

An Assistive and Assessment Smart Eye for Visually Impaired People using CAFFE Model

Vaibhav Sontakke, Siva Shanmugam G



Abstract: *Those who are visionless or visually compromised and rely heavily on others, like their family or friends. Whereas in acquainted settings, they're going to face several obstacles whereas conducting their everyday activities. A bit like we tend to love democracy and knowledge it, so that they can expertise it too. They ought to not be set apart merely as a result of capable otherwise and particularly in today's world wherever we tend to are technologically thus advance. The planned work object discovery system identifies, calculates the gap to our camera from an illustrious entity in a picture and scans the important world objects from binary pictures or film, wherever the entity might belong to any category or cluster, like folks, cars, vehicles, etc. to finish this duty of detection an object in a picture or video, I used OpenCV packages, Caffe model, Python, and NumPy. This investigation work discovers however deep learning techniques are wont to notice a live object, find an item, reason an item, extract options and show information and lots of additional, in footage and videos use OpenCV and the way to use the Caffe model, and conjointly why select the restaurant prototype over alternative frames. To form our deep learning-based period factor detector with OpenCV, we did like to access webcams and apply factor discovery to every frame effectively.*

Keywords: *Object Detection, Deep Learning, Caffe Model, Thing Discovery, Distance from Camcorder to Thing, OpenCV, Raspberry Pi, Camcorder Module.*

I. INTRODUCTION

The objective of this work is to source a wise life for the blinded by employing an object discovery system to help them in their everyday lives. I used Deep learning factor discovery that features Mobile Nets, Single Shot Detectors, and OpenCV. Factor recognition was aimed toward unique items from the scene and predicting the several bounding boxes—the classifier developed by employing a standard.

Open supply deep learning platform, known as Caffe. I tend to use model Caffe to notice and classify things. Caffe model may be a pre-trained model capable of recognizing twenty-two distinct things like a vehicle, sofa, human, dog,

stairs, etc. The perceived pictures that are input for process withstand few phases like edge discovery, the discovery of greyscale, and color sense. I tend to might choose the output image in either greyscale or black and white or color RGB looking on our demand. The thing is to seek out the (xi, yj) coordinates of the items within the pictures. We'd like to match the detected factor with the prevailing factor in the model when obtaining the detected factor within the image model. The findings are going to be rumored with the preciseness of the factor detected vs. the present factor (for example, 98.72 percent of the individual detected).

II. RELATED WORK

In recent years, many related works [2]-[6] were studied and developed for caring, helping, and protecting the visually impaired people. Elmannai and Elleithy [2] reviewed the current development status of the sensor-based assistive devices/aids and gave some possible future directions. Khlaikhayai et al. [3] proposed an intelligent walking stick, which was applied for elderly and blind people to achieve safety and navigation. This work integrated wireless sensors with an ad hoc network. This intelligent walking stick can perform and realize the special events such as blind rally and disable people networks. Bai et al. [4] developed smart guiding glasses as an electronic travel aid (ETA) for providing indoor traveling for the visually impaired group. Compared to other existing works, this work implemented a multi-sensor fusion-based obstacle avoidance method, which was integrated a depth sensor with an ultrasonic sensor to overcome the problems of detecting small obstacles. Bai et al. [5] proposed a wearable smart glasses-based indoor navigation device, in which a dynamic sub-goal selecting scheme was adopted to guide the blind people to the destination and help them bypass obstacles simultaneously. Lee et al. [6] presented an indoor positioning system, in which image and ultrasound sensors were mounted on a pair of wearable glasses. This system can real-time recognize specific color-coded markers and observe obstacles for detecting distance of 15 meters. However, most previous works [4]-[6] for the visually impaired people were based on wearable glasses and applied to the indoor environment. It is difficult, nearly impossible, for blind people to walk without an aid. Even with an aid such as a white cane or a guide dog, it can sometimes be inconvenient, uncomfortable, and perhaps inaccurate in avoiding obstacles. This machine will establish such color-coded markers in period and observe obstacles to detection fifteen-meter distance. Even so, for the visually impaired folks earlier works [4]-[6] supported wearable glasses and applied them to the indoor climate.

