment and distinguishing people in front of them as an object. This is very useful for a blind person that is on their own because recognizing the any real-world running objects which are useful in daily life for blind people is quite challenging. Devices like a glass openly considered to perform this task already exist on the market but the current trend in technological development is to take advantage of cameras found in smartphones, portable computing devices (tablet-like devices), and wearable devices that offer a number of complementing solutions for the user [1]. Today, in the digital world mobile applications are being developed technocrats and some have already been released into the market to help social. However, but it is found that there is a lack of advancement in the current technology is possible and necessary.

In this paper we present a simple and robust method for the identification of real-world objects, which is part of a wearable glasses for helping the blind. The scope of this glasses is that it guides the visually impaired person in reading independently in an efficient manner. These days if they must read anything, firstly it needs be written in braille script for them to feel and understand it. Whereas this glass will help them interpret anything just with help of a click. The functionalities of this device are:

- It detects the human face which is standing in front of it. We first need to scan and load it into the database.
- It also detects the currency in the same way.
- It can convert written text into phonetic text.

Fig. 1 shows basic Aiding system for the blind and its use for object recognition. In that A) camera mounted on a wearable frame B) Object C) Headphone and D) Cable connecting the wearable system to the processing unit.



**Figure 1:** Aiding System for Visually Impaired People [1]

The rest of the paper is divided as follow. Section 2 presents the related work. The State-of-Art of proposed system discussed in Section 3. The methodology which includes the CNN technique and OCR model is presented in Section 4. Section 5 consists of various simulation results. The paper is concluded in Section 6.

## II. Related Work

In this section, first we present a systematic survey of the state-of-the-art electronics glass architectures and its working. In [1], authors propose a novel method for automated identification of banknotes' denominations based on image processing is presented. This method is part of a wearable aiding system for the visually impaired, and it uses a standard video camera as the image collecting device. The method first extracts points of interest from the denomination region of a banknote and then performs an analysis of the geometrical patterns so defined, which allows the identification of the banknote denomination. Experiments were performed with a test-subject in order to simulate real-world operating conditions. A high average identification rate was achieved by this method [1].

Advancement of Technology has replaced humans in almost every field with machines was the first point of the authors [2]. By introducing machines, banking automation has enormously reduced human workload very proficiently. It is imperative to take care while handling currency, which is

reduced by automation of banking. The identification of the currency value is hard when currency notes are blurry or damaged or withered. Convoluted designs are included to improve and increase security of currency. This certainly makes the task of currency recognition very difficult. To correctly recognize a currency, it is very imminent to choose the proper features and suitable algorithm. In this particular method [2], Canny Edge Detector is used for segmentation and for classification, NN pattern recognition tool is used which gives 95.6% accuracy.

In [3], authors describe the development of a human face recognition system (HFRS) using Multilayer Perceptron Artificial Neural Networks (MLP-NN). It is trained with a set of face images until it is in a "learned" state. The network is able of classifying the face input into its class [3]. In the case of the subject face is not one of those trained, the network will register it as anon. This system takes the face image as input from video camera, is also developed to notice the presence of an object in front of the camera and to search for the human face area automatically. The identified face area will then be used as the inputs to the neural network to perform recognition [3].

The paper [4] presents a more hardware-based solution to the problem of the visually weakened people in which an android system integrated with camera and USB laser is attached to the chest of the user. Device detects the classified image and its distance from the user providing with it a voice output.

Authors in [5] present a novel technique where visually impaired people are able to capture image in their android phone and through an application which is running on the deep learning principles. They have used VGG16 Model and used Flickr8k dataset [5] to train the model in generating output. So, system takes an image runs through a CNN module which encodes the image in different aspect and then using RNN and LSTM it generates a caption for the image. RNN is for the generating the highest probability of a word and LSTM for the complete caption.

In [6], authors make an effort to prepare a complete solution which helps blind or visually impaired people to navigate through an area which has obstacles. They have included an RGB-Depth camera which is used at 30 degrees to get the location of the obstacles using the floor segmentation method. Here they had an issue with transparent objects like glass or french door. So, to remove this issue they have also included ultraviolet sensors which transmit 8 cycles of ultrasonic burst at 40 KHz and wait for the reflected burst. It also has AR gear so that people who have not completely lost their vision can be helped. It uses audio for guiding the user. It also has feature of tactile feedback. Moreover, this paper has provided very good comparison on various data acquisition techniques.

