OBJECT DETECTION FOR VISUALLY IMPAIRED PEOPLE USING SSD ALGORITHM

K.Vijiyakumar[1]
Assistant Professor
Department of Information Technology
Manakula vinayagar Institute of Technology, Puducherry, India
vijiya.kumar@gmail.com

K.Ajitha[2], A.Alexia[3],M.Hemalashmi[4],S.Madhumitha[5]
Department of Information Technology
Manakula vinayagar Institute of Technology, Puducherry, India
ajitha18041999@gmail.com, ria26aroquianadin@gmail.com, hemalakshmi0709@gmail.com, vilasri1968@gmail.com

Abstract— **Visually impaired people are unaware of the danger that they are facing in their life. They may face many challenges while performing their daily activity even in their familiar environments. Vision is the necessary human senses and it plays the important role in human perception about surrounding environment. Hence, there are variety of computer vision products and services which are used in the development of new electronic aids for those blind people. In this paper we designed to provide navigation to those people. It guides the people about the object as well as provides the distance of the object. The algorithm itself calculates the distance of the object. Here it also provides the audio jack to insist them about the object. Here we are using SSD Algorithm for object detection and calculating the distance of the object by using monodepth algorithm.**

IndexTerms —Blind; Object Detection; Object Recognition; Image Processing.

I. INTRODUCTION

Man-made brainpower essentially alludes to a fake formation of human-like knowledge which can learn reason, plan, see, or procedure normal language. AI and Artificial Intelligence are assuming significant job in around the world. How about we start by a case of Virtual Personal Assistants (appeared in Fig1) which have been recognizable to us.

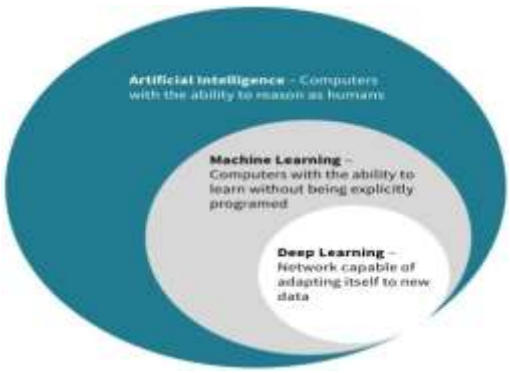

Figure1: Artificial Intelligence

FUTURE SCOPE

- Artificial Intelligence is here to stay and is going nowhere. It digs out the facts from algorithms for a meaningful execution of various decisions and goals predetermined by a firm.
- Artificial Intelligence and Machine Learning are likely to replace the current mode of technology that we see these days, for example, traditional programming packages like **ERP** and **CRM** are certainly losing their charm.
- Firms like Facebook, Google are investing a hefty amount in AI to get the desired outcome at a relatively lower computational time.
- Artificial Intelligence is something that is going to redefine the world of software and IT in the near future.

MACHINE LEARNING

In 1959, Arthur Samuel characterized the AI as a "Field of study that enables PCs to learn without being unequivocally programmed"."The objective of AI is to assemble PC frameworks that adjust and gain as a matter of fact". Tom Dietterich

AI is the subset of Artificial Intelligence. It is the field worried about the plan and improvement of calculations that permit PCs to expand work dependent on exact information. For example, from sensors and databases. A significant focal point of the AI is to consequently figure out how to perceive complex examples and settle on canny choices dependent on those information (appeared in Fig 2)

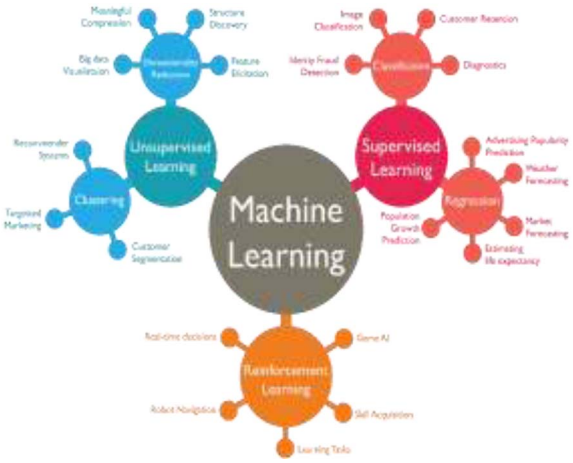

Figure2: Machine learning

By and large, calculation assumes a significant job in Machine Learning Algorithms can be separated into various classifications:

- Supervised learning
- Unsupervised learning
- Semi-supervised learning
- Reinforcement learning

Supervised learning

Supervised learning algorithms are prepared with named models (input) where the necessary yield is known.

