iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44511

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

A Comparative Study on Semantic Segmentation Algorithms for Autonomous Driving Vehicles

Shivam Chhabra[1], Rahul Rohilla[2]
[1, 2]*Department of Computer Science and Engineering, Dronacharya College of Engineering, Gurgaon Haryana, India*

Abstract: ***Semantic segmentation refers to the process of classifying each pixel in an image for better understanding and analysis of image. This is how machines look at the real world and identifies different objects. These are the times when autonomous vehicle industry is blooming and establishing new heights. There are so many research studies going on around semantic segmentation that are advancing to break boundaries in the world of computer vision. Despite of so much of progress made in the field in the recent years, autonomous vehicles needs more improved and efficient models to ride on the roads. In this research paper we compare the currently proved popular choices of models for semantic segmentation with respect to autonomous vehicles on different parameters to create an in-depth analysis on which all models and their variations improve and affect the quality of real time segmentation of the real world. The popular model architecture choices that were assessed and compared in this research were Fully Convolutional Network (FCN), U- Net and DeepLab. The data used for analysis was taken from Lyft Perception Challenge on Lyft Perception Challenge on Udacity.***

I. INTRODUCTION

During Image Segmentation an image is partitioned into different segments. The main objective of this process is to simplify or transform the image into something that is easy to understand and analyze. In machine vision it also helps machines to identify and understand the relationship between two similar objects located at two different points in an image. The output of this process is an image that has different partitions for each object located in that image. These partitions are differentiated by variation in texture, color, or intensity. Over the past years, image segmentation has found application in many different fields of real world such as in medical imagery to locate tumors, measure tissue volumes, and virtual surgery simulation [1]. Similarly, an autonomous vehicle also uses semantic segmentation and instance segmentation to accomplish different objectives such as pedestrian detection, vehicle detection and road lane detection [2]. Deep learning has revolutionized semantic segmentation and led to better results for more effective and efficient applications. Before Deep learning Networks, machine learning algorithms faced a major challenge of extracting features from an image, but the continuously growing field of deep learning never stumbled upon having a separate step to extract features. Deep learning state of the art algorithms are capable of achieving semantic segmentation faster than the classical approaches and with higher accuracy [7].

II. LITERATURE SURVEY

The first prominent work in this field was fully convolutional networks(FCNs) [1]. FCN was proposed an end-to-end method to learn pixel-wise classification, where transposed convolution was used for upsampling. Skip architecture was used to refine the segmentation output, that utilized higher resolution feature maps. That method paved the road to subsequent advances in the segmentation accuracy. Multi scale approaches [2][3], structured models [4][5], and spatio-temporal architectures [6] introduced different directions for improving accuracy. All of the above approaches focused on accuracy and robustness of segmentation. Well known benchmarks and datasets for semantic segmentation such as Pascal [7], NYU RGBD [8], Cityscapes [9], and Mapillary [10] boosted the competition toward improving accuracy. However, little attention is given to the computational efficiency of these networks. Although, when it comes to applications such as autonomous driving this would have tremendous impact. There exists few work that tries to address the segmentation networks efficiency such as [11][12]. The survey on semantic segmentation [13] presented a comparative study between different segmentation architectures including ENet [12]. An autonomous vehicle uses semantic segmentation and instance segmentation to accomplish different objectives such as pedestrian detection, vehicle detection and road lane detection [13]. An autonomous vehicle in order to deliver a safer drive requires lots of intelligence in a fraction of second. The speed of approaching object, turns, road signs, pedestrians, footpaths and much more needs to detected and assessed instantly for a real world experience.[14][15]. Yet, there is no principled comparison of different networks and meta-architectures. These previous studies compared different networks as a whole, without comparing the effect of different modules. That does not enable researchers and practitioners to pick the best suited design choices for the required task.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

