Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 5, 4167-4173

Deployment And Layout Of Deep Learning-Based Smart Eyewear Applications Platform For Vision Disabled Individuals

Himanshu Kaushik[1] , Manpreet Bajwa[2] , Anupam Kumar Sharma[3] , Prashant Vats[4] , Shipra Varshney[5] , Sandeep Singh[6] , Siddhartha Sankar Biswas[7]

[1]Department of Computer Science & Engineering, Faculty of Engineering & Technology, SGT University, Gurugram, Haryana, India. khimanshuofficial@gmail.com
[2]Department of Computer Science & Engineering, Faculty of Engineering & Technology, SGT University, Gurugram, Haryana, India. manpreet_feat@sgtuniversity.org
[3]Department of Computer Science & Engineering, Dr. Akhilesh Das Gupta Institute of Technology & Management, New Delhi, India. anupam.sharma@adgitmdelhi.ac.in
[4]Department of Computer Science & Engineering, Faculty of Engineering & Technology, SGT University, Gurugram, Haryana, India. , prashantvats12345@gmail.com,
[5]Department of Computer Science & Engineering, Dr. Akhilesh Das Gupta Institute of Technology & Management, New Delhi, India. shipra_vin@yahoo.com
[6] Department of Computer Science & Engineering, Faculty of Engineering & Technology, SGT University, Gurugram, Haryana, India. sandeep_feat@sgtuniversity.org
[7]Department of Computer Science & Engineering, School of Engineering Sciences & Technology, Jamia Hamdard University, New Delhi, ssbiswas@jamiahamdard.ac.in

ABSTRACT:

This study suggests a mechanism to help people with visual impairments navigate and move about. Nowadays, multiple services are possible at the same moment. Some of these will be discussed later in this essay. Nevertheless, no dependable and cost-effective alternative has yet been presented to replace the existing technology used by the visual impairments in their everyday tasks. This report begins with examining the problem, followed by an examination of the original problem resolution. Following that, it looked at some of the newest advancements and research in the field of helpful technologies. Finally, it suggests developing and implementing a system to assist the physically handicapped. A Raspberry microprocessor, cameras, batteries, eyeglasses, headphone, backup charger, and connections are all included in the suggested gadget. Use the camera to capture subjects. The phone's RCNN technique and deep learning modules will be used to finish computer vision and analysis. The ultimate output, on the other hand, is sent from the headgear to the vision impairer's hearing. The study offers methods as well as answers to the aforementioned issues. The research work may be applied to real-world scenarios and is suitable for people with visual impairments. For picture data processing, the gadget presented in this project employs an area accumulating neural net and a Microcontrollers. The process utilizes the Tesseract module from the Python language to do OCR (optical character recognition) and offers an output to the user. Later in this section, the prescribed tactics and effects are outlined.

KEYWORDS: AI, CNN, Smart Eyewear, Optical character recognition, Saliency Extraction.

I. INTRODUCTION

Vision is the most basic requirement of each and every person being, beginning with one of the most basic aspects of human psyche. According to the Health Organization (WHO), about 1.3 billion of people suffering from eyesight issues globally in 2018. A total of 39 million individuals are blind, with 246 billion having normal vision [7]. People with poor eyesight rely on the hands of others to satisfy their everyday requirements, and they are frequently affected by or compromised by this condition. We are attempting to offer solutions that can assist the visually handicapped in solving their daily challenges in the face of modernity and the technology revolution. This article will continue with a thorough explanation of the gadget. A blind person can't even walk without assistance. His safety is constantly in the hands of your guardians, but they may face difficulties. The increasing number of individuals with impairments throughout the world has piqued academics' interest in developing various technologies that will allow people with a disability to utilize this technology in the same way that regular folk do to fulfill their everyday chores. This generation of digital eyeglasses may capture the stereoscopic phone camera input and give comments to users via earbuds, allowing them to travel alone in a new area. Blind persons can be taught to view objects using programmed alternative sensory devices in this way. Wearable tech is computing gadgets that are placed over the glasses. Their screen travels with the user's phone, allowing customers to see the screen no matter where they are or how they are positioned. As a result, smart glasses and lenses may alter or improve concentration regardless of where the person is or what they are gazing at. There are 3 separate approaches for altering how the viewer perceives visual data. This paper takes advantage of contemporary technology's capabilities to investigate how it may replace the physically impairer's present traditional mobile phones and, as a consequence, improve the quality of life of these people all around the world. Intelligence Eyeglasses with Supported Sight were created to assist patients with severe visual impairments in navigating new environments, overcoming obstacles, and gaining more freedom [15]. The fact that most blind persons would have had at least partly impaired vision is the basis for eyeglasses.