Manuscript received on April 02, 2020.

Revised Manuscript received on April 20, 2020.

Manuscript published on May 30, 2020.

* Correspondence Author

Vaibhav Sontakke*, School of computer science and Engineering, Vellore Institute of Technology Vellore, India.
Vaibhav.sontakke2018@vitstudent.ac.in

Siva Shanmugam G, School of computer science and Engineering, Vellore institute of technology Vellore India.

Email: Sivashanmugam.g@vit.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

An Assistive and Assessment Smart Eye for Visually Impaired People using CAFFE Model

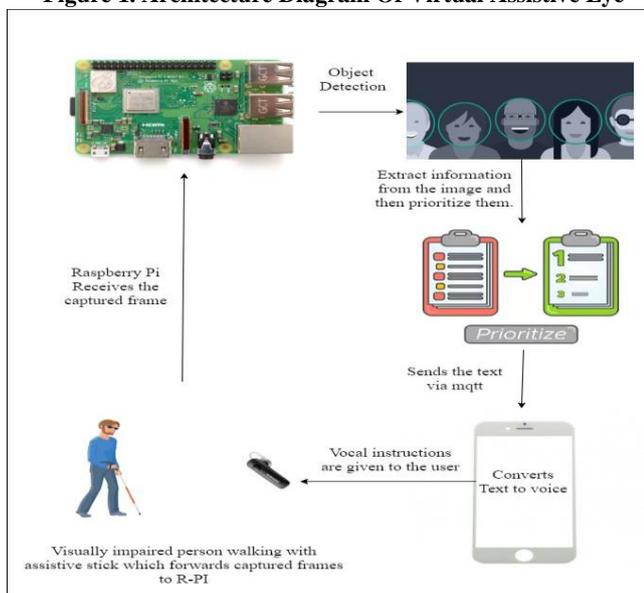
There's additional, and we tend to lack a full set of practical aids conjointly.

Hence, during this work, we tend to mix a good wearable glass with an intelligent-walking follow to produce an intelligent help system for visually impaired and blinded folks to supply walking safety functions like fall-down / collision warning and visually impaired/blinded folks identification of obstacles. We are walking while not aid is robust, virtually not possible for blinded folks. And with an aid like a white cane or a working dog, obstacles might usually be awkward, unpleasant, and even unreliable to avoid.

III. PROPOSED ARCHITECTURE

The camera placed on the stick is capable of taking appropriate wide-angle pictures. The camera continuously senses pictures, and those pictures work as an input to the image processing activity. Image processing activity then performs image recognition to identify the objects, and corresponding results are sent to the Raspberry Pi microprocessor installed into the stick. The results are calculated with the help of code written in the R-Pi. Based on the results, corresponding instructions are sent to the user via Bluetooth device, and the user may act accordingly. The distance from the different objects is also calculated in the R-Pi. When the situation of the emergency arises, the user can use the buttons given on the stick to send the message of threat to concerned people.

Figure 1. Architecture Diagram Of Virtual Assistive Eye



To reduce the load on the processor, the idea of splitting the stages is implemented here. Instead of converting the results of Image Recognition directly into the speech, it is converted into text first and then sent to the handheld device. After that handheld device using the TTS approach sends the output to Bluetooth devices.

The discovery system at the start identifies the target in real-time through the camera. Factor Discovery within which the strategy of distinctive and recognizing real-world factor instances from footage or photos, like a vehicle, bike, TV, flowers, and humans. Growing factor's distinctive options or distinctive points aid within the categorization of the thing. Identification of things is that the one massive sub-territory of

applied science that helps distinguish things in a very series of pictures or videos. There are such a lot of economic algorithms that though the target is partly barred, it is wont to establish the target. Throughout the previous year's different approaches or techniques are introduced. I tend to use the Caffe deep learning system here, which is employed to coach prototypes with deep algorithms within the neural network.