In [7], authors have not prepared any hardware. It is more of a software-based paper. They have used Alexnet for training and testing. They have used two approaches 1) Transfer Learning and 2) Feature Extraction. Their major purpose was to make a hospitals and clinic for blind or visually impaired people friendly. So, authors have created their own database by taking real time pictures of a hospital doors, signs etc. They created the labels and gaussian noise images themselves. In this regard a pre-trained Alexnet model has been fine-tuned using the database they generated. Majority of the images were taken for training and remaining were taken for test set. Learning rate was kept 0.0001, epoch of 20 and minimum batch-size of 128. They have made comparison between TL and FEX. They are able to get good precision of almost up to 98%.

In [8], It has used 6-axis MEM (micro-electromechanical sensor module). Infrared transceiver has been used to detect object. It also has Wi-Fi and Bluetooth connectivity in the proposed solution. With the help of MEM, it detects an object and in case the blind person collides with an object then a message will be sent to guardian or caretaker for the same, and that location will be marked. For messaging and location purpose authors have also generated an application.

Object detection performance, as measured on the canonical PASCAL VOC [9] dataset, has improved tremendously in the recent years. The better performing techniques are complex ensemble systems that typically combine numerous low-level image features with high-level context. In [9], authors propose a simple and scalable object detection algorithm that helps to refit mean average precision (mAP) by more than 30% relative to the previous best result on VOC 2012—achieving a

mAP of 53.3%. Their approach combines two insights: first of all, one can apply high-capacity CNN to bottom-up region proposals in order to localize and segment objects and secondly when labeled training data is very less, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, yields a significant performance increase. Since they combine region proposals with CNNs, they call their method R-CNN [9]: Regions with CNN features.

### III. State-of-Art Proposed System

#### Single Board PC: NVIDIA Jetson TK1 Developer Kit:

This hardware is much useful as it is very handy when it comes to training and running CNN models. It is also lightweight, easy to wear and portable. It also supports CUDA which will help in training the model more efficiently and quickly. The NVIDIA Jetson TK1 development kit is a full-featured platform for Tegra K1 embedded applications. With the power of 192 CUDA cores, it will use to develop cutting-edge solutions in Computer Vision, robotics, medicine, security, and automotive. Fig. 2 Shows wearable NVIDIA Jetson developer kit.



Figure 2: NVIDIA Jetson TK1 Development Kit

#### Camera: SainSmart IMX219 AI Camera

This camera module is designed for NVIDIA Jetson Nano Board. It is with 8-megapixel IMX219 sensor and features the 85-degree field of view and supports 1080p at 30 fps, 720p at 60 fps and 640 x480p 60/90 video recording. With the power interface, it also supports 3.3V output power supply. Fig. 3 shows lightweight high-resolution camera which works very great in real-time.



Figure 3: SainSmart IMX219 AI Camera

#### System Block Diagram and System Flow Chart

Fig. 4 shows block diagram explains the basic view of the hardware connection in the simplest manner. Flow chart of system gives a brief understanding of the process which is followed shown in fig. 5.

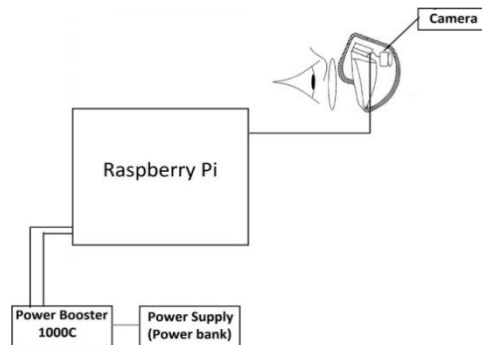


Figure 4: System Block Diagram

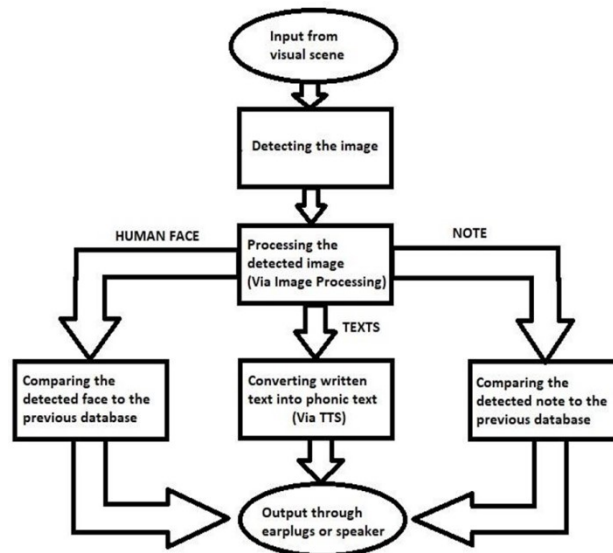


Figure 5: Working Process of System Flow Chart

#### IV. CNN Model and OCR

CNN stands for Convolutional Neural Network; it is one of the easiest yet a very powerful neural network which can be used for object detection or classification process. The term "deep" is used as there are many hidden layers in the neural network. It can contain up to 150 layers. These layers are trained using a large labelled dataset. One of the easiest ways of tagging the database is to save the different images which are to be classified or detected in different folders. This process makes them different classes. This way the model learns features directly from the data and it does not require manual feature extraction. A CNN convolves features with input and uses 2-dimensional layers; making this architecture well suited for 2-dimensional data such as images. It has the automatic feature extraction which is learned when the model is fed a collection of images. The working of CNN shown in fig. 6.