2. RELATED WORK ON CNN AND DEEP LEARNING APPLICATION FOR FEATURE EXTRACTION

Object tracking technologies are now trying to make the world and more rational place. We are developing wearable technology for blind individuals as a phase of this effort, which will assist individuals in navigating their everyday lives as well as identifying things. This paper discusses how to identify things indicated on the highway or in indoor areas using a plastic foundation [1] approach. We propose a method for classifying things found on the ground. Technical features necessitate a lot of energy and manpower. Conversely, supervised learning may be viewed as a type of presenting wherein we deliver an unfiltered input signal in terms of text, image, or video. The vast processing capacity of processing units as GPU is used in deep learning algorithms (GPUs).

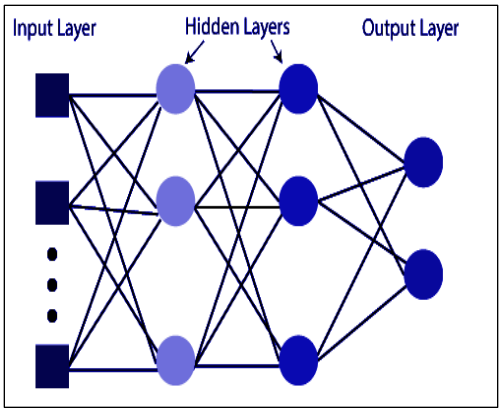

Figure 1: Simple Multilayer Perceptron

Multilayer Perceptron (MLP), as seen in the illustration, examines all values of the same relevance and ignores the spatial position of the input function. In a picture, neighboring pixels have comparable interpretations and contribute substantially to the picture's inherent meaning. To categorize the picture, we may assist the neural network in some manner [16] in obtaining the spatial configuration, which could also aid in improving the accuracy of classification. Sound neural networks are seen in this image. In the

picture above, the convolutional algorithm is connected to the input image. Two-dimensional convolutional filtering is used.

