International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-2, July 2020

AI Powered Glasses for Visually Impaired Person

Nirav Satani, Smit Patel, Sandip Patel

Abstract*: This dissertation presents a system that can assist a person with a visual impairment in both navigation and movability. Meanwhile, number of solutions are available in current time. We described some of them in the later part of the paper. But to date, a reliable and cost-effective solution has not been put forward to replace the legacy devices currently used in mobilizing on a daily basis for people with a visual impairment. This report first examines the problem at hand and the motivation behind addressing it. Later, it explores relative current technologies and research in the assistive technologies industry. Finally, it proposes a system design and implementation for the assistance of visually impaired people. The proposed device is equipped with hardware like raspberry pi processor, camera, battery, goggles, earphone, power bank and connector. Objects will be captured with the help of camera. Image processing and detecting would be done with the help of deep learning, R-CNN like modules on the device itself. However, final output would be delivered by the earphone into the visually impaired person's ear. The research work contains the methodology and the solutions of above mention problem. The research works can be used in practical use cases, for visually impaired person. The system proposed in this project includes the use of a region based convolutional neural network as well as the use of a raspberry pi for processing the image data. System includes tesseract library of programming language python for OCR and give output to the user. The detailed methodology and result are elaborated later in this paper.

Keywords: OCR, R-CNN, Transfer learning.

I. INTRODUCTION

Starting with most crucial part of human psychology, Vision is an essential need of any person. Considering the fact given by the World Health Organization (WHO), in 2018 nearby 1.3 billion people in the world are suffered by visual problems. Among them, about 39 million people are blind, and roughly 246 million people have light vision [7]. Visual impaired person usually depends on others hand to fulfill their daily needs, and many times suffer or compromise a lot due to this illness. Considering modern era, and revolution in technology we try to introduce an equipment which can help to face daily life problems of visually impaired person. Elaboration of this equipment is mention further in this paper.

Could you imagine how the lifetime of an individual 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 inworld attracts the concern of researchers to invent various technologies, aiming that these technologies can useful the disabled people to perform their daily 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. This smart glass can assist them while walking alone in new environments by taking inputs through a stereo camera and providing feedback to the person through headphones. so, people blind can be trained to visualize objects using sensory substitution devices programmed. Smart glasses are computing devices worn ahead of the eyes. Evidently their displays move with the user's head, which results in the users seeing the display independently of his or her position and orientation. Therefore, the technology like smart glasses or lenses can be use which can alter or enhance the wearer's vision no matter where he/she is physically located and where he/she looks. There are three different paradigms of the way to alter the visual information a wearer perceives. Leveraging the potential of modern technologies, this report explores how much they can be used to replace the current legacy devices used by visually impaired people for mobility and ultimately improve the quality of life for these people worldwide.

II. A REVIEW OF RELATED WORKS

Till now so many works had done for the blind and visually impaired person by making developing many different kinds of assistive technologies to aid them in navigation. Some of the works are described below.

1.1 THE ASSISTED VISION SMART GLASSES

The assisted vision smart glasses were designed to aid individuals with very low vision to navigate unfamiliar settings, recognize obstacles, and achieve a higher degree of independence [15]. The glasses are based on the fact that most blind individuals have at least some dim vision. The assisted vision smart glasses are intended to capitalize on this sight.

Manuscript received on May 25, 2020.
Revised Manuscript received on June 29, 2020.
Manuscript published on July 30, 2020.
***** Correspondence Author
Nirav Satani*, Information Technology, Parul University, Vadodara, India. Email: niravsatani24@gmail.com
Smit Patel, Information Technology, Parul University, Vadodara, India. Email: niravsatani24@gmail.com
Sandip Patel, Information Technology, Parul University, Vadodara, India. Email: niravsatani24@gmail.com

The glasses are made of OLED displays, with a gyroscope, a GPS module, an earpiece, a compass, and two small cameras. Incoming data is processed, and then used in a myriad of ways, for instance, brilliance can be used to show depth.