Unsupervised learning

Unsupervised learning algorithms work with unlabelled models (input) where the necessary yield is obscure.

Semi-supervised learning

It consolidates both name and unlabelled contributions to produce an adept capacity or classifier.

Reinforcement learning

Reinforcement learning algorithm is fomented with how astute operators should act in a situation to expand some documentation of total prize

Other algorithms are

- Learning to learn
- Development learning
- Transduction etc.

ADVANTAGES OF MACHINE LEARNING

- Easily identifies trends and patterns
- No human intervention needed (automation)
- Continuous Improvement
- Handling multi-dimensional and multi-variety data
- Wide Applications

DEEP LEARNING

Deep learning is a computerized reasoning capacity. Deep learning is a subgroup of AI that has capacity of gaining from information which are unlabelled. Deep learning has extended connected at the hip with the computerized time, which has acquired a fly of information all structures the world over. These information, are drawn from sources like web based life, web crawlers, online business stages, and among others (shown in Fig 3). This huge measure of information is promptly convenient through fine tech applications like distributed computing.

Deep learning calculations attempt to learn (different degrees of) portrayal by utilizing a progressive system of numerous layers. On the off chance that we gracefully the framework huge amounts of data, it starts to get it and react to it in helpful way.

WHY IS DEEP LEARNING USEFUL?

- Features which are found out are anything but difficult to obtain, brisk to learn
- Deep learning gives a truly adaptable, widespread, learnable system for protrayal world, visual and semantic data.
- Can learn unaided and directed
- Efficient for the start to finish joint framework learning
- Use a lot of preparing information

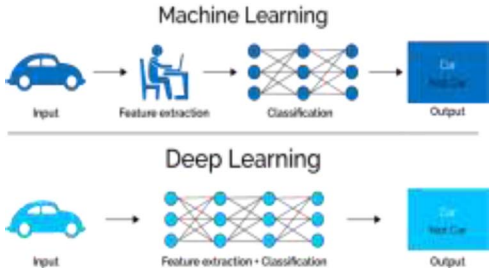

Figure3: Deep learning

ARCHITECTURES

1. **Deep Neural Network** – It is a neural network with a certain level of complexity (having multiple hidden layers in between input and output layers). They are capable of modeling and processing non-linear relationships.
2. **Deep Belief Network(DBN)** – It is a class of Deep Neural Network. It is multi-layer beliefnetworks.
3. **Steps**
   a. Learn a layer of features from visible units using Contrastive Divergence algorithm.
   b. Treat activations of previously trained features as visible units and then learn features offeatures.
   c. Finally, the whole DBN is trained when the learning for the final hidden layer is achieved.
4. **Recurrent** (perform same task for every element of a sequence) **Neural Network** – Allows for parallel and sequential computation. Similar to the human brain (large feedback network of connected neurons). They are

ISBN 978-1-7281-6202-7

able to remember important things about the input they received and hence enables them to be more precise.

II. LITERATURE REVIEW

In this paper they grew new methodology for distinguishing a different article in a solitary picture utilizing Convolutional neural systems. In this model they first develop a region proposals using Edge box algorithm. Then it passed to the CNN algorithm and it passed to softmax to extract the images which is the last layer of the CNN algorithm. At last they get a mean average precision (mAp) for each class. The accuracy of this model is 37.38%. After that they reduce the complexity of the problem using R-CNN (Region Convolutional neural network). In second time they create selective search using Edge box algorithm and it directly passed into softmax which is the last layer of the CNN algorithm. When compared with first second one gives higher accuracy then first one [2].

In this paper they detect the pattern of the images using Convolutional neural network. When compared with other it gives high accuracy then that of others. They used various layer in this paper. In this paper they discussed one example such as they introduce the key characteristics using CNN algorithm. They using Imaging and Computer vision to do task such as imaging and recognition. They also give high accuracy the that of other technology [5].

In this paper they are modelling a visually intelligent agent. They typically focused on the computer vision. They first extract all the characteristics and is passed as a dataset. In this paper they show variety of metrics from the visually intelligent agent in many situations. In this model they encodes and represent the other domains. In a particular they show strong task on the surface estimation and scene classification using dog modelling task as representation learning using YOLO algorithm [3].