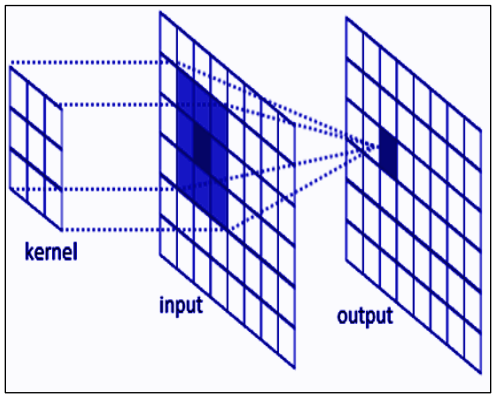

Figure 2: Convolution Operation

In the picture above, the convolutional algorithm is connected to the input image. Two-dimensional convolutional filtering is used. It's vital to keep in view that we'll require a kernel for a picture with exceedingly large inputs. The described primarily comes into the picture at this point. In actuality, a multilayer perceptron neural network typically contains a greater number of ordered variables than a congregation of similar neural circuits or better. The filtering had already collaborated with the picture to detect the boundaries and curvature in the device's perception [20]. The spatial picture structure is exploited by abstract neuronal networks, and they may be utilized to handle a variety of sorting and extrapolation problems [8]. The competitional neural networks were designed for the very first time by Lecun et al., however, the processing capability of the machines at the moment was sufficient to include the successful algorithms. CNN is made up of modern layering and groups, as well as quasi activation algorithms and multilayered percepts. The contouritic function, as previously said, strives to discover a certain shape or functional boundary, the clustering of the 3 levels picture for more high-level features with the precise size of something like the filter, and the quasi-input layer produces the network. Sophisticated algorithms and structures that are entirely interconnected strive to categorize the image. For better precision, a depth neural network [14] must be educated with a large amount of data. However, a competing neural network performs better when there is a lot of information. Moreover, if the data gathering is insufficiently vast, the system may result in a minority. It's considered an overstatement when the network learns a specific border, form, or connection since it only works on the trained pictures and doesn't transfer effectively. When identifying a roadway, a creature, or a human, the information must be trained and tested number of photographs from each category. Designers wanted images from the "actual" environment acquired by sensor equipment in their gadgets. As a result, the information had to be gathered individually. This takes a deal of time and resources, and anyway, it can't get the photos inside the order that the difficulties of picture categorization work (millions of images). With the first layer, we utilize a standard convolution layer with the convolution level and highest speed layered. Furthermore, following a flattening layer, which converts any n-dimensional matrices to a 1D matrices, several surface multilayered sensory levels, and lastly, an output unit with a sequence of values equal to the number of classes. In actuality, multilayer neural networks typically contain a greater number of structured variables than a congregation of similar neural circuits or better. The filtering had already collaborated with the picture to detect the boundaries and curvature in the device's perception [20]. The spatial picture structure is exploited by abstract neural network models, and they may be utilized to handle a wide variety of sorting and extrapolation problems [8].

3. PROPOSED APPROACH FOR DEEP LEARNING-BASED SMART EYEWEAR APPLICATIONS PLATFORM

The project's machine intelligence eyeglasses are supplementary electronic spectacles that collect and analyze pictures using artificial intelligence techniques. It will contain the following elements in this context. The Raspberry Pi foundation is a PI family microprocessor card [14]. It may be thought of as a single-board computer that runs on the Linux operating system. For Cordless connection One may establish a WiFi access node for Web access with the Pic Microcontroller 3's wireless System and Headset configuration. The Rpi featured a specific connector for connecting

the LCD touchscreen interface, which is completely absent from the display. The Raspberry Pi 3 also features a professional camera doorway that allows users to attach the cameras to the PI cards without even any issues. Consider, for example, the 6 Gb of ram MicroSD Card slot devoted towards the Pi System. The table includes a power cable and a resume mechanism. The gadget is extremely dangerous due to its 5V voltage supply. One can use a USB cable that connects to a laptop or may use a battery or an AC/DC converter to charge it. For the Camera module for Raspberry Pi, a Source population is recorded via the webcam. The webcam is in charge of capturing images. Furthermore, this required information will be sent to the Raspberry Pi CPU for processing. The cameras will be mounted on top or side of the spectacles. For the Source of power, It can be used to provide Pic Microcontroller mobile power bank with uninterrupted power supply power. The computer's energy requirements have been met by using DC power from the power source. For the voice-recognition component, the user will be informed when earphones are being utilized for output by this module. Using earphones, the Raspberry Pi's input will be transmitted to an individual's ear. Based on the feedback of the speech of objects surrounding them, Researchers can identify photos, collect features from data, and then translate this text into an element [19] to supply the speakers with profound learning architectures like CNN, Reinforcement Learning, and Generative adversarial. To identify and recognize differences, we developed a Cnn architecture to categorize pictures and an OCR Triangular prism package to retrieve picture text, as well as a face recognition package and a matching algorithm.