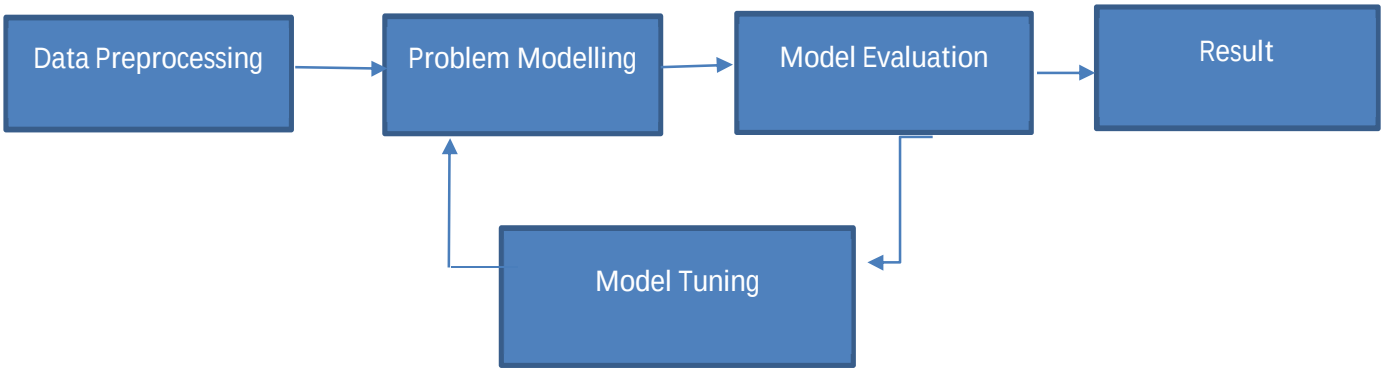
III. PROPOSED METHOD

In this research we trained, tested, and compared three popular algorithms that are used widely for the task of image segmentation for autonomous vehicles, namely FCN, U- Net, DeepLabV3. While this is an accuracy sensitive task, we need higher accuracy in less time for real- time application. As sematic segmentation is classification of pixels task, we used Categorial Cross Entropy as the primary metric.

$$\text{Loss} = -\sum_{i=1}^{\text{output size}} y_i \cdot \log \hat{y}_i$$

formula for Cross- Entropy Loss
[16]

Here $y_i\hat{}$ is the predicted value, and y_i is the target value
Layout for our research is:



A. Dataset Collection

The dataset is collected from Kaggle.com. Kaggle is a repository for datasets for tasks in machine learning, deep learning, data analytics and visualizations. The dataset is called Semantic Segmentation with CARLA. CARLA is a self- driving car simulator. The dataset was released as a part of Lyft Perception Challenge on Udacity. There are 1000 images. It consists of RGB images and their corresponding semantic segments. Semantic Segments are of road, trees, vegetation, vehicles etc. These RGB images are taken from various trips in 'town 1' of CARLA self- driving car simulator. The test dataset is taken from cvlibs.net. it is Kitti semantic segmentation dataset of 208 images. We used these images to evaluate our models.

B. Data Preprocessing

Data preprocessing not only reduces time complexity and space complexity of deep learning algorithm but also has proven to give better results. In order to achieve this at first, we have rescaled our all-training sample images to size of 256 x 256 along with their semantic segmented images. This step was followed by normalizing the images. By normalizing pixel values, we are setting our input values between 0 and 1. Normalizing data helps in speeding up the learning process and get to the convergence point faster.

C. Problem Modelling

For training we have trained three models Full Convolutional Network, U- Net and DeepLab. All of the three models are trained using transfer learning method. Under transfer learning we have fine tuned weights of the backbones of these architectures namely Vgg16 feature extractor for FCN and ResNet 50 Architecture for U- Net and DeepLab. The weights of these backbone feature extractors are frozen, and the training and tuning is done of the weights of the forward layer. Transfer learning is used to speed up the training process and make use of the learning that was done on greater computing machines and large amount of data. All these models are trained on Google Collab GPU. Also Early Stop is applied while training with respect to validation loss metric and patience parameter set to 40.