III. PROBLEM STATEMENT

Vision is important sensory organ for human. Visually impaired people suffer many problems both in indoor and outdoor. In existing system, the object is detected up to 15cm but it fails to detect the object in wider distance. So, in this project we are extending the distance for the object detection.

IV. EXISTING WORK

The work that is being present in the system is based on the new technologies that helps to improve the visually impaired people. Our project is used to detect the obstacles that is used to reduce the navigation difficulties for the impaired people. In this existing system, a webcam was used to capture the object that the blind people are facing every day. For object detection, after the webcam captured the object, the feature was extracted and compared in database. If the object matched with the database, a speech was generated by the system to assist the visually impaired people to identify the object. This prototype was able to recognize visual objects and presented the detection information as sound. A webcam was able to capture the object and the microcontroller was used to perform the object detection and recognition by retrieving the Caffe model. The trained model is able to retrieve the sound-based visuals from the Microsoft Azure cloud storage through Wi-Fi. The system could perform the object detection under offline condition if the previous Caffe model has been downloaded and stored in the local storage of Raspberry Pi 3.

V. PROPOSED WORK

In this paper we present a keen living for the outwardly hindered individuals by helping them in their everyday life utilizing object identification framework. The detection system will first detect the object through the camera in real time. The video will then be processed by making comparison with the trained models in the cloud database. Normally a visually challenged people use cane as a guide to protect them from obstacles. Especially to detect the object near their legs like stairs, chairs, etc. in the indoor environment. But in the outdoor environment it's somewhat difficult to detect the objects with the walking cane. In this project we designed to provide full navigation to those people. It guides the people about the object as well as provides the distance of the object. The algorithm tracks the distance of the object. Here it also provides the audio jack to insist them about the object (shown in fig 3). Here we are utilizing SSD Algorithm for object location and for computing the separation of the article by utilizing monodepth calculation. The blind people use the real time camera on their daily using things so that the camera records the video of their path and thus passed as input to this project. The algorithm processes the input image and provide voice message to the impaired people. The object gets detected by the technique. And it matches the object with the database images to confirm the obstacles that comes into the way. After that processing it gives the voice instruction. So, the visually impaired people get the direction. These are the features provided by the designed project.

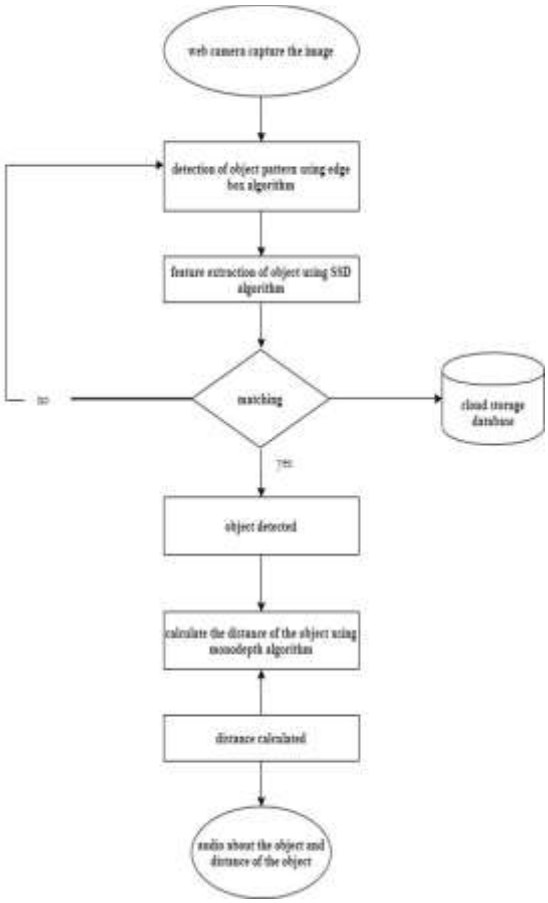

Figure4 Working of proposed work

VI. SSD WORKING

• **Single Shot:** this implies the undertakings of article limitation and grouping are done in a solitary forward go of the system

• **Multi Box**: this is the name of a procedure for bouncing box relapse created by Szegedy et al. (we will quickly cover it in the blink of an eye)

• **Detector**: The system is an article finder that additionally groups those identified items

SSD (Single Shot Detector) is a well known calculation in object recognition. In SSD we should need to make a solitary effort to identify various articles in the picture, and the local proposition organize (RPN) it need two shots, one for producing district recommendations, another for recognizing the item for every proposition. Furthermore, accordingly the SSD calculation is a lot quicker when contrasted and two shot RPN. The SSD engineering originates from Single shot this is only the errand of item restriction and grouping done in a forward system.