Since most visually impaired persons can distinguish brilliance from dullness, the glasses can brighten anything close to the wearer so that they can discern obstacles and persons. The GPS module can provide directions, and the gyroscope assists the glasses in calculating perspective changes as the wearer moves. The camera can also work with the computing module to help read markers along the way (cit).

1.2 THE AI GLASSES

AI glasses integrate a host of features, including artificial intelligence, ultrasound techniques, and computational geometry to create an essential aid for visually impaired persons [3]. By linking glasses with GPS technology of a tablet, along with stereo sound sensors, the prototype can issue spoken commands, recognize currency denominations, read signs, etc. The estimated cost of AI glasses is between $1000 and $1500.

1.3 ENVISION GLASSES

Recently Google and envision company introduced AI glasses which helpful for the visually impaired person. The cost of this equipment is around 1,50,000₹. This glass is also known as Google smart glass [4].

III. AI GLASSED FOR VISUALLY IMPAIRED PERSON

The AI glasses for this project is an electric supplementary eye goggles, equipped with image capturing and recognition using Artificial Intelligence technologies. In this regard, it will have the following components:

1.4 A RASPBERRY PI

The Raspberry PI 3 is a microcontroller board n PI series [14]. It can be considered as a single-board computer that works on a LINUX operating system. According to a 2018 article, the board not only has tons of features it also has terrific processing speed making it suitable for advanced applications. here wireless connectivity is needed. Raspberry PI 3 has wireless LAN and Bluetooth facility by which you can setup WIFI HOTSPOT for internet connectivity. Raspberry PI 3 had dedicated port for connecting touch LCD display which is a feature that completely omits the need to monitor. Raspberry PI 3 also has dedicated camera port so one can connect camera without any hassle to the PI board. Take the 16GB micro SD card and dedicate it specifically for PI OS. Board has a reset button and a power jack. With an operating voltage of 5V, the tool is highly user-friendly. It can be connected to a computer directly through a USB cable or powered up with a battery, or an AC-DC adapter.

1.5 RASPBERRY PI CAMERA MODULE

The camera is used to take the input from outside. The camera is responsible for taking the images of the objects. Further, this input data transfer to the raspberry pi processor for computation. Position of the camera would be at the top or side of the glasses.

1.6 POWER SUPPLY

To provide the constant power to the raspberry pi portable power bank would be used. DC supply through the power bank fulfils the electricity requirements of the processor.

1.7 A VOICE PLAYBACK MODULE

This module will inform the user of the Earphone is used for output purposes. Output generated by the raspberry pi will be delivered to the person's ear with the help of earphone.

BACKGROUND ON CNN AND MACHINE LEARNING

Today, computer vision-based applications are making the planet a far better and convenient place. during this project, we are making smart glasses for visually impaired, which may help them navigate also as identifying objects in day to day life. this text explains a deep learning-based [1] method for classification of objects given during a road or indoor setting. We propose one method to classify the objects on the road. during this method, we train a convolutional neural network.

In practice, machine learning requires features and to extract features, one must do feature engineering. Feature engineering takes tons of diligence and manpower. Instead, deep learning may be a sort of representation learning, where we feed the input signals in its raw form, be it speech, image, video. In our application, we'd like to classify objects in any given image. Deep learning methods exploit the tremendous computational power of Graphics Processing Units (GPUs). At each level of the deep learning architectures, we introduce a non-linearity which helps in solving very complex problems.

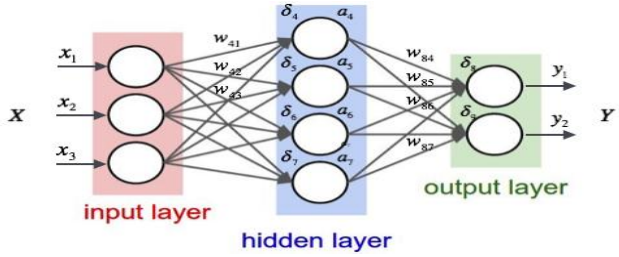


Figure 1: Simple Multilayer Perceptron

As shown in the image, multi-layer perceptrons consider all the features of equal importance and do not consider any spatial position of the input features. While in the image the neighboring pixels have similar values and they contribute significantly in the semantic information of the image. In order to classify the image, if somehow, we can help the neural network [16] get the spatial structure, then it can help in achieving better classification accuracy. This is where convolutional neural networks come into the picture.