IV. ART OF PROPOSED WORK

CAFFE Model behaviorally a robust framework which attained various benefits to our mechanism. They are listed below,

Expression: prototypes and optimizations are defined as plaintext schemas instead of code.

Speed: for research and industry alike, speed is crucial for state-of-the-art prototypes and massive data.

Modularity: new tasks and settings require flexibility and extension.

Openness: scientific and applied progress call for shared code, reference prototypes, and reproducibility.

Community: academic research, startup prototypes, and industrial applications all share strength by joint discussion and development in a BSD-2 project.

The proposed work is related to deep learning-based items discovery which consists of three main thing discovery techniques that are preferred to be encountered:

Speed R-CNNs (in 2015 presented by Girshick et al.)

YOLO (You Only Look Once) YOLO presented by (Redmon and Farhadi, 2016) SSDs (Single Shot Detectors) presented by Liu et al., 2016. Speed R-CNNs is presently the foremost popular methodology for observing artifacts exploitation deep learning but, such procedures are exhausting to know (specifically for deep learning learners), exhausting to apply and tough to apply.

Additionally, there is a speeder application of R-CNN's the rule is also terribly slow, within the order of seven independent agencies. After there are findings for expected actions to employ, YOLO, as a result of such rule, is far faster, ready to method 39-89 independent agency on a Titan X GPU. YOLO's super-fast version may stand up to one hundred fifty-five independent agencies. The difficulty of YOLO is that they plant plenty of precision are expected. SSDs, which were first created by Microsoft firm, replicate a compromise of each. The rule is straighter than quicker R-CNNs (and I'd say higher explained within the original seminal paper). The entire images placed in the work are an example of factor Discovery with Single Shot Detectors (SSD). By combining the MobileNet architecture and the Single Shot Detector (SSD) framework, we achieved a fast, efficient deep learning-based method on object detection.

4.1 Single Size Multi-Box Detector SSD

It distinct the output space of the bounding boxes into a collection of default boxes over various side ratios and scales per position of the operating map. The network produces scores for the inclusion of every kind of factor in each default box, whereas predicting and generates box changes to raised suit the form of a factor [9]. Additionally, the network incorporates predictions from many operate maps with totally different resolutions to manage artifacts of varied sizes in a very natural method.

4.2 Mobile Nets

MobileNets: Efficient (deep) neural networks
Mobile Nets are wont to produce light-weight, deep neural networks. It's supported by a simplified design that uses separable convolutions of depth-wise style. Mobile internet uses three deeply separable convolutions that use eight times less computing than regular convolution with merely a minor reduction preciseness. Needs and use cases embody seeing, classification of fine grains, face options, and large-scale localization.

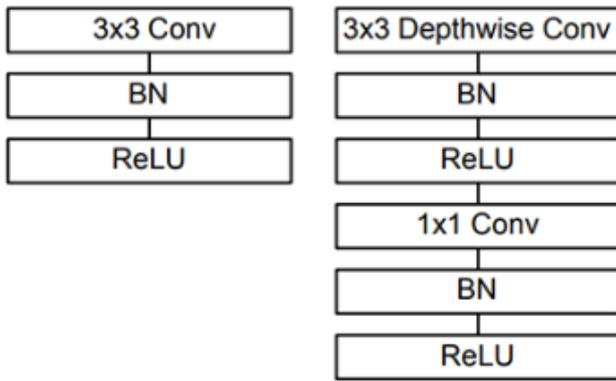


Figure 2. Normal convolutional stages with batch normalization and ReLU (Left). Level-wise independent convolution with layers charted by batch normalization and ReLU (Right).

