ISSN 2319 – 1953
International Journal of Scientific Research in Computer Science Applications and Management Studies

An Efficient Semantic Segmentation using Neural Network based E-Net Architecture

Dr. G. Rama Subba Reddy
Associate Professor, Dept. of CSE, Mother Theresa Institute of Engineering and Technology, Palamaner, India
subbareddy1227@gmail.com

***Abstract-* The demand on home automated gadgets, augmented wearables and self-driving vehicles are high in recent years. These technologies use the concept of semantic segmentation. This methodology includes by taking each and every pixel, that is each pixel might belongs to any group and also be monitored in the real-time applications. Even though there is greater data sets availability and various algorithms of Machine Learning (ML) outlines the performances of such applications. For classifying the images in space along with various considerable segmented regions there exists many Neural Networks. It includes with SegNet or Complete CNN to do multiple class classification. In this paper work, we used recently proposed ENet model (Efficient Neural Network) which is specifically modelled for the tasks which requires less latency operations. This model is 18x times faster, used least number of flops i.e. 75x less and utilizes minimum number of parameter i.e.79X less. The recommended method practically utilizes Cityscapes database and the outcomes are compared with other traditional techniques. We also provided the performance measurements by using ENet on the embedded systems, which is mainly required to improve software's that could make ENet very speedy. We have not used any post processing, as it reduces the CNN's performance, but however you may include as a step to attain many accurate outcomes.**

***Keywords* – E-Net architecture, semantic segmentation, automated driving, remote sensing, Machine Learning.**

I. Introduction

In the fields of image processing and computer vision, the semantic segmentation is treated as one of the important applications. It is usually applied in various fields such as in medical and intelligence transportation. Particularly various data sets are provided to researchers to investigate their algorithms. Semantic segmentation was studying from many years. With the availability of Deep Neural Networks (DNN), the semantic segmentation results a best progress. On the basis of computer vision's background, Semantic segmentation as a major challenge performs the process of segmentation on any kind of image. During segmentation process it divides an image into different regions including objects such as sceneries, dog and human as well [1]. Along this segmentation is somehow depth while comparing with detection of objects because there is no need to use detection in segmentation. Specifically, mankind does image segmentation without object detection. This is the major part to study the process in visual manner which produces a robust method and to use that method for improving the available techniques in computer vision [2].

At present, the field of computer vision is facing the issue regarding semantic segmentation. In deeper situations, it is considered as high-level task which deals the way to analyse the overall situation. It is intended to study the situation concerning the problem existing in computer vision is highlighted by the fact that more the applications results in obtain awareness over imagery. Some applications include vehicles with self-driving, communication between a human and system etc. With the emergence of the concept of deep learning many issues regarding semantic segmentation are solved by applying deep architectures, often Convolutional Neural networks, extends the rest of the models with better rate of accuracy and efficiency [3].Such algorithm includes with detection of roads markings, finding tumours, and detecting medical instruments in surgery, colon based cryptographic segmentation. Various segmentation applications related to medical field can be stored. In contrast, non-semantic segmentation will group the pixels of image together based on common characteristics of single objects. Hence it is defined in improper way like the others.With this it results in two issues and those are i) neighbouring pixels of same class might belonging to different object's instances and ii) the regions are not belonging to the similar object instance [4]. Semantic segmentation is very powerful when compared to conventional segmentation. The primary purpose of semantic segmentation is self-driving cars which are included in this study. The rest of the applications involves Geo Sensing, Autonomous driving, Face segmentation, Fashion i.e. parsing cloths, Precision farming. The aim of this paper is finding, studying on what that image has at pixel level and is shown in Figure 1 (left: input image, right: segmented image). The previous algorithm only performs either binary or multi segmentation but do not assign the labels of that classes of multiple kinds. For this image or a video stream is given as an input and as an output there given an image which has different regions with various classes [5].

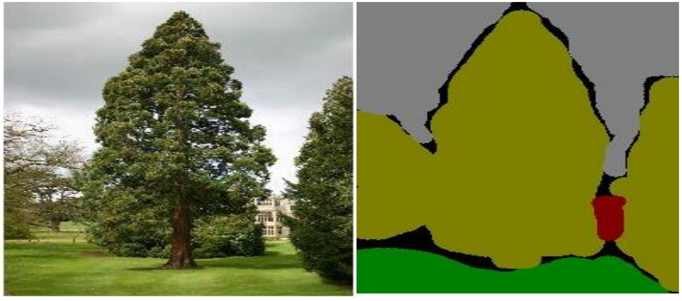


Figure 1: Semantic Segmentation

II. Recent Works

The Cityscapes Dataset [6] is like a benchmark that concentrates on semantic analysing the images of urban streets. It includes 30 classes in and over 5000 images that are fine annotated which are gathered from the 50 cities. In addition to this, the observed time period is took many months. Figure. 2 represents a fine-annotated image.

