Various Methods for Object Detection Based on Deep Learning

Arlin Maria Scari, Neena V V

Abstract: ***The growing technology in the world made-up the deep learning method, which classifies different vehicles from a video. In deep learning technology use different strategies such as RCNN, Fast RCNN, RPN, faster RCNN, YOLO, SSD. All methods offer various accuracy of the identification of the vehicle. The convolutional natural network determining an object detection task exploitation in deep learning. Object detection is very important in AI as well as in videos using pc vision. Through this paper demystifies the important role of deep learning supported by CNN for object detection. And the methodology offers additional correct result. Deep learning techniques shows the development of object detection in various area and the different technics are assessed during this paper.***

Keywords: Machine Learning, Deep learning, CNN, RCNN, Fast RCNN, Faster RCNN, YOLO, SSD.

I. INTRODUCTION

Vehicle detection is incredibly necessary within the traffic scene to classify the vehicle supported its structure and different options by newest deep learning technology. Robert working is based on AI, which is the latest technology and same as human intelligence processes by machines, particularly laptop systems. These methods embrace automatically learn, reasoning and self-correction. The important use of AI is to make a knowledgeable system, speech recognition and machine vision etc. In machine Learning, it's seen as a set of computing. And the deep learning is also same as machine learning method that coach the systems how do the human gain knowledge naturally using examples [21]. Deep learning shows important role behind the driverless cars because it is a key technology to detect the objects, like to acknowledge a stop sign or to inform apart a pedestrian from a post [21]. And Deep Learning could also be a subdivision of machine learning, which involved algorithms by the structure and performance of the human brain brought up as artificial neural networks.

The goal of machine-controlled surveillance work and watching systems is to get rid of the requirement of human labour for natural vision primarily based tasks which will be performed by a pc or an automatic system. The applications of pc vision systems have conjointly applied in numerous public areas like roads, airports and retail sectors. One such form of vision systems is within the task of watching and analysing scenes of road traffic, with a specific interest in watching highways and intersection. Such a system is needed for effective real-time traffic management systems. which will find changes in traffic characteristics in an exceedingly timely manner, permitting regulators and authorities the power to respond to traffic things [5] quickly. The core of any such system which will be used to effectively detect exact object and classifying the moving vehicles from the video [16].

A. Artificial Intelligence

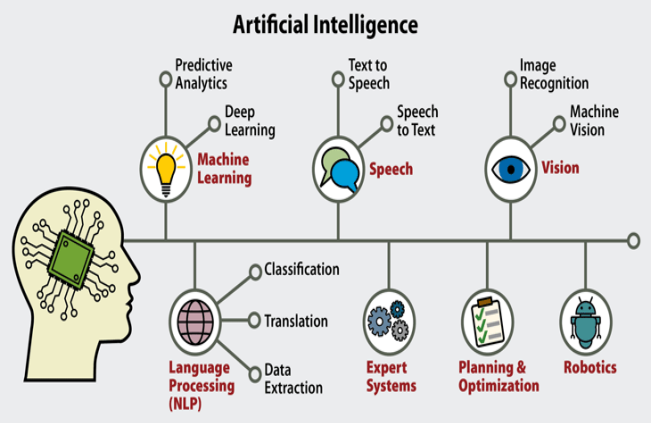

Fig. 1. Artificial Intelligent [21]

Artificial Intelligent is known as machine intelligence and here the intelligence exhibit by machine, the human displays natural intelligence. Similarly, the artificial intelligence is using to describe systems with autonomous such as robots (or computers), while it can do the activities like human being can do with their thoughts, such as understanding some situation and make solution for that. The AI field is originated in the case, where the human knowledge could be so absolutely depicted that a machine can act in that way. This raises theoretical contentions about the idea of the autonomous. It is generally referred to as machine learning, here is the intelligence incontestable by machines in distinction to the natural information displayed by humans. Intelligence is to describe machines (or computers) with their activities like human do learn the problem and find solution for the problem. The AI is supported on the claim that human intelligence can be thus exactly delineated that a machine is often created to simulate it [21]. Which is raising the theoretical arguments concerning the character of the mind and also the ethics of making artificial beings blessed with human-like intelligence that area unit problems.