We sometimes use an existing spec, like VGG or Res-Net, once building factor discovery networks and instead use it at intervals the factor discovery pipeline. The matter is that such network architectures are in order on 199-499 MB vast. Due to their sheer scale and ensuing range of computations, network architectures like they may be unsuitable for facility uses systems. Instead, we can employ Mobile-Nets proposed by Howard et al., 2018, a different piece by investigators from Microsoft. It tends to decision such "Mobile-Nets" setups as a result of they're planned for limited facility systems like such as mobiles. Mobile-Nets are distinct compared to standard CNNs by exploiting depth-separable convolution (Figure a pair of above).

The fundamental conception after level-independent convolution that divides convolution about 2 stages: A three-dimensional convolution thorough. A point-specific convolution of one range l followed us. This conjointly helps them to cut back the number of parameters in our network. The matter is that we tend to lose precision— Mobile-Nets isn't sometimes as precise as because of the prominent members; however, there is abundant extra facility effectively. If we tend to mix each the Mobile-Net design and, therefore, the SSD (Single Shot Detector) system, it hits a quick, effective, deep-method for factor discovery.

4.3 Model Classification

We might distinguish things supported by a variety of various characteristics like naming a fruit. We tend to see its form and color generally. Likewise, we can distinguish various things that supported their type: size color, movement, and texture. We tend to outline them by movement and texture just for quickly moving things. Since we cannot see their type and color. When a factor is outlined, we tend to should classify the factor. Classification is achieved by effort gradients or patterns within the same category of factors if that pattern is contained within the check thing then allotted to its category.

a) Thing Discovery using Shape

A combination of the image and scene-based attribute factor like a picture blob (a massive binary thing) field, the blob ratio within the hopping box, and a television camera magnification is applied to the invention elements at the input. Classification is applied to the virtual image, considering the blob. The findings are bestowed by employing a bar chart.

b) Thing Discovery using Motion

Classification is usually not required once a standard image is taken as an input that has no things in Motion. A robust clue wont to establish any moving body is a human's non-rigid Motion. Whereas the human Motion is distinguished from a factor in Motion looking on this handy clue.

c) Thing Discovery using Shape

While color isn't a correct single-handed life to acknowledge and track artifacts, the less machine value of color-centric algorithms makes the color a tight operate to oppress. As an example, the color histogram-based mostly technique is employed for period automotive identification.

d) Thing Discovery using texture

The texture-based approaches to movement-founded approaches alike with the help of texture pattern recognition toil. It provides improved accuracy by the employment of decussate least standardization of variations. However, it will take over time, which might be stable by employing alternative profligate techniques.

V. Flow Diagram

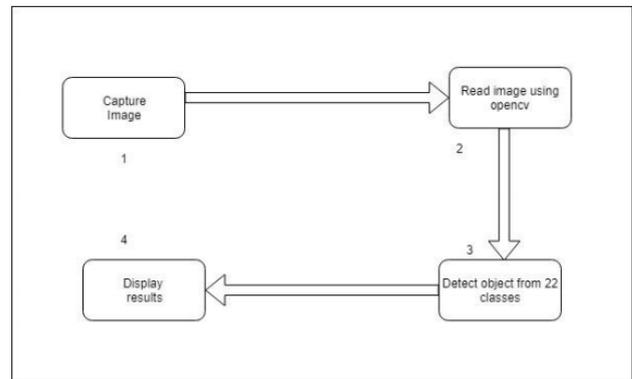


Figure 3. Object Discovery Process

- Step1:- Seizure copy from the spycam
- Step 2:- Study the picture into OpenCV2 using OpenCV Integral operation
- Step 3:- Enable thing discovery using the Caffe prototype
- Step 4:- Display the Image with localization and the Classified label

5.1 Mathematical Model

Whatever the process involved proving the proposed word scientifically is essential. Then only we can taste the actual sweet of the work. Initially, the work started with finding the distance from camcorder to thing/marker using Python and OpenCV.I will use triangle similitude to calculate the distance from our camcorder to a known point or marker. Something like this goes for the triangle: Consider a pointer or an entity through a defined thickness W. This marker also positioned specific length T after the camcorder. Using our camcorder, we take an image of our target and calculate then the outward thickness in dots K.