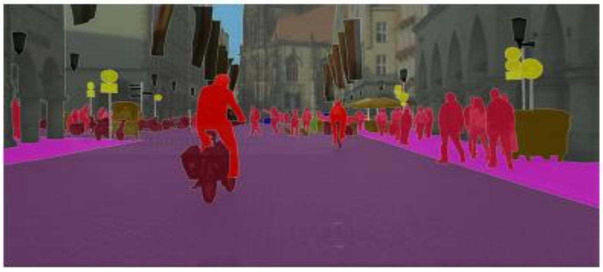


Figure2: A fine annotated image from Cityscapes

The deconvolution network utilized in [7] comprises of deconvolution and also un-pooling layers, which detects labels of class, based on pixel wise and predicts the segmentation masks. In contrast FCN in the paper [8], it is applied on every object proposals in order to get the instance-wise segmentations integrated for the last semantic segmentation. In Paper [8] there was already applied the same strategy which is so called as Atrous Convolution' or 'Hole Convolution' or 'dilated convolution'.In [9], there presented a single module that uses dilated convolutions for valuating multi-scale contextual data in proper way. The model relies on dilated convolutions which support expansion of exponential receptive field with no loss in resolution or also in coverage. Because dilated convolution contains artefacts, paper [10] develops a model named as dilated residual networks (DRN) to ignore those artefacts and then enhances the network's performance. In the domain of semantic image segmentation paper [11] segments a network with the conventional multi-scale image input and sliding pyramid pooling which can increases the performance effectively. This architecture holds the context of patch-background.Likewise, Deep lab implements an image pyramid structure [12] that extracts some multi-scale features by feeding the various resized input images in heterogeneous deep network. At the end of every deep network, the resultant features are combined to do classification on pixel-wise. CNN is considered as feature extractor [13]. Nevertheless, the data in this layer is not suitable for high prediction. In contrast, the previous layers may be exact in localization, but unable to capture semantics.

1. *Semantic Segmentation*

Semantic segmentation models are usually modelled and presented in this paper. Besides we also presented complete state-of-art-methods. Hence it allows us to deploy the other models. Since almost all these have same underlying requirements, settings and flow. Followed the same model of feature extraction along with the multi-scale processing. Hence it is easy to implement and train from end-to-end. Your option of using depends on your specification regarding accuracy or speed or memory.The structure of the M-CNN network is shown in Figure 3 and semantic segmentation from natural language expression in Figure 4.Figure 5 shows the representative segmentation pipeline of traditional architecture.

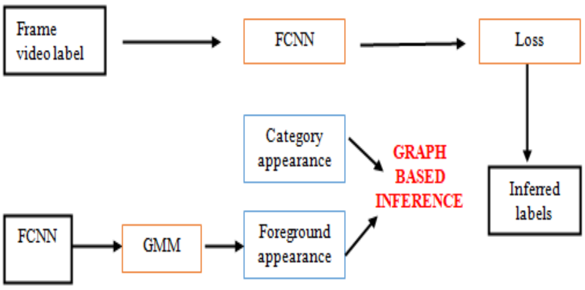


Figure 3: Overview of M-CNN framework.

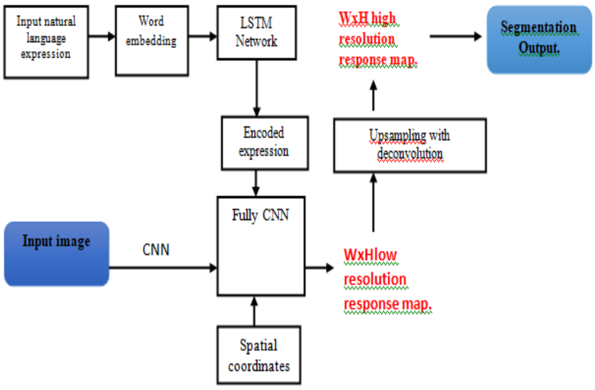

Input natural language expression
Word embedding
LSTM Network
WxH high resolution response map.
Segmentation Output.
Encoded expression
Upsampling with deconvolution
Input image
CNN
Fully CNN
WxHlow resolution response map.
Spatial coordinates

Figure 4: Segmentation from natural language expression.