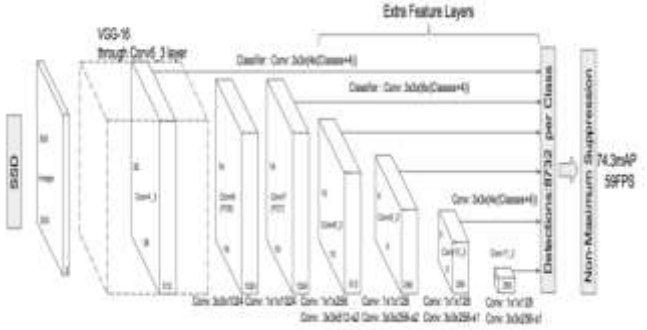

Figure 5: Block Diagram of SSD

In the above chart, SSD design expands on the revered VGG-16 engineering. VGG-16 is the Visually Graphic Group and 16 originates from the convolutional neural system it comprises of convolutional layer, max pooling, completely associated layer, extra layer. In this design the completely associated layer is been disposed of due to VGG-16 was utilized as the base system as a result of its solid execution in top notch picture arrangement assignments and its fame is to move learning in improving the outcomes. A portion of the convolutional layers were included rather than the first VGG completely associated layers. In each ensuing layer is to used to diminish the size of the information and furthermore remove include at numerous scales.

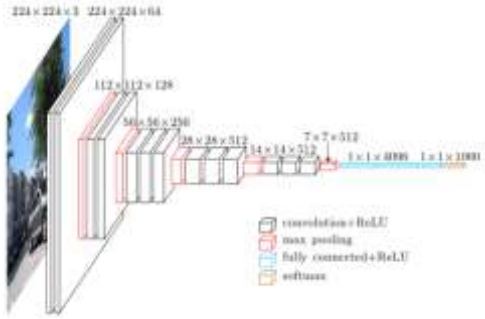

Figure6: Working of SSD

VGG ARCHITECTURE

MULTIBOX

MultiBox is utilized to make various forecasts containing limit boxes and certainty scores. SSD doesn't appoint locale proposition organize. It processes both the area and class scores utilizing little convolutional channels. The bouncing box relapse procedure of SSD is enlivened by Szegedy's proposition. The wok done on MultiBox a convolutional arrange is utilized. The 1x1 convolutions help in dimensional decrease since the quantity of measurements will go down however the "width" and "stature" will stay same.

MuliBox's misfortune work is joined with two basic parts which advanced toward SSD:

Certainty Loss: This is utilized to gauge the sure of the system were the article is processed in the bouncing box

Area Loss: this is utilized to gauge the distance away the system's anticipated bouncing box starting from the earliest stage to the preparation set

multibox _ misfortune = certainty _ misfortune + alpha * area _ misfortune

ISBN 978-1-7281-6202-7

The term alpha causes us in adjusting the commitment of the area misfortune. In profound learning, the objective is to discover the qualities which lessen the misfortune work, in this way bringing the expectation closer.

Edge Box

The utilization of article recommendations is a compelling late methodology for expanding the computational effectiveness of item location. We propose a novel technique for creating object bouncing box recommendations utilizing edges. Edges give a scanty yet enlightening portrayal of a picture. Our principle perception is that the quantity of forms that are completely contained in a jumping box is demonstrative of the probability of the crate containing an item. We propose a basic box objectness score that quantifies the quantity of edges that exist in the case short those that are individuals from shapes that cover the case's limit. Utilizing effective information structures, a great many competitor boxes can be assessed in a small amount of a second, restoring a positioned set of a couple thousand top-scoring recommendations. Utilizing standard measurements, we show results that are altogether more precise than the present status of-the-craftsmanship while being quicker to process. Specifically, given only 1000 proposition we accomplish over 96% article review at cover limit of 0.5 and over 75% review at the all the more testing cover of 0.7. Our methodology runs in 0.25 seconds and we furthermore show a close to constant variation with just minor misfortune in precision.

- Probablity that there is an item,
- Height of the jumping box,
- Width of the jumping box,
- Horizontal arrange of the inside purpose of the bouncing box,
- Vertical organize of the middle purpose of the bouncing box.