An Assistive and Assessment Smart Eye for Visually Impaired People using CAFFE Model

That permits them to originate our camcorder's professed central duration L:

$$L = (K \times T) / Q \quad (1)$$

As a sample, assume there we put a standard size of 9 x 12-inch broadsheet (flat; $Q = 12$) $T = 23$ inches before the camcorder and shot an image. Then there is an observation on the thickness of the broadsheet size through the picture, and then it is found that the obtained thickness of the broadside is $K = 259$ dots.

My focal length F is then:

$$L = (259px \times 23in) / 12in = 496.42$$

Then we extended camcorder together nearer and beyond from the pointer, and then we can employ triangle similarity to evaluate the coverage length of the thing about camcorder:

$$T = (K \times L) / K \quad (2)$$

Another time, to ensure high real, assume we extend our camcorder 4 ft (i.e., 46 inches) far from the pointer and shoot an image of equal size of the broadsheet. Using a spontaneous picture treating, we can conclude that the observed size of the broadsheet is now 200 dot. Working about the expressions we now obtain:

$$T = (12in \times 549.56) / 185 = 46in$$

Or roughly 46 inches, which is 2 feet.

VI. EXECUTIONAL SETUP

This model will specialize in pictures for the implementation half and ask for identification and localization of factors similarly as we can take into account the gap of a thing from the television camera.

1. OpenCV-Python

I used python3 because of the programming language since it consists of various laptop vision libraries and is often used. The whole project is enforced on the operative raspberry-pi raspbianOS. We tend to download the OpenCV image from OpenCV version three.3 of the official Github project.

2. Reading an Image

We are using inbuilt cv2 function `cv2.imread()` to read the image. We can capture the image by using the following command:- `raspistill -o ImageName.jpg`

`CV.IMAGEREAD_COLOR`: Is used to load color images and is used by default if nothing else specified

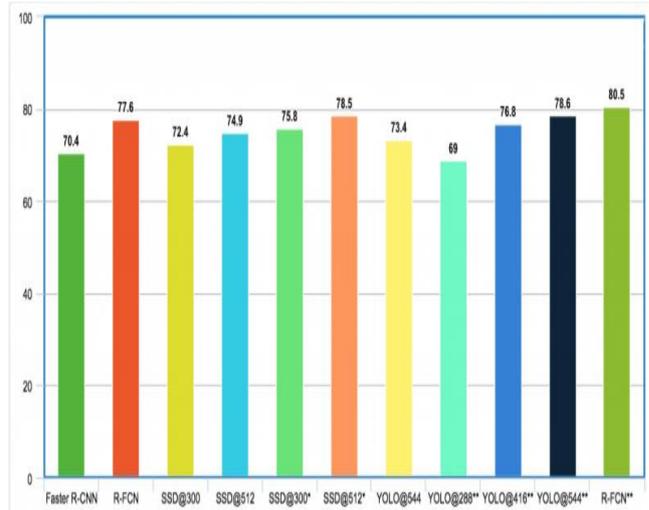
`CV.IMAGEREAD_GRAYSCALE`: Is the image is loaded in grayscale mode. Useful for edge discovery algorithm.

`CV.IMAGEREAD_UNCHANGED`: Is Image is loaded as it is.

3. Attribute recognition and description

They were interpreting the attributes as what are a number of the essential attributes the image uses? However helpful, is it to outline specific attributes? Identification of edges (Okay, Corners square measure standard attributes, however, finding them in a real matrix of pixels).