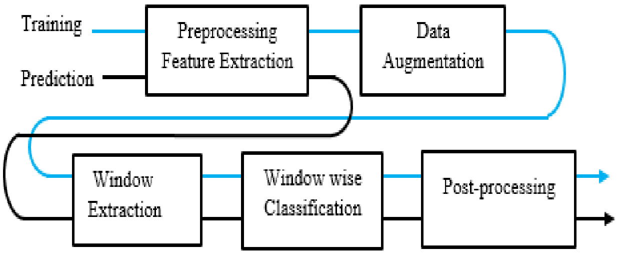


Figure 5: Basic Structure of Semantic Segmentation

1.1 Algorithms

Clustering Algorithms - These algorithms can be applied directly on the pixels. There are two clustering algorithms; one is k-means and other is mean-shift algorithm. K-means is general-purpose clustering algorithm that requires clusters group which are provided earlier. Initially it randomly positions k-centroids in a feature space. Mean-shift algorithm it assigns each data pixel to its neighbouring centroid and shifts that centroid towards cluster's center and continues the process until the required criteria get satisfied [14]. *Graph Based Image Segmentation* – These algorithms interpret some pixels such as vertices and weights of the edge is considered as a measure of colour deviation [15]. *Random Walks* - This relates to algorithms of graph-based image segmentation. This kind of image segmentation usually works as: seed pixels are located in image for various objects in that. *Watershed Segmentation* - This takes an image with a Grayscale and also interprets the same image as a height map. Minimum values are stated on catchment basis and a maximum value which exists between the both catchment basins is said as a watershed. It represents that those regions must keep in dark on images with Grayscale. It starts to finish the total basins from lower point. A watershed can be detected when there is meeting of dual basins is done. This paradigm exits when it reaches highest point.Initially these are presented in [16]. This classifier type applies some techniques called ensemble learning where there will be training of many classifiers done together with the combination of its integration. Another technique of the ensemble learning is bagging. The classifiers are considered as a decision tree in the context of Random Decision Forests. SVMs- There are known for well-examined classifiers which are to be explained with five central ideas. For such thoughts, the training data is represented as (P_i, Q_i) where P_i is a feature vector and Qi∈ {−1, 1} the binary label for training example i∈ {1... n}.

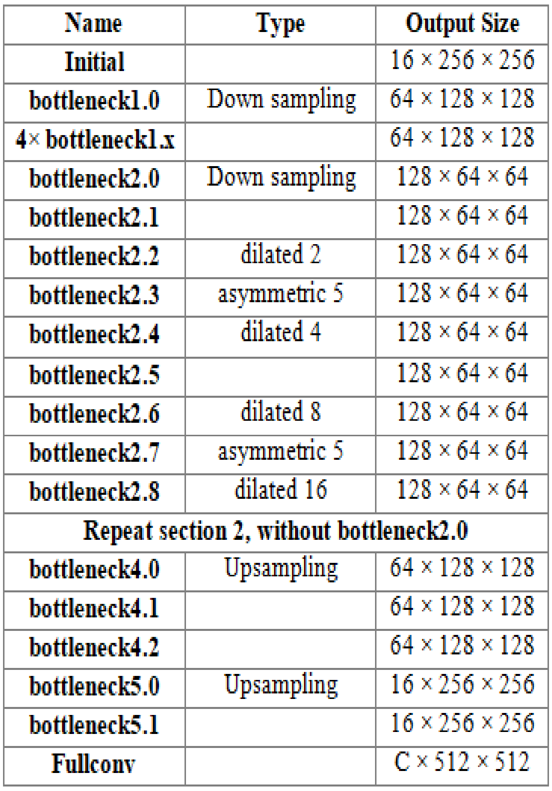
3.2 Neural Networks

Artificial Neural Networks (ANN) are the classifiers derived from biologic neurons. Each single neuron (artificial) contains inputs that are weighted and summed up too. Next neuron try to apply the activation function for that weighted sum and provides output. Various neural networks regarding ideas over regularization, best algorithms of reduction and existed models and so on. We are not providing the summary of its description here. But we highlighted few of the breakthroughs. Alex Krizhevsky et al. in there may be high optimization towards parameters that must be learned when there is issue in imagery. A major thought is that clever regularization called as a dropout training. It keeps the outputs of neurons in random way to zero while training.