Revised Manuscript Received on November 15, 2019

Arlin Maria Scaria,Computer Science and Engineering, Vimaljyothi Engineering College,Email: arlin.maria18@gmail.com.

Neena V V Computer Science and Engineering, Vimaljyothi Engineering College,Email: neenaphalgun@vjec.ac.in

Fig one shows that however, the AI is expounded to human Brain and also the activities occur in it, where to create synthetic brain beings endowed with human intelligence, issues that have been explored given that historical instances through myth, fiction and philosophy.
Fig 1 shows how the AI is related to human Brain, and the activities occur in it.

B. Machine Learning

It could be a subarea of AI. that have the power to find out mechanically while not expressly programmed. There are Machine learning classification:
i) Supervised machine learning algorithms, it will apply what has been learned within the past to new information mistreatment tagged examples to predict future events. Where the algorithmic program may check with its output with the right, meant output and notice errors to switch the model consequently.
ii) Here Unsupervised machine learning algorithms are used when the information used to train is neither classified nor labelled, and it hugely differs from supervised machine learning. It may or may not provide the correct output.
iii) Semi-supervised machine learning algorithms, which in between supervised and unsupervised learning. Here use labelled data training and unlabelled data training. There a small amount of labelled data and a large amount of unlabelled data. This method helps to improve learning accuracy.

C. Deep Learning

It is also same as Machine Learning, as part of a broader family of machine learning methods based on artificial neural networks. Currently, Deep Learning with a convolutional architecture (CNN) have emerged as a popular method for solving problems related to visual object recognition either in images or videos and has given the state of the art performance in various optical recognition tasks, such as image classification, object detection and localisation and image segmentation [16]. Learning can be supervised, semi supervised or unsupervised.

II. OBJECT DETECTION METHODS

The methods for object detection generally fall into either machine learning-based approaches or deep learning-based Approaches. For Machine Learning approaches, it becomes necessary first to explain features using among the methods are given below. Then use a technique such as support vector machine (SVM) was traditional method which do the classification. On the other side, deep learning methods that can perform end-to-end detection of objects without specific characteristics and are typically based on convolutional neural networks (CNN). In Deep Learning Region Proposals methods are R-CNN, Fast R-CNN and Faster R-CNN, the other techniques are Single Shot MultiBox Detector (SSD), and You Only Look Once (YOLO). Here the object detection, it involves detecting the objects in an image or video along with their location, typically using a bounded box. So the particular region is selected with the bounding box.

A. CNN

CNN is the convolutional neural network and it is also known as ConvNet. Now a day, this method has been widely used to all kinds of vision problems in system and image processing. There the images are classified and detect the object from the image. So the role of CNN is the feature extractor from the image and it became like a deep neural network [18]. Similar to Artificial Neural Network, the CNN has input layer and output layer and various hidden layers. The CNN is developed from the biological process like the connectivity pattern of neurons. CNN also include pooling layers. In CNN the input layer consists of the given image model where the number of neurons are equal to the number of features. In the hidden layers contain the inputs from the first layer are fed into hidden layers and it is depending upon the model data size. From this output from each layer generates a matrix multiplication and which make a nonlinear network Then the output layer contain the functions occur like softmax and sigmoid, then convert into probability score. [17].

B. RCNN

RCNN is Region convolutional neural network. This algorithm reduce the drawback of CNN, where difficult to run many patches using sliding window detector. So here is the solution for the problem as using object proposal algorithm. Selective search algorithm is the object proposal algorithm in RCNN, that generates up to 2000 region proposals from one image during test time. It can reduce the count of bounding box. These region proposals are then warped using an affine transform into 227x227 size images, that are then fed into a pre-trained neural network [19]. Instead of