D. FCN (Vgg16 backbone)

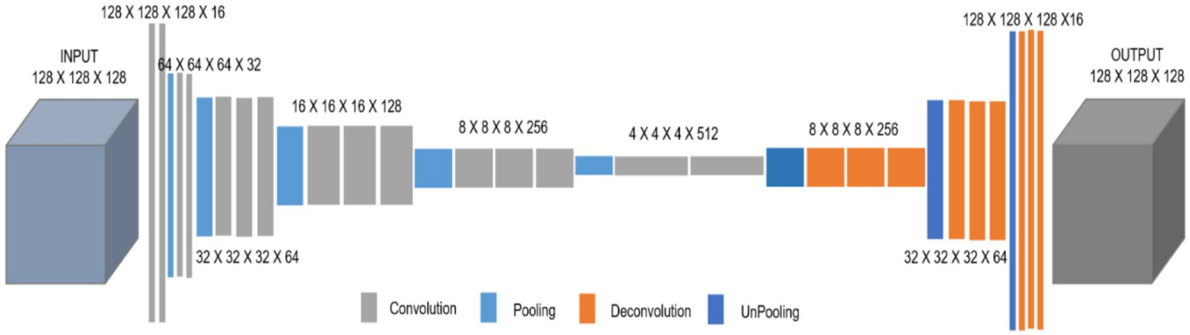


Figure - Architecture of FCN

Vgg16 proves its feature extraction power on imagenet dataset by replacing AlexNet large sized feature kernels with multiple 3 x 3 kernels [17]. This is then used in the Fully Convolution Network. In a FCN multiple convolutional layers are stacked upon as in an upscaling and downscaling fashion. The architecture downscales the image and then upscales. There are skip connections in the architecture that are utilized during upsampling to build proper segments of those features that were lost during downsampling.[18]

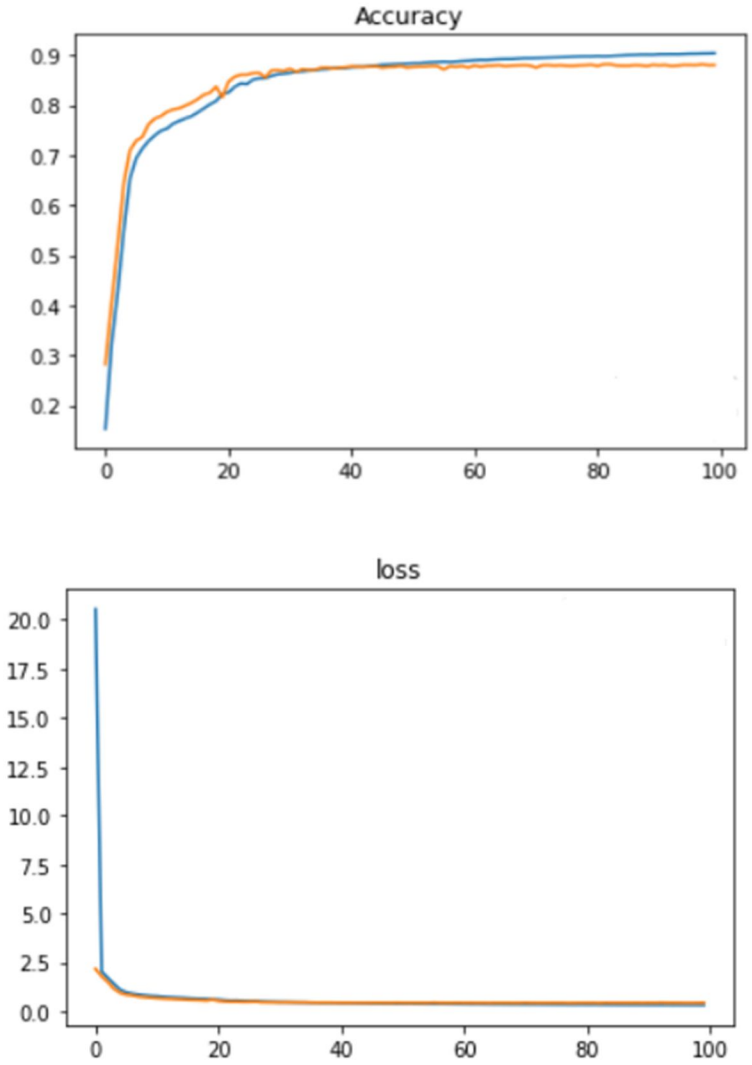