The method we used may be broken down into four basic steps:

3.1 Information gathering and algorithmic coaching: To build a CNN model with good accuracy, we need to gather as many pictures as possible in the period allotted. We gathered photographs from the sessions we required and Then created a Cnn architecture from them. Keras, a Language package that uses Tensorflow upon that backend, was used for the modeling. After gathering data, we partition the batch data for training and use the steepest descent approach for improving the objective functions.

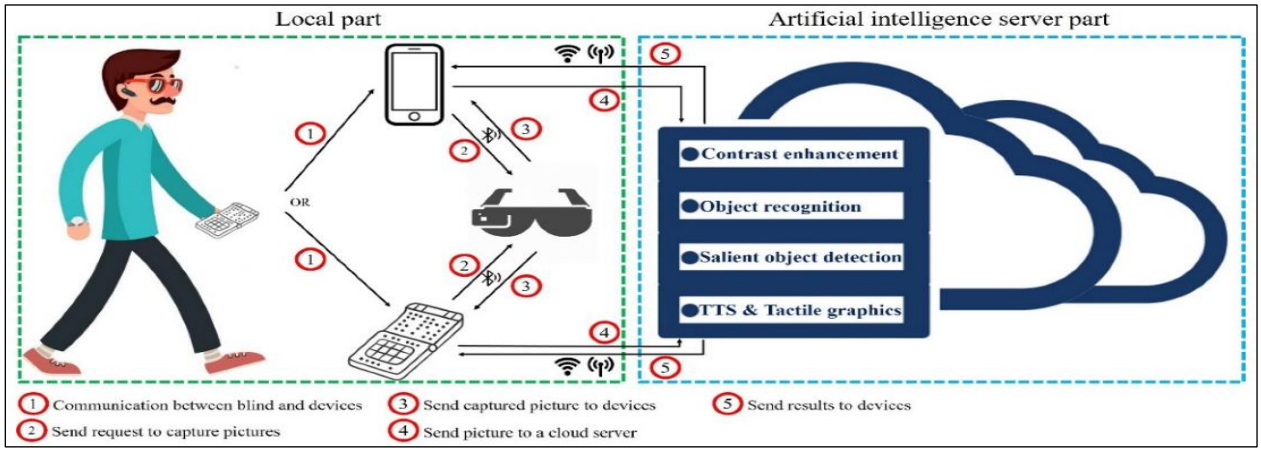

Fig. 3 To show the proposed designated wearable sensor technology using AI and CNN

3.2 Modeling implied argument: The information will be processed and stored to storage after retraining. To speed up the going investigation, we construct the network and keep it in storage throughout the test stage (intimation). The viewfinder photos may be directly determined in the Python OpenCV package and then sent to the charging mechanism. The framework enables us to have individual opportunities in each class. As a result, the picture fits under any of the preceding categories.

3.3 Tesseract and voice text: To transform photos into text, designers utilize Google's tesseract package [11]. By using the photos acquired from the webcam straight into the tesseract model, as

in the second step. To allow faster inference, we retain this information loaded in storage. Following receipt of the information, users can transform the written data into pronunciation.

3.4 Identification of the identity of the human being in whom the model is formed: Researchers utilized the OpenCV Face Detection and recognition Library to identify the identity of the human individual wherein the modeling is produced [17]. We may start by educating employees and then enhance as time goes on. A superior facial recognition method is used to estimate the position of the head. The facial arteries (presents) then are retrieved to use the face recognition system. Then we contrasted the companies operating of the expression we generated to use the lighted proximity to the consolidations of the facial we created using these face legal entities. Whichever class the image belongs to is determined by the smallest distance metric between both the database and the broad range. When a caution event occurs (such as a car arriving or a road approach, a human approaching, a creature approaching, or stairs approaching), the eyeglasses activate a topic Lookup and notify users with a message for the term "at the front." Anything other than that, it provokes and gives the customer the outcome (that also is in front of the consumer, where to get off using intrinsic GPS, and what were the patient thinks, the synopsis of the picture (sophisticated learning), and much more demanding to integrate into such an Iot platform) whenever the user is urged (clicking a toggle or having to talk of activity).