6.1 Comparison between all primary object detection



In the above picture, we can observe that R-FCN outperforms the accuracy.

- Let us take a look at the practical code implementation so we can get an overview to implement this single-shot object detection algorithm.
- The first step is to load a pre-trained object detection network with the OpenCV's deep neural network module.
- This will allow us to pass input images through the network and obtain the output bounding box (x, y)-coordinates of each object in the image.
- Now we write the code to print the name of the detected object and their confidence scores.
- At last, we look at the output of MobileNet Single Shot Detector for our input images.
- SSD with MobileNet provides the best accuracy tradeoff within the fastest detectors.
- SSD is fast but performs worse for small objects comparing with others.
- For large objects, SSD can outperform Faster R-CNN and R-FCN inaccuracy with lighter and faster extractors.

VII. RESULT

Figure 4 shows the program output. It can observe that everything is marked and placed with a black rectangular border, and therefore, the category mark is shown at the highest with the preciseness of that factor, as shown within the category mark.

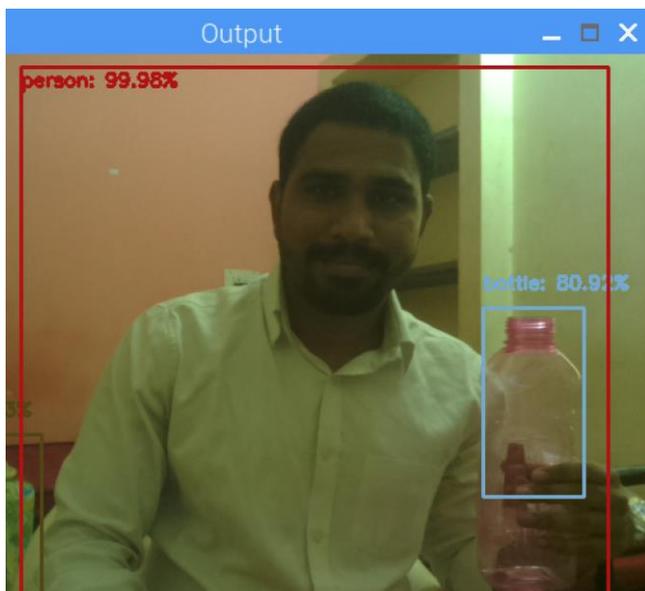


Figure 4. The output of Thing discovery

Figure 5 shows the performance of the other program went to find the gap between the television camera and the detected factor. In this case, the tablet is 3.71 feet far away from the television camera, as we can see.



Figure 5. The output of distance to camcorder thing

VIII. CONCLUSION

The new program guarantees a higher life for its customers. The majority of this system's users arrange to be students with these disabilities. Deep learning-based identification of artifacts has become a hotspot of labor in recent years. This project begins with generic pipelines for factor discovery to supply base architectures for alternative connected tasks. With this, the opposite 3 vital tasks are accomplished, particularly factor discovery, face discovery, and pedestrian discovery. Discovery of things with deep learning, and OpenCV and Secure, rib OpenCV video streams. Noise within the television camera sensing element and Lightning condition can affect the result. As a result of it will trigger a drag in factor recognition. The top result's a deep learning-based detector of artifacts which might include method regarding 6-8 independent agency. After I took the photographs for this instance, my tapeline was somewhat slack in it, that the measurements are off by regarding one in.. I conjointly took the pictures in haste and not one hundred percent on high of the tapeline feet labels, which contributed

to the one in mistake. That said, the similitude to constellation still holds, and you'll be able to use this approach to live the gap to your television camera from a factor or marker only.