By and by, there are two sorts of standard article discovery calculations. Calculations like R-CNN and Fast(er) R-CNN utilize a two-advance methodology - first to recognize districts where items are required to be found and afterward identify protests just in those areas utilizing convnet. Then again, calculations like YOLO (You Only Look Once) and SSD (Single-Shot Detector) utilize a completely convolutional approach in which the system can discover all articles inside a picture in one pass (thus 'single-shot' or 'look once') through the convnet. The district proposition calculations as a rule have somewhat better exactness however more slow to run, while single-shot calculations are progressively productive and has as great precision and that is the thing that we are going to concentrate on in this area.

To follow the guide beneath, we expect that you have some fundamental comprehension of the convolutional neural systems (CNN) idea. You can invigorate your CNN information by experiencing this short paper "A manual for convolution number juggling for profound learning".

Grid cell

Rather than utilizing sliding window, SSD separates the picture utilizing a network and have every lattice cell be liable for recognizing objects in that locale of the picture. Recognition questions essentially implies foreseeing the class and area of an article inside that district. In the event that no item is available, we consider it as the foundation class and the area is overlooked. For example, we could utilize a 4x4 lattice in the model beneath. Every framework cell can yield the position and state of the article it contains.

Presently you may be thinking about imagine a scenario in which there are various articles in a single framework cell or we have to identify numerous objects of various shapes. There is the place stay confine and open field come to play.

Anchor box

Every lattice cell in SSD can be relegated with different stay/earlier boxes. These grapple boxes are pre-characterized and every one is liable for a size and shape inside a network cell. For instance, the pool in the picture underneath compares to the taller grapple box while the structure relates to the more extensive box.

SSD utilizes a coordinating stage while preparing, to coordinate the fitting stay box with the jumping boxes of each ground truth object inside a picture. Basically, the grapple box with the furthest extent of cover with an article is liable for foreseeing that item's class and its area. This property is utilized for preparing the system and for foreseeing the distinguished items and their areas once the system has been prepared. By and by, each grapple box is determined by an angle proportion and a zoom level.

Aspect Ratio

Not all items are square fit as a fiddle. Some are longer and some are more extensive, by changing degrees. The SSD design permits pre-characterized viewpoint proportions of the stay boxes to represent this. The proportions boundary can be utilized to indicate the diverse perspective proportions of the grapple boxes partners with every lattice cell at each zoom/scale level.

Zoom level

It isn't important for the stay boxes to have a similar size as the lattice cell. We may be keen on finding littler or bigger articles inside a matrix cell. The zooms boundary is utilized to indicate how much the grapple boxes should be scaled up or down regarding every framework cell. Much the same as what we have found in the grapple box model, the size of building is commonly bigger than pool.

Receptive Field

Receptive field is characterized as the locale in the information space that a specific CNN's component is taking a gander at (for example be influenced by). We will utilize

"highlight" and "actuation" reciprocally here and treat them as the straight blend (once in a while applying an initiation work after that to increment non-linearity) of the past layer at the relating area. As a result of the convolution activity, highlights at various layers speak to various sizes of locale in the information picture. As it goes further, the size spoke to by an element gets bigger. In this model underneath, we start with the base layer (5x5) and afterward apply a convolution that outcomes in the center layer (3x3) where one component (green pixel) speaks to a 3x3 area of the information layer (base layer). And afterward apply the convolution to center layer and get the top layer (2x2) where each element compares to a 7x7 district on the info picture. These sort of green and orange 2D cluster are additionally called highlight maps which allude to a lot of highlights made by applying a similar component extractor at various areas of the info map in a sliding window fastion. Highlights in a similar element map have the equivalent open field and search for a similar example yet at various areas. This makes the spatial invariance of ConvNet.

CONVOLUTION

The term convolution alludes to the scientific mix of two capacities. This is utilized to create a third capacity. It blends two arrangements of data. The convolution is performed on the info information with the utilization of a channel or bit at that point produce a component map.

ReLu

The ReLu (Rectifier Linear Unit) alludes to the initiation work for the yields of the CNN neurons. This capacity will yield the information straightforwardly in the event that it is certain in any case the yield will be zero.

POOLING

A pooling layer is another structure square of a CNN. Its capacity is to dynamically lessen the spatial size of the portrayal to decrease the measure of boundaries and calculation in the system. Pooling layer works on each component map autonomously. The most widely recognized methodology utilized in pooling is max pooling

FULLY CONNECTED

Toward the finish of a CNN, the yield of the last Pooling Layer goes about as contribution to the supposed Fully Connected Layer. There can be at least one of these layers.