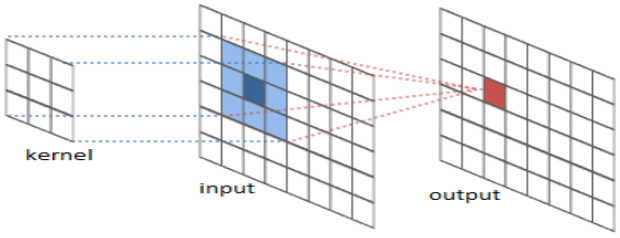


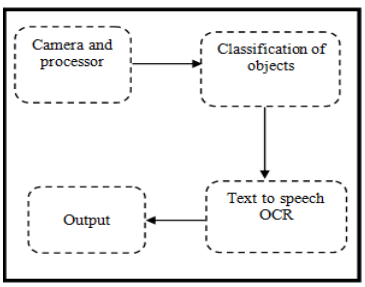
Figure 2: Convolution Operation

In the image above, a convolution function [2] is applied to an input image. We use 2-dimensional convolution filters. Let's understand the convolution function a bit. Here, the kernel is basically a convolution filter. Suppose, if we want to find a diagonal in the input image.

So, the right type of filter would be all ones on the diagonal and negatives on other places. Now, we can find diagonals in the image, by convolving it through the image. Here, one important thing to keep in mind is that, we need only one kernel for an arbitrarily large input image. Here, parameter sharing comes into the picture. In practice, multi-layered neural networks generally have orders of times a greater number of parameters than an equivalent or better performing convolutional neural network. In earlier computer vision eras, filters were convolved with the image to identify edges and curves [20].

Convolutional neural networks exploit the spatial structure of an image and we can use it to solve many classification and regression problems [8]. Convolutional Neural Networks was first proposed by Lecun et al. But, at that time, the computational capability of the computers held back the algorithm from thriving. CNNs became popular in 2012 when AlexNet proposed by Krizhevsky et al. won the ILSVRC 2012 challenge in classifying ImageNet images. Here, 2 GPUs were used in parallel by them. In general, CNN consists of convolutional and pooling layers, non-linear activation functions, and multi-layer perceptrons. As explained above, convolutional functions try to find a specific shape or edge of the function, pooling layers down sample the image to get more semantic details with the same filter size, non-linear activation function makes the network learn complex functions and fully connected layers try to put up altogether to classify the image. Deep learning neural networks [14] need to have a lot of data to train themselves for better accuracy. But, if the data is in abundance, in practice, a convolutional neural network works better. And if the dataset is not sufficiently large then the network may result in overfitting. Whenever the network learns a very specific type of edge or shape or connection, where it works on training images, but cannot generalize well then it is called overfitting. In the classification of road, animal or a person the data needs to be trained on many images of such classes. In our project, we needed 'real' world images, i.e. images captured from a camera mounted on the head. So, the data needed to be collected manually. This requires a lot of time and effort and still, it cannot come to the order of images in which image classification challenges work (millions of images). We use a typical convolutional layer with the initial layer being stacked convolutional and max-pooling layer. And after that a flatten layer, which unrolls any N-Dimensional matrix to a 1-D array, followed by some simple multi-layer perceptron layers and finally the output layer with a number of units equal to the number of classes.

IV. SYSSTEM DESIGN



Work flow diagram

Here are the steps involved in image detection and identification. The Raspberry PI board is the main part of the system. It controls the other system components. When the object is captured by the camera, it will send the information to the microcontroller. The microcontroller, in turn, will detect the object using artificial intelligence. Then, the result is transmitted to the earphone, which is contacted with a blind person.

APPROACH

These are smart glasses which give voice feedback of the things around you when prompted or when in caution. With the development of advanced deep learning architectures such as CNNs, RNNs, and GANs, we can classify images, extract information from the images and then convert this text to speech [19] in order to feed into the speaker. We train a CNN model to classify images, Tesseract OCR library to get a text from images and face recognition library and a matching algorithm to detect and recognize faces.