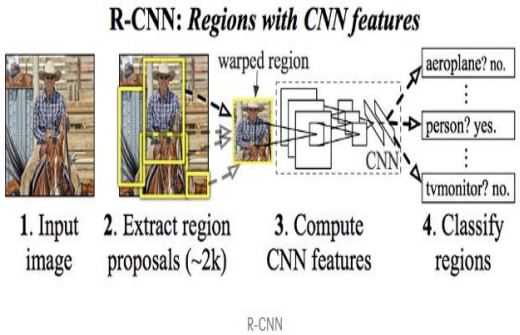

Fig. 2. RCNN [1]

functioning on an enormous variety of regions, the RCNN rule proposes a bunch of boxes within the image. And checks if any of those boxes contain any object RCNN uses selective search to extract these boxes from an image (these boxes are called regions). The selective search algorithm proposes the bounding box. These 2000 candidate region proposals square measure crooked into a square and they are transmitted into a convolutional neural network producing as output a feature vector [17], [21]. Selective search uses the features use to generate possible location of the object. Then the boxes feed into CNN and SVM can predict the class.

C. Fast RCNN

This technique solved some of R-CNN's drawbacks in building a quick object detection algorithm [1].

Fast RCNN is the combination of RCNN and SPP. Spatial Pyramid Pooling (SPPnet)RCNN become as very slow method, because selective search takes more time for 2000 region proposals.

Only one time the CNN representation calculated and used for patch generation using selective search. Here features map occur based on pooling type operation. SPP generate a fixed size layers [3]. Fast RCNN is end to end training. From the

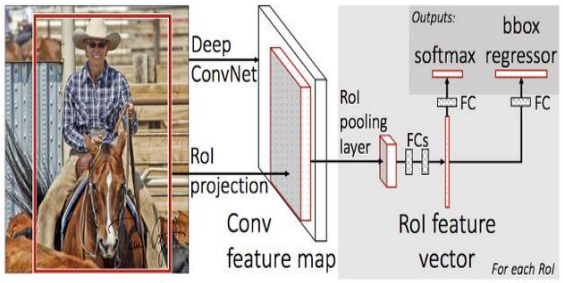

Fig. 3. Fast RCNN [1]

convolutional feature map, it identifies the region of object proposals and warps them into squares, and by using a RoI pooling and it has a fixed sixe then feed into a FC layer [7], [11]. Softmax layer is used by ROI feature vector to predict the region's class and also the bounding box offset values [7], [21]. All the time, 2000 regional suggestions do not want to fed to the CNN. Instead, for each image, one time only the convolution operation is performed, and it generates a feature map [7].

Through a Fast R-CNN network, (fig.3) to identify the object proposals of a image. Initially the image passes through convnet and fed into pooling layer then get the feature map from the input image. Then, all object proposal RoI layer makes a fixed-length feature vector from the feature map [1]. A pooling layer of region of concern (RoI) then removes a fixed length feature vector from the feature map. Each feature vector is fed into a series of fc layers that made branching into two Sibling output layers [10]. One generates softmax estimates of probability over K object classes with background class and next one generates four real value figures for each object class. Each set of four values encodes refined bounding-box positions for one among the K class [1]. While it trains itself, it added bounding box to neural network. Here fig 4 shows the test time of Fast RCNN takes less time compare with other methods. It shows that Fast RCNN is better than RCNN and SPPnet [13] in the case of time taken for test time and training time.

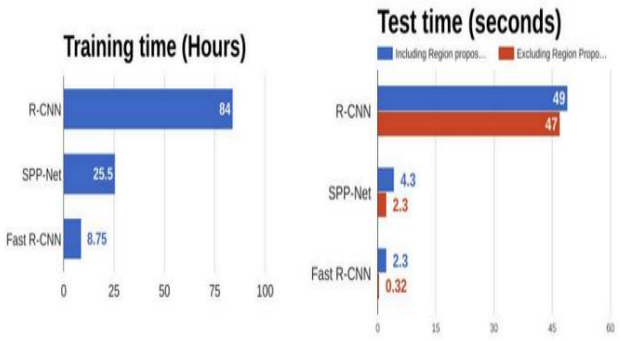

Fig. 4. Comparison of object detection algorithms [21].