IV. E-NET ARCHITECTURE

The network architecture of E-Net is shown in Figure 6. It contains multiple stages, and are highlighted by horizontal lines and the starting digit after every block. The image size of output image is of 512 X 512. This architecture adopts ResNet[17] describes a single main branch and extensions with convolutional filters which separate from it and then merge back with element wise addition as given in Figure 7.b.

In our paper, we used deep learning ENet architectures to perform semantic segmentation. It helps in applying on both the images and videos as well. The key benefit with ENet is that it is 18 times speedy, need only some parameters and provide better accuracy when compared with the rest of the models. The model size is only 4mb. In the case of execution period single pass consumes 0.2 seconds in CPU.



Name	Type	Output Size
Initial		16 × 256 × 256
bottleneck1.0	Down sampling	64 × 128 × 128
4× bottleneck1.x		64 × 128 × 128
bottleneck2.0	Down sampling	128 × 64 × 64
bottleneck2.1		128 × 64 × 64
bottleneck2.2	dilated 2	128 × 64 × 64
bottleneck2.3	asymmetric 5	128 × 64 × 64
bottleneck2.4	dilated 4	128 × 64 × 64
bottleneck2.5		128 × 64 × 64
bottleneck2.6	dilated 8	128 × 64 × 64
bottleneck2.7	asymmetric 5	128 × 64 × 64
bottleneck2.8	dilated 16	128 × 64 × 64
Repeat section 2, without bottleneck2.0		
bottleneck4.0	Upsampling	64 × 128 × 128
bottleneck4.1		64 × 128 × 128
bottleneck4.2		64 × 128 × 128
bottleneck5.0	Upsampling	16 × 256 × 256
bottleneck5.1		16 × 256 × 256
Fullconv		C × 512 × 512

Figure 6: E-Net Architecture (Image size 512 X 512)

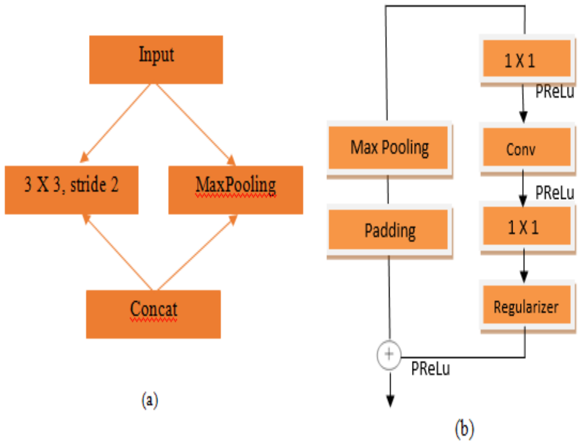
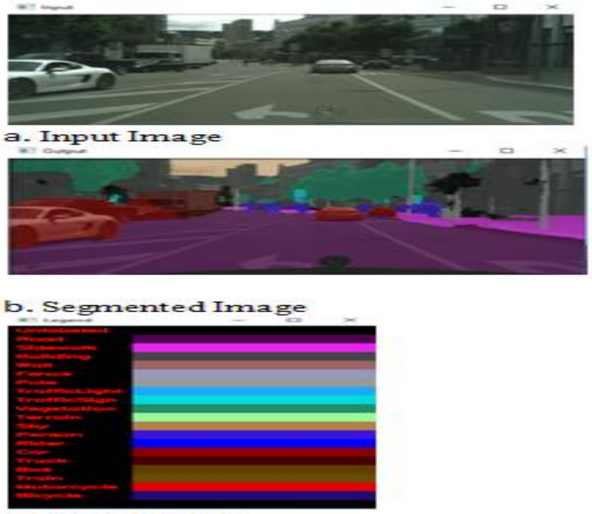


Figure 7: (a) Initial Module (b) Bottle-neck module

Other models which performs similar task are Alex Net, VGG-16, Google Net and ResNet. This model is trained based on the instances of classes like Road, Side-walk, Buildings, and Vehicles and so on. We used the packages such as numpy, argparse, imutils, time and OpenCV in our work. ENet architecture utilizes very fewer parameters, consumes less space around 1 MB of memory. The primary phases of semantic segmentation involve classification, localization or detection and eventually semantic segmentation. This method can be applied on both images and videos.