FUTURE WORK

One of the items within the spotlight is increasing the preciseness of image recognition. Another priority space is to slash the reasonable device price and facilitate as many folks as doable. Additionally, the app upgrades options like Step Discovery, Slope Discovery, and stress also will get on creating the stick water-proof. The scholars with these impairments are the meant cluster for the Hearing incapacity Support theme. The future analysis involves introducing distinctive features of expression. There is also quite one speaker speaking; thus, it's exhausting to differentiate between them. Besides, this may even be adding a feature wherever the user can respond with the speech. As this system's spectrum is wide open, there are lots of options to feature here. The innovations are applied with the appearance of the most recent technology, and therefore the lives of individuals with many skills are created easier, improved, and higher.

REFERENCES

1. World Health Organization (WHO), "Visual impairment and blindness Research," Oct. 11 2017. [Online] Available: URL: <http://www.who.int/mediacentre/factsheets/fs282/en/>.
2. W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: current status, challenges, and future directions," *Sensors*, vol. 17, article 565, 2017.
3. R. Khlaikhayai, C. Pavaganun, B. Mangalabruks, and P. Yupapin, "An intelligent walking stick for elderly and blind safety protection," in *Proceedings of the 2nd International Science, Social Science, Engineering and Energy Conference 2010: Engineering Science and Management*, *Procedia Engineering*, vol. 8, pp. 313-316, 2011.
4. J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, "Smart guiding glasses for visually impaired people in the indoor environment," *IEEE Trans. on Consumer Electronics*, vol. 63, no. 3, pp. 258-266, 2017.
5. J. Ramprabu, T. Gowthaman, "Smart Cane for Visually Impaired People," in *International Journal of Computer Science and Information Technologies*, vol. 4, no. 1, pp. 24-28, January 2013.
6. 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.
7. 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 a visual marker," *IEEE Consumer Electronics Magazine*, vol. 7, no. 3, pp. 12-20, 2018.
8. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). A focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).
9. Dai, J., Li, Y., He, K., & Sun, J. (2016). R-CNN: Object detection via region-based, fully convolutional networks. In *Advances in neural information processing systems* (pp. 379-387).
10. Rahul, Y., & Nair, B. B. (2018, September). Camera-Based Object Detection, Identification, and Distance Estimation. In *2018 2nd International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE)* (pp. 203-205). IEEE.
11. Singh, B. S. (2018). IoT based Smart Healthcare Applications for People with Disabilities. *Asian Journal for Convergence in Technology (AJCT)*.
12. Tan, H. X., & Tan, H. P. (2018, February). Early detection of mild cognitive impairment in the elderly through IoT: Preliminary findings. In *2018 IEEE 4th World Forum on the Internet of Things (WF-IoT)* (pp. 207-212). IEEE.

An Assistive and Assessment Smart Eye for Visually Impaired People using CAFFE Model

13. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," NIPS, 2015, pp. 91–99.
14. P. Viola, M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Computer Vision and Pattern Recognition, 2001, pp. 1-9.
15. T. Guo, J. Dong, H. Li, and Y. Gao, "Simple Convolutional Neural Network on Image Classification," IEEE 2nd International Conference on Big Data Analysis (ICBDA), 2017, pp. 721-724.
16. S. S. Liew, M. Khalil-hani, and S. A. Radzi, "Gender classification : a convolutional neural network approach," Turkish Journal of Electrical Engineering and Computer Sciences, vol. 24, no. 3, 2016, pp. 1248–1264

AUTHORS PROFILE



Vaibhav Sontakke is currently doing M. Tech in Computer Science and Engineering from Vellore Institute of Technology, Vellore. And he also held B.E Electronics and Telecommunication Engineering degree from Priyadarshini College of Engineering, Nagpur. I have developed an interest in machine learning from my graduation. The interest in machine learning made me choose computer science engineering in my post-graduation.



Siva Shanmugam G, working as an Asst. Prof.& Research Scholar at SCOPE, VIT University, Vellore, TN, India. He has had 6 years of teaching experience and currently doing his Ph.D. research on Green Cloud computing. His interest includes Machine learning, Distributed Systems, Web Security, Sensor Networks, Mobile, and Internet Computing.