Tesseract is on the most powerful and advanced OCR software can be used easily and provides almost error-free results. It is one the most efficient way to convert image to text. It can also be

executed from command line interface.

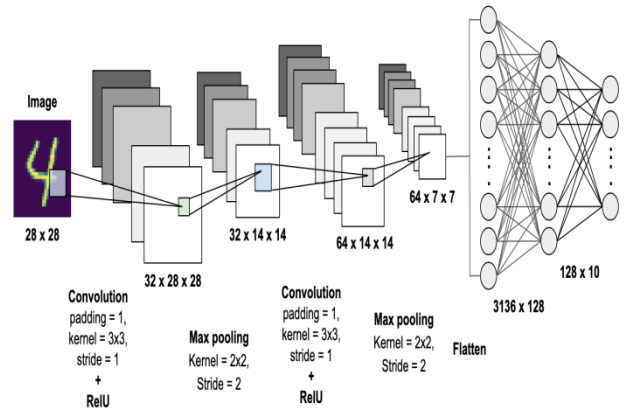


Figure 6: CNN Model

### Dataset

First, to train neural network model, it needs to make a dataset. There are ways to make a database, it can even download a whole database or even download weights of a pre-trained model for the prediction purpose, here we have made a data base using images from the source IMAGE-NET. For that we have captured the images using the camera for face recognition part. For the currency recognition we have used the online database.



Figure 7: Database

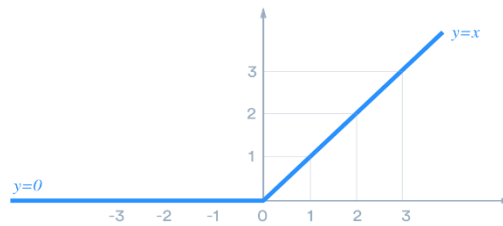
### Loading a Dataset and Information of the Model

**Loading a Dataset:** We are able to make a dataset of 2000 images which has 500 of both the categories. (Here we have used only 500 per category otherwise it can be made of even more images.)

**Model:** Now will move forward towards building a convolutional neural network. Input of my model is 3x100x100. (3 as RGB and 100x100 is the size of my image). We have used 4 layers for both the face and currency recognition. Given image gives more information regarding the model. Max pool layer helps in making the feature detection independent of noise and trivial changes like image rotation or tilting, shown in fig. 8. It is based on the sliding window concept, in which it applies a statistical function over the values which are specified as window also known as kernel. Max pooling takes the maximum value within the filter of convolution. ReLU is short for REctified Linear Unit and is a type of activation function shown in fig. 9. Mathematically, it is defined as  $y = \max(0, x)$ .



**Figure 8:** Maxpooling Method in CNN



**Figure 9:** RELU Activation Function

Linear layer is used to convert the data into linear function which then makes it much easier for classification.

## V. Simulation Results

To implement this model, we have used following simulation tool shown in table 1.

Table I. Simulation Tools

Model	Simulation Tools
CNN	Pytorch
Computer Vision	OpenCV
OCR	Tesseract

We have defined various parameters such as learning rate, momentum of that learning rate, type of optimizer, type of loss function etc. Training the model is the most important step. Here we have trained the model twice in the same epoch that is in one iteration of loop with the training dataset and cross-validation dataset. Adding further we found the accuracy and loss curve to fine tune our model.

Once the model is trained, we can use it to predict the different faces and currency. We have used the same model for the both the purpose as it works perfectly in both the instances. For currency we have used dataset of 10, 20, 50, 100 rupees notes (Indian Currency). Here, 25-epoch, Adam optimizer with 0.001 learning rate and Cross Entropy Loss used for face and currency recognition model. Accuracy of face detection and currency recognition is around 95% and 90%, respectively shown in fig. 10 and fig.11.