The approach taken by us can be divided into four major steps:

1.8 DATA COLLECTION AND TRAINING THE ALGORITHM:

To train a CNN model to get some respectable accuracy, we need to collect as many images as possible in the given time frame. We collected images of the classes we needed and then trained a CNN model on it. The programming was done using Keras which is a Python library using Tensorflow in the backend. After data collection, we divide the training data into batches and then optimize the loss function using gradient descent algorithm.

1.9 INFERENCE OF THE MODEL:

After training the model is then serialized and saved to the disk. At the testing (inferencing) stage we load the model and keep it in the memory in order to make the inference process faster. Images captured by the camera can be read directly into the python library openCv and then fed to the model we loaded.

The model gives us the individual probabilities of it being in each class. Then the image can be classified into any of the classes above.

1.10 TESSERACT AND TEXT TO SPEECH:

We use Google's tesseract library as an Optical Character Recognition tool to convert image to text [11]. As same as in the second section, we use the images captured from the camera directly into the tesseract model. We also keep this model loaded in the memory to make the inference faster. It gets the data and the text data then can be converted to text to speech.

1.11 FACE RECOGNITION:

We also used openCV's face recognition library to recognize the name of the human on which the model is trained [17]. We can train some users on the initial base and can improve along the way. The location of the face is determined using a pre-trained face detection model.

Then the face features (face embeddings) are extracted using the face recognition model. Then we use these face embeddings and compare these to the face embeddings we have trained using Euclidean distance. The lowest Euclidean distance with the highest frequency among the dataset determines in which class the image belongs.

Glasses usually sit idle and observe the surroundings, when any cautionary event occurs (such as a car is coming or road ahead, a person ahead, animal ahead, stairs) it triggers internal thread and alarm the user with a text to speech output saying "thing ahead". Otherwise, when prompted by the user (by pressing a button or saying an action word) it triggers and gives user the output (who is in front of the user, where to go for a place using inbuilt GPS, what are the things around the user, image description (advanced deep learning) and harder to incorporate in an IoT device.). I will use Intel Movidius NCS for the computation and a Raspberry Pi.

V. RESULT AND DISCUSSION

The contents Images of the objects are captured by the camera and captured information sent to backend application and later backend application gives the result with the help of Artificial based system. This generated result is tended to accurate and its transfers to the person's ear with the help of the sound module. Some related clicks of the working procedure, as a result, are mention below:

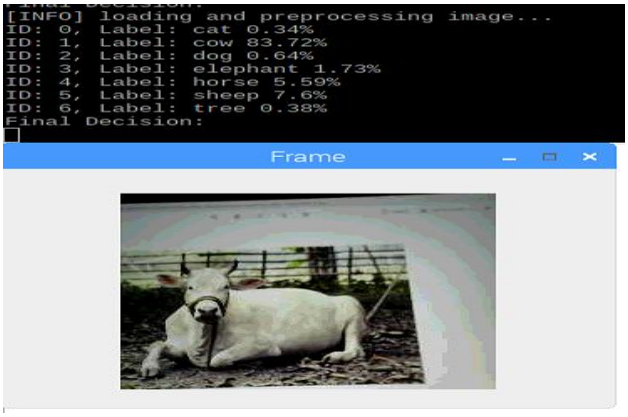


Figure 3: Cow Detection

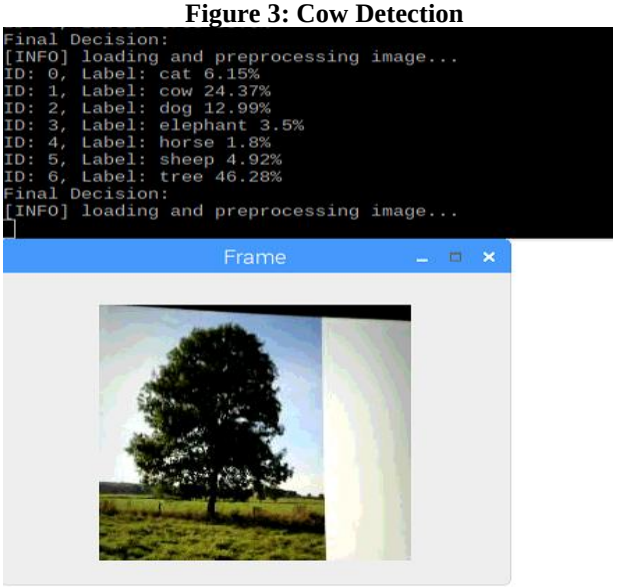


Figure 4: Tree Detection

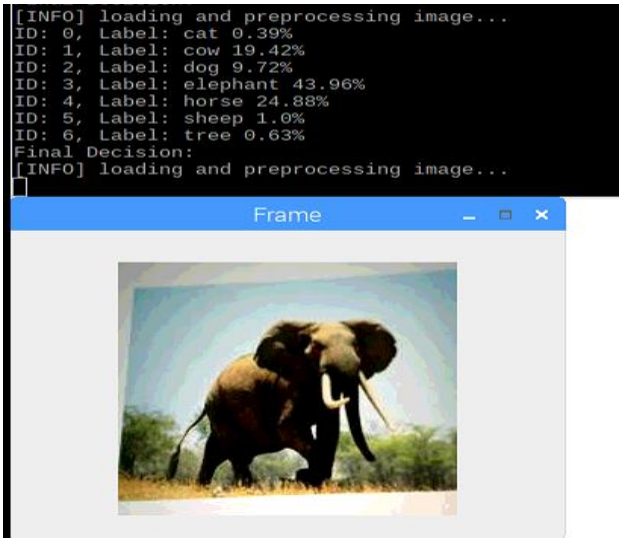


Figure 5: Elephant Detection

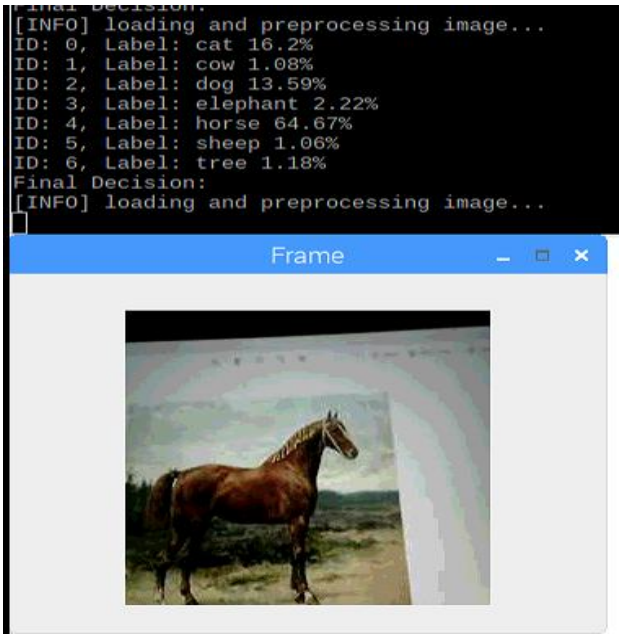


Figure 6: Horse Detection

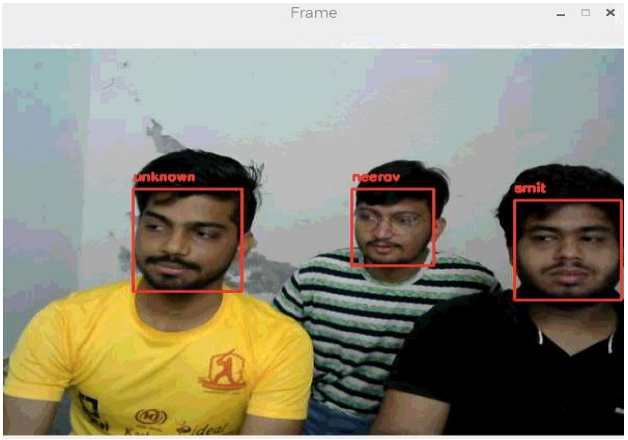


Figure 7: Face Recognition

TABLE 1: RESULT

Object	Accuracy	Result
Cow	83.72%	Cow detected
Tree	46.28%	Tree detected
Elephant	43.96%	Elephant detected
Horse	64.67%	Horse detected
Faces	100%	Face recognized