D. Faster RCNN

Both R-CNN & Fast R-CNN use selective search algorithms to determine region of proposed object. These methods are become slow as well as the time-consuming process affecting the network performance. Faster RCNN proposed the new method as RPN (Region Proposal Network) used for network to extract the candidate box and generate region of interest. To compared with the RPN the Selective [1] Search algorithm method to extract fewer candidates, Then the Faster RCNN become more efficient. Faster RCNN is the combination of Fast RCNN and RPN [22]. Faster RCNN bring-up the idea of anchor boxes. Here RPN is mainly used for proposing the region as well as the object detection.

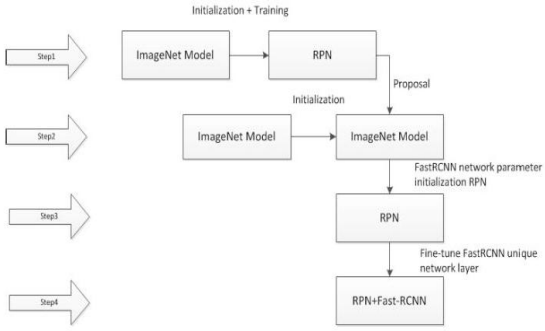

Fig. 5. RPN network and Fast R-CNN training flow chart [22].

Classifier and regression will be examined the RPN output. RPN predict the chances of bounding box of image. Based on RPN dataset are trained first. Then generate a ground truth box from the anchors. Then preparing information is the anchor, get from the over prepare and the ground-truth boxes. The issue got to illuminate here is how use the ground-truth boxes to

name the grapples. The essential thought here is that need to name the stays having the higher covers with ground-truth boxes as frontal area, the ones with lower covers as foundation. Clearly, it needs few changes and compromise to separate closer view and foundation. Here it can check the region of interest here within the execution and labels for the anchors. After RPN, ROI pooling is used here to get the fixed number

feature map from input.

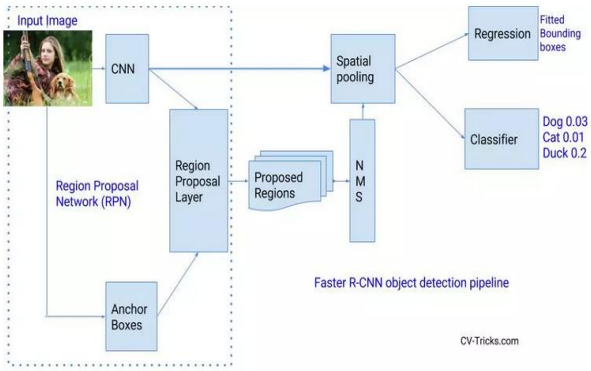

Fig. 6. Faster RCNN architecture [21].

Faster RCNN is faster than Fast RCNN and RPN. RPN predicts the probability of it being background or foreground. Which has 3 parts, in first part,

the image is pass through CNN and here extract the appropriate features from image by trained filters. The second part shows RPN can predict where the object is occupied then predict the bounding box. The third part exhibit the region a class of the image.

E. YOLO

YOLO is very quick technology for object detection. Here do not need a complicated pipeline because the regression problem of the frame detection [2]. Here the neural network of a new picture run at test time to foresee revelations. The base of systems runs at forty-five outlines per moment inside the execution on a dataset as Titan X GPU, and a fast form runs at over one hundred and fifty fps on the dataset. This implies that in periods with but twenty-five milliseconds of idleness we'll stream video. In expansion, YOLO pick up more exactness than other real-time frameworks [1]. YOLO causes the picture worldwide when making projections [1]. YOLO is learning generalized object depictions. YOLO performs elevated detection approaches such as DPM and R-CNN by a right margin when trained on natural images and tested on design. Because YOLO is extremely generalizable, when applied to fresh domains or unexpected input [4], it is less probable to break down. YOLO predicts per grid cell various bounding boxes. YOLO split the image into a grid or matrix of M x M and each of the grid foresee S number of bounding boxes. Which shows accuracy and identify the object using anchor box. Classification score of each box is predicted by this method in every training. You can combine both the classes to calculate the probability of each class being present in a predicted box [4].

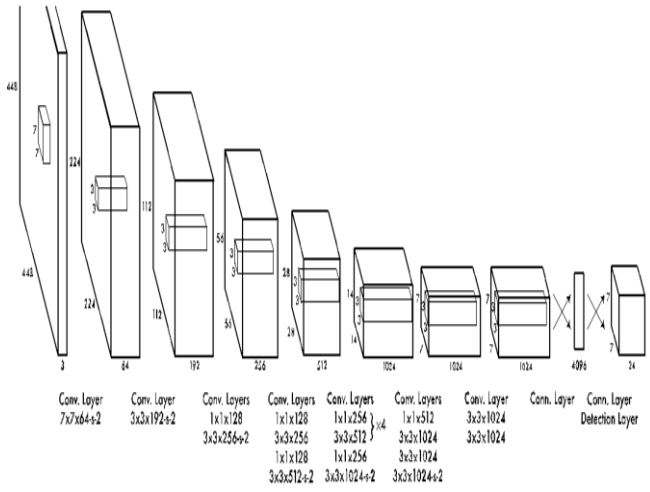

Fig. 7. yolo [4]

F. SSD

SSD is represent as single-shot for the object identification. It is also a pipeline structure. It is faster than the previous state-of-the-art for single-shot detectors (YOLO). Considerably significantly more accurate, in fact as correct as slower techniques
that perform specific region proposals and pooling
(including quicker R-CNN). The core of SSD is redacting class scores and box offsets for a hard and fast set of default bounding boxes victimisation tiny convolutional filters applied to feature maps [8]. PASCAL VOC, COCO, and ILSVRC are the training datasets, which ensure that SSD has competitive accuracy methods. Here it utilizes an additional step for object identification and is much faster, while providing a unified framework for both training and inference.

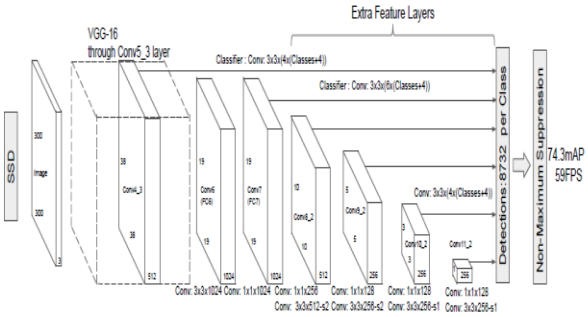

Fig. 8. SSD [8]

In case of SSD only needs an input image and ground truth boxes for each object during training. The SSD technique is based on a convolutional feed-forward network that produces a collection of bounding boxes and scores for the presence of object class cases in those boxes, followed by a non-maximum extraction stage to produce the final detections. Convolutional
predictors for detection every else feature layer (or optionally. AN existing feature layer from the bottom network) will turn out a hard and fast set of detection predictions employing a series of convolutional filters. These are shown on top of the SSD network architecture (figure 8) for a feature layer of size m x n with p channels, the basic element for predicting potential detection parameters is a 3 x 3 x p small kernel
that either produces a score for a category or an offset shape relative to the default box coordinates [8]. It generates an output value at each of the m x n places where the kernel is applied. The yield values of the bounding box offset are evaluated against a default.

III COMPARISON TABLE

Methods	Ref	Pros	Cons
RCNN	[21]	Increase the precision of the bounding box by predicting the presence of an object within the region proposals .	Difficult to selecting a huge number of regions. It takes 47 seconds for each test image. It is very slow.
Fast RCNN	[1], [21]	Object presence is predicted based on selective search algorithm. Faster than RCNN	Bottlenecks in it affecting its performance.
Faster RCNN	[5], [15], [21]	The feature map to identify the region proposals. And good accuracy. Combination of Fast RCNN and RPN	For predicting the region proposals use separate network
YOLO	[4]	Within a single CNN predicts the bounding boxes. And differ from region proposed algorithm.	Difficult to identify small objects within the image.
SSD	[8]	Good balance between speed and accuracy. And predicts bounding boxes after multiple convolutional layers.	Accuracy is less.

IV CONCLUSION

Object detection is considered as important in identifying the vehicle from videos. For object detection, there different methods are using based on deep learning. This approach provides three different classifications like based on RPN, SSD, YOLO.

In this paper shows the role of deep learning, the various methods for object detection and show the learning capacity of methods in deep learning. Different methods provide various accuracy for the images. Based on deep learning RCNN, Fast RCNN, Faster RCNN, YOLO, SSD are the different methods shows the accuracy of images and here Faster RCNN is provide more accuracy compared with the whole methods.

REFERENCES

1. Girshick, Ross. "Fast R-CNN." In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448, 2015.
2. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In Advances in neural information processing systems, pp. 1097-1105, 2012.
3. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition." IEEE transactions on pattern analysis and machine intelligence 37, no. 9: 1904-1916, 2015.
4. Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi." You only look once: Unified, real-time object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788, 2016.
5. Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." In Advances in neural information processing systems, pp. 91-99, 2015.
6. Dai, Xuerui. "HybridNet: A fast vehicle detection system for autonomous driving." Signal Processing: Image Communication 70: 79-88, 2019.
7. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556, 2014.
8. Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector." In European conference on computer vision, pp. 21-37. Springer, Cham, 2016.
9. Zhou, Bolei, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba."Object detectors emerge in deep scene cnns." arXiv preprint arXiv:1412.6856, 2014.
10. Erhan, Dumitru, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov."Scalable object detection using deep neural networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2147-2154, 2014.
11. Long J, Shelhamer E, Darrell T. "Fully convolutional networks for semantic segmentation." Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S. 1.]: CVPR Press, 1109-1123, 2015.
12. K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features, in Proc. 10th IEEE Int. Conf. Comput. Vis, pp. 14581465, 2005.
13. Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories." In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 2, pp. 2169-2178. IEEE, 2006.
14. Taigman, Yaniv, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. "Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1701-1708, 2014.
15. Pathak, Ajeet Ram, Manjusha Pandey, and Siddharth Rautaray. "Application of deep learning for object detection." Procedia Comput. Sci 132: 1706-1717, 2018.
16. Arinaldi, Ahmad, Jaka Arya Pradana, and Arlan Arventa Gurusinga. "Detection and classification of vehicles for traffic video analytics."Procedia computer science 144 259-268, 2018
17. Chahyati, Dina, Mohamad Ivan Fanany, and Aniati Murni Arymurthy."Tracking people by detection using CNN features."Procedia Computer Science 124: 167-172,2017
18. Dai, Xuerui. "HybridNet: A fast vehicle detection system for autonomous driving."Signal Processing: Image Communication 70 : 79-88,2019
19. Arinaldi, Ahmad, Jaka Arya Pradana, and Arlan Arventa Gurusinga. "Detection and classification of vehicles for traffic video analytics." Procedia computer science 144: 259-268, 2018
20. Uijlings, Jasper RR, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. "Selective search for object recognition." Internationaljournal of computer vision 104, no. 2: 154-171,2013
21. https://towardsdatascience.com/r-cnn-fast-r-cnn- faster-r-cnn-yolo object- detection-algorithms-36d53571365e
22. Suhao, Li, Lin Jinzhao, Li Guoquan, Bai Tong, Wang Huiqian, and Pang Yu. "Vehicle type detection based on deep learning in traffic scene." Procedia computer science 131 (2018): 564-572, 2018

AUTHORS PROFILE

Arlin Maria Scaria has been pursuing M-Tech in Computer Science and engineering from Vimal jyothi Engineering College, Chemberi since 2018 and she has completed B-Tech in Information Technology from K.M.C.T. College of Engineering for Women.

Neena V.V, Received M.Tech degree in Computer Science and Engineering from Anna University, Chennai and B.Tech degree in Computer Science and Engineering from Kannur University, Kerala. Now working as an Associate Professor in Computer Science and Engineering Department, Vimal Jyothi Engineering College, Chemperi, Kannur, Kerala. Currently pursuing PhD from Amrita University, Coimbatore, Tamilnadu. My current research interests include Machine Learning, Edge Computing, Edge Analytics, Computer Architecture.