SOFT MAX

A softmax layer , is an extra layer in the CNN arrange which give high exactness by including the foundation subtleties of the information.

$$P(y = j \mid \theta^{(i)}) = e\theta^{(i)} \frac{e\theta^{(i)}}{\sum_{j=0}^{k} e\theta_k^{(i)}}$$

$$\theta = w_0 x_0 + \cdots + w_k x_k = \sum_{i=0}^{k} w_i x_i = w^T x$$

SSD OBJECT DETECTION

The SSD object detection composes of 2 parts:

1. Feature Extraction
2. To detect the object

1. Feature Extraction

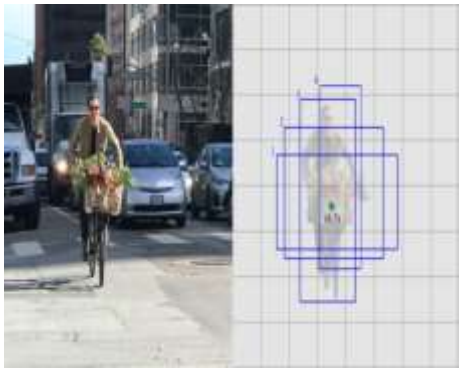

Figure7: Feature Extraction of SSD

From the object pattern the features are extracted using SSD algorithm. SSD algorithm works on the principle of CNN algorithm.

2. TO DETECT OBJECTS

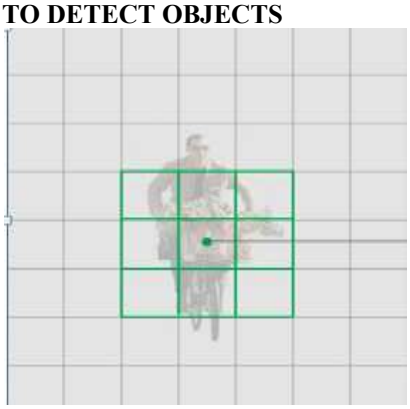

Figure8: Convolution Filter of SSD

By applying Convolutional neural network we can detect the object.

VII .CONCLUSION

In this project, an object detection system for visually impaired people based on SSD algorithm in real time has been proposed. The system has retrieved the trained model from the cloud database to perform object detection in real time. The proposed system is beneficial for the visual impaired people for better living quality to detect the object as well as calculating the distance of the object.

References

[1] Bc. Jan Hadáček, "Application of a Camera in a Mobile Phone for Visually Impaired People." Masters thesis, Czech Technical University in Prague, May 2017.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[2] "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, IEEE transactions, Dec 2016.

[3] “You Only Look once: Unified, Real-Time Object Detection.” J Redmon, S Divvala, R Girshick, A Farhadi, IEEE transactions, May 2016.R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[4] “SSD: Single Shot MultiBox Detector Wei Liu.”, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, IEEE transactions, Jan 2016.

[5] Image recognition: By Samer Hijazi, Rishi Kumar, and Chris Rowen, IP Group, Cadence “Using convolutional neural network for image recognition”.

[6] Vicky Mohane, Chetan Gade “Object Recognition for Blind people Using Portable Camera” WCFTR World conference 2016.

[7] Meghajit Mazumdar, Dr. Sarasvathi V, Akshay Kumar “Object Recognition in Videos by Sequential Frame Extraction using Convolutional Neural Networks and Fully Connected Neural Networks” International Conference on Energy, Communication, Data Analytics and Soft Computing 2017.

[8] Yide Ma, Dong Hwan Kim, and Sung-Kee Park “Region-Based Object Recognition by Color Segmentation Using a Simplified_PCNN” IEEE transactions on neural network and learning system, Vol, 26 No. 8 Aug 2015.

[9] Mingjie Liang, Huaqing Min, Ronghua Luo, and Jinhui Zhu, “Simultaneous Recognition and Modeling for Learning 3-D Object Models From Everyday Scenes” IEEE transcation on cybernetics, Vol 45 No. 10 Oct. 2015.

[10] http://cocodataset.org/#home Show Context

[11] https://www.kaggle.com/jessicali9530/coil100

[12] http://www.vision.ee.ethz.ch/en/datasets/

[13] https://blog.statsbot.co/real-time-object-detection-yolo-cd348527b9b7