4. RESULTS AND DISCUSSIONS

The sensor collects the contents of the item, as well as the captured data, which is communicated to the computer program, which then displays the outcome using an intelligent machine. The ensuing product is typically precise, as is the transmission from the audio modules to the person's ear. To enjoy the benefits of this aesthetic value, headsets with aided eyesight were created. An OLED screen, an accelerometer, Gps receiver, microphone, compasses, and miniature camera were all included in the spectacles. Analyze the data that comes in and do it in a variety of ways, such as using luminance to signify depth. Although most visual impairments persons can tell the difference between dark and bright light, eyeglasses may be used to highlight any item near the operator and help them discern between barriers and individuals. So, because a person travels, the GPS may provide directions, and the accelerometer enables the eyeglasses to compute the change in the viewpoint of vision. The camera can also be used in conjunction with the computational module to help with the understanding of the text. Eyewear AI combines several features, such as machine learning, acoustic techniques, and mathematical modeling, to produce a valuable resource for those with visual impairments [3].

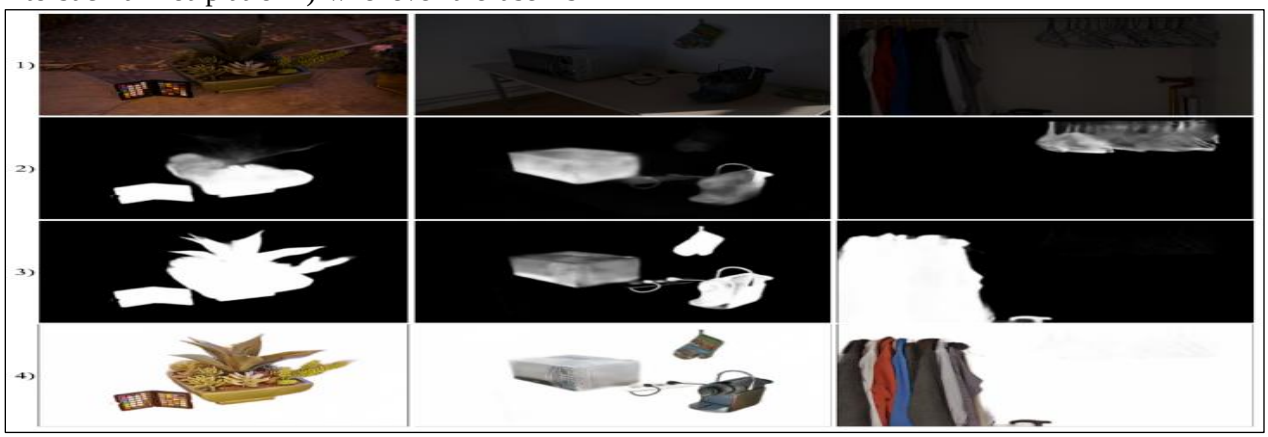

Fig.4 On the tough LOL samples, the outcomes of the saliency extract algorithm. (1) photos with limited lighting; (2) saliency identification after image improvement; (3) saliency identification after contrast enhancement; (4) prominent items in full - colour spectrum

The saliency identification model was visually evaluated with a reduced picture initially and then after using the low-light addition to address. Figure 4 displays the darkish pictures that were examined, such as a flower vase, garments, and a microwave, in the first row. The prominent object extraction is shown in the middle row well before the reduced light-enhancing procedure. The outcomes of visual feature identification following the picture improvement approach and prominent entities in full-color capacity by increasing the ternary masked approach are shown in the 3rd and 4th rows, correspondingly. Because the backgrounds and forefront are comparable, the saliency separation findings from darkish photos demonstrate inaccurate extraction, as illustrated in the middle row of Figure 4. The suggested saliency identification approach, on the other hand, can mitigate these disadvantages. The suggested saliency selection approach, on the other hand, can mitigate these disadvantages. We enhanced the contrast between both the backdrop and the item using the reduced light picture progress in implementing and therefore easily retrieved several elements. Furthermore, improving low-light picture brightness improves the accuracy of applying the edge detection method to identify the interior borders of prominent items.