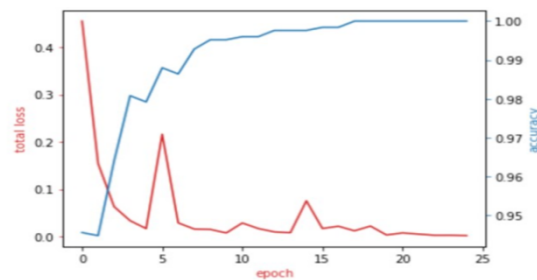


Figure 10: Face Recognition

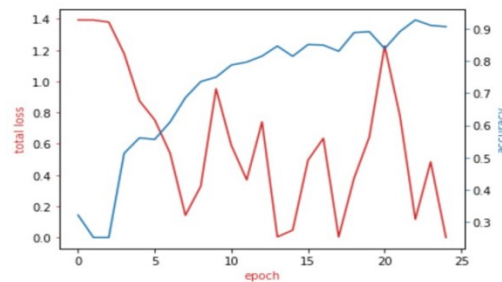


Figure 11: Currency Recognition

## VI. Conclusion and Future work

This paper shows a novel methodology, CNN with OCR for recognition of face and currency denominations via image processing. The proposed method is fast, efficient and robust with respect to the way in which the object is presented to the system. Simulation is conducted over 500 images. Here, database is trained using CNN, which gives 95% and 90% classification accuracy for face detection and currency recognition respectively. In the future we have planned to design improvements with the objective of making it even more efficient and robust by currencies of different countries can be added which makes it applicable globally and instead of using OCR a model, RNN and NLP engine can be made to complete the task of image to text.

## References

- [1] Rojas-Domínguez, A., Lara-Alvarez, C., & Bayro-Corrochano, E. (2014). Automated banknote identification method for the visually impaired. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8827, 572–579. [https://doi.org/10.1007/978-3-319-12568-8\\_70](https://doi.org/10.1007/978-3-319-12568-8_70).
- [2] Patel, V. N., Jaliya, U. K., & Brahmhatt, K. N. (2017). Indian Currency Recognition using Neural Network Pattern Recognition Tool. 2(April 2014), 67–72.
- [3] M. H. Ahmad Fadzil and H. Abu Bakar, "Human face recognition using neural networks," *Proceedings of 1st International Conference on Image Processing, Austin, TX, 1994*, pp. 936-939 vol.3, doi: 10.1109/ICIP.1994.413708.
- [4] Baljit Kaur and Jhilik Bhattacharya "Scene perception system for visually impaired based on object detection and classification using multimodal deep convolutional neural network," *Journal of Electronic Imaging* 28(1), 013031 (8 February 2019). <https://doi.org/10.1117/1.JEI.28.1.013031>.



- [5] Sans R.K., Joseph R.S., Narayanan R., Prasad V.M., James J. (2019) AUGEN: An Ocular Support for Visually Impaired Using Deep Learning. In: Pandian D., Fernando X., Baig Z., Shi F. (eds) Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB). ISMAC 2018. Lecture Notes in Computational Vision and Biomechanics, vol 30. Springer, Cham. [https://doi.org/10.1007/978-3-030-00665-5\\_124](https://doi.org/10.1007/978-3-030-00665-5_124)
- [6] J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu, "Smart guiding glasses for visually impaired people in indoor environment," in IEEE Transactions on Consumer Electronics, vol. 63, no. 3, pp. 258-266, August 2017, doi: 10.1109/TCE.2017.014980.
- [7] Bashiri, Fereshteh S., Eric LaRose, Jonathan C. Badger, Roshan M. D'Souza, Zeyun Yu and Peggy L. Peissig. "Object Detection to Assist Visually Impaired People: A Deep Neural Network Adventure." ISVC (2018).
- [8] L. Chen, J. Su, M. Chen, W. Chang, C. Yang and C. Sie, "An Implementation of an Intelligent Assistance System for Visually Impaired/Blind People," 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2019, pp. 1-2, doi: 10.1109/ICCE.2019.8661943.
- [9] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.
- [10] "Visual impairment and blindness Research", Oct. 2017, [online] Available: <http://www.who.int/mediacentre/factsheets/fs282/en/>.
- [11] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: current status challenges and future directions", Sensors, vol. 17, 2017.
- [12] J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu, "Virtual-blind-road following-based wearable navigation device for blind people", IEEE Trans. on Consumer Electronics, vol. 64, no. 1, pp. 136-143, 2018.
- [13] C.-W. Lee, P. Chondro, S.-J. Ruan, O. Christen and E. Naroska, "Improving mobility for the visually impaired: a wearable indoor positioning system based on visual marker", IEEE Consumer Electronics Magazine, vol. 7, no. 3, pp. 12-20, 2018.
- [14] Chaitanya Tejaswi, Bhargav Goradiya, Ripal Patel. A Novel Approach of Tesseract-OCR Usage for Newspaper Article Images. Journal of Computer Technology & Applications. 2018; 9(3): 24-29p