TABLE 2: SYSTEM DEFICIENCIES AND REMEDIATION

System deficiencies	Remediation
One drawback of this system is it is comparatively slow.	By using other faster algorithm and highly efficient processor, speed can be improved.
Limited objects detection possible due to raspberry pi limited efficiency.	The highly efficient processor can solve this problem.
This system may not be accessible for use in the underdeveloped nation because of cost and unfamiliar with technology.	By organizing events to give a demonstration of the new technology and financial aid can cure this problem.

VI. CONCLUSION

A conclusion the proof of concept version can be used in a production environment. Our proposed device is able to detect the objects, and recognize the faces but it has limitation because of lower processor and therefore, it still needs some improvements. The use of transfer learning can make the classification highly accurate but slow the performance down, conversely, quick response is essential for the convention of users. To address this issue, one needs a faster computing device which ideally should be low-powered. Apart from this, Other better option is to use cloud computing.

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. “Very deep 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. Raspberry Pi information, 26 April 2018, https://components101.com/microcontrollers/raspberrypi-3-pinout-features-datasheet
13. R. Girshick, "Fast R-CNN," IEEE International Conference on Computer Vision, Santiago, 2015
14. 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
15. Sandnes, Frode Eika. "What do low-vision users really want from smart glasses? Faces, text and perhaps no glasses at all." International Conference on Computers Helping People with Special Needs. Springer, Cham, 2016.
16. Takefuji, Yoshiyasu. Neural network parallel computing. Vol. 164. Springer Science & Business Media, 2012.
17. 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.
18. World Health Organization (WHO). “Blindness and Vision Impairment.” WHO, 11 October 2018, https://www.who.int/news-room/factsheets/detail/blindness-and-visual-impairment
19. Xing Luo, Oscar. "Deep Learning for Speech Enhancement: A Study on WaveNet, GANs and General CNN-RNN Architectures." (2019).
20. Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. “Gradient-based learning applied to document ´ recognition”. In: Proceedings of the IEEE 86.11 (1998), pp. 2278–2324

AUTHORS PROFILE

Nirav satani is currently pursuing a bachelor’s degree in the field of Information and Technology at the Department of Computer Science at Parul University, India. He is always keen on computer and technology. He did several use cases during his academics. For instance, he did smart signal-based project which was selected as best visionary-project of the year 2018 by his university. He did competitive coding too. He took a part in various online coding competition and he also won the coding challenge (in group) which held on university level. Deep learning, Machine learning and Data Science are the favorite domain of him. However, he also cleared competitive exam namely GATE (Graduate Aptitude Test in Engineering) in 2020. In future, he willing to works on data science domain. **Email**: niravsatani24@gmail.com

Smit Patel is currently pursuing a bachelor’s degree in the field of Information and Technology at the Department of Computer Science at Parul University, India. He has done a number of projects during his academics. He interested in machine learning, deep learning. However, he also like to build a website. He had done an internship as full stack developer. He willing to work as full stack developer and contributing in deep learning domain in future. Mathematics and visualization of the data are the interested leisure activity of him. He also did competitive programming on various online platform. He took a part in various online coding competition and he also won the coding challenge (in group) which held on university level. **Email:** patelsmit110799@gmail.com

Sandip Patel is currently pursuing a bachelor's degree in the field of Information and Technology at the Department of Computer Science at Parul University, India. He has been involved in a number of roles and projects during his academics. He did several projects by his self. Some of the projects he has worked on include Density based Management of traffic Management System, HR Management System, and the project related to the Job costing. His research interests are in the fields of Convolutional Neural Networks, Transfer learning and Deep Learning Algorithms. He has prior Experience of six Months in IT industry of India. His favorite academic activity is Competitive coding. He took a part in various online coding competition and he also won the coding challenge (in group) which held on university level. **Email:** sp910101@gmail.com