5. CONCLUSIONS

The theoretical version demonstrates of conclusions can be employed in production environments. The gadgets we recommend can identify objects and recognize faces, but they have limits owing to the lowest CPU and need to be improved. The application of domain adaptation can provide an accurate note, but now with poor performance, and conversely, the customer conventions require a quick reaction. To solve this problem, we needed a fast-computing gadget that's also light on electricity. This study offers a key technology device for BVI persons that includes computer intelligence and learning techniques to identify objects, extract prominent objects, and recognize the language. The suggested solution is automated and works on a web server artificial intelligence and machine learning. To aid BVI in a night-time setting, it identifies and distinguishes items from reduced and shadowy photos. Deep neural networks were used to enhance the classic smart glass technology, which now includes saliency separation for interactive visuals and language processing for text-to-speech. Since smart building systems rely on deep learning methods, they need more resources and storage in integrated devices. As a result, we included an artificially intelligent system to assure performance advantage and address energy issues.

References:

1. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks." In: Advances in neural information processing systems. 2012, pp. 1097–1105

2. Frieden, B. Roy. "Image enhancement and restoration." Picture processing and digital filtering. Springer, Berlin, Heidelberg, 1975. 177-248.

3. Ilag, Balu N., and Yogesh Athave. "A-Design review of Smart Stick for the Blind Equipped with Obstacle Detection and Identification using Artificial Intelligence." International Journal of Computer Applications 975: 8887.

4. J. Donahue, R. Girshick, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," IEEE Int. Conf. on Computer Vision and Pattern Recognition, Columbus, 2014.

5. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016, pp. 770–778

6. Karen Simonyan and Andrew Zisserman. "Profound convolutional networks for large-scale image recognition." In: arXiv preprint arXiv:1409.1556 (2014).

7. Li, Xiang, et al. "Cross-Safe: A computer vision-based approach to make all intersection-related pedestrian signals accessible for the visually impaired." Science and Information Conference. Springer, Cham, 2019.

8. Liu, Fayao, Chunhua Shen, and Guosheng Lin. "Deep convolutional neural fields for depth estimation from a single image." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

9. Matthew D Zeiler and Rob Fergus. "Visualizing and understanding convolutional networks." In: European Conference on Computer Vision. Springer. 2014, pp. 818–833

10. Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. "Learning and transferring mid-level image representations using convolutional neural networks." In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014, pp. 1717–1724.

11. Mithe, Ravina, Supriya Indalkar, and Nilam Divekar. "Optical character recognition." International journal of recent technology and engineering (IJRTE) 2.1 (2013): 72-75.

12. R. Girshick, "Fast R-CNN," IEEE International Conference on Computer Vision, Santiago, 2015

13. Sakhardande, Jayant, Pratik Pattanayak, Mita Bhowmick. "Smart Cane Assisted Mobility for the Visually Impaired." International Journal of Electrical and Computer Engineering, vol. 6, no. 10, 2012, pp. 9-20

14. Sandnes, Frode Eika. "What do low-vision users want from smart glasses? Faces, text, and perhaps no glasses at all." International Conference on Computers Helping People with Special Needs. Springer, Cham, 2016.

15. Takefuji, Yoshiyasu. Neural network parallel computing. Vol. 164. Springer Science & Business Media, 2012.

16. Viraktamath, S. V., et al. "Face detection and tracking using OpenCV." The SIJ Transactions on Computer Networks & Communication Engineering (CNCE) 1.3 (2013): 45-50.