a. Input Image

b. Segmented Image

c. Legend colors

Figure 8: Semantic Segmented Result

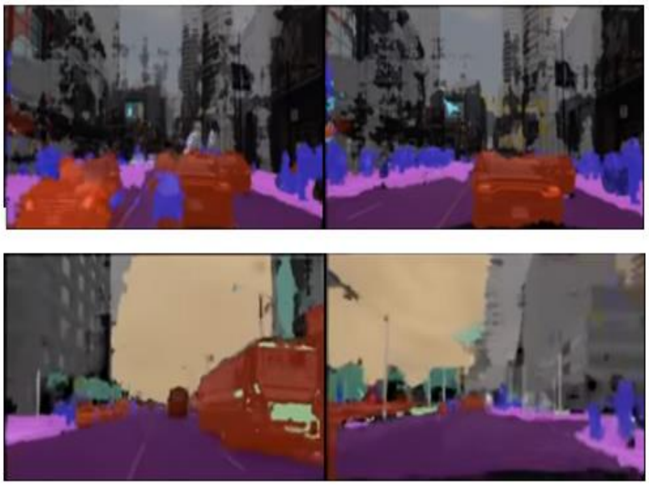


Figure 9: Semantic segmented Frames from Self driving car

In Figure 8, we have taken a sample image and the result of semantic segmentation of this work is shown in Figure 8.b. Similar to this, we can get the legendary colours (up to 20) for the image as shown in Figure 8.c. This work can also applicable to videos, the sample frames shown in Figure 9.

V. Conclusion

Semantic segmentation is very important in the analysis of image content and also object detection in the same image. The Difference between segmentation and semantic segmentation is that, traditional algorithms are used for segmenting the input images into different regions without knowing anything about. For instance, usual graph cuts, super pixels and so on. Many other applications includes Geo Sensing, Autonomous driving, Face segmentation, Fashion i.e. parsing cloths, Precision farming. In this paper, traditional works of semantic segmentation are discussed. For this, to earlier work there used neural networks to object detection in real times. Besides we also concentrated on feature detection that can be used for pipeline segmentation to obtain the outcomes of segmentation. Later, we provided the work related to deep learning by using E-net architecture to perform segmentation. Additionally we included the outcomes of E-net architecture for the automatic driving vehicle. Enhancing the efficiency and performance of those paradigms are considered as our further work in future.

References

[1]. Cohen A, Rivlin E, Shimshoni I, Sabo E, "Memory based active contour algorithm using pixel-level classified images for colon crypt segmentation". Comput Med Imaging Graph 43:150–164, 2015

[2]. Wu Z, Shen C, Hengel A High-performance semantic segmentation using very deep fully convolutional networks, 2016a.

[3] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu. Semantic segmentation with second-order pooling. In ECCV, 2012.

[4] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. DeCAF: A deep convolutional activation feature for generic visual recognition. In ICML, 2014.

[5] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In Computer Vision–ECCV 2014, pages 818–833. Springer, 2014.

[6]Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke , Roth S, Schiele B (2016) The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3213–3223

[7] Noh H, Hong S, Han B (2015) Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE international conference on computer vision, pp 1520–1528

[8] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2016b) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv preprint arXiv:1606.00915

[9] Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122

[10]. Yu F, Koltun V, Funkhouser T (2017) Dilated residual networks. arXiv preprint arXiv:1705.09914

[11]. Lin G, Shen C, van den Hengel A, Reid I (2016a) Efficient piecewise training of deep structured models for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3194–3203

[12] ChenLC,YangY,WangJ, XuW,YuilleAL(2016c) Attention to scale:scale aware semantic image segmentation.In:Proceedings of the IEEE conference on computer vision and pattern recognition,pp3640–3649

[13]Hariharan B, Arbeláez P, Girshick R, Malik J (2015) Hypercolumns for object segmentation and fine-grained localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 447–456.

[14]. J. A. Hartigan, Clustering algorithms. John Wiley & Sons, Inc., 1975.

[15]. P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," International Journal of Computer Vision, vol. 59, no. 2, pp. 167–181, 2004. [Online]. Available: http://link.springer.com/article/10.1023/ B: VISI. 0000022288.19776.77

[16]. T. K. Ho, "Random decision forests," in Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on, vol. 1. IEEE, 1995, pp. 278–282. [Online]. Available: http://ect.bell-labs.com/who/ tkh/publications/papers/odt.pdf.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv preprint arXiv: 1512.03385, 2015.

[18] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 648–656.

[19] S. Chetlur, C. Woolley, P. Vandermersch, J. Cohen, J. Tran, B. Catanzaro, and E. Shelhamer, "cudnn: Efficient primitives for deep learning," arXiv preprint arXiv:1410.0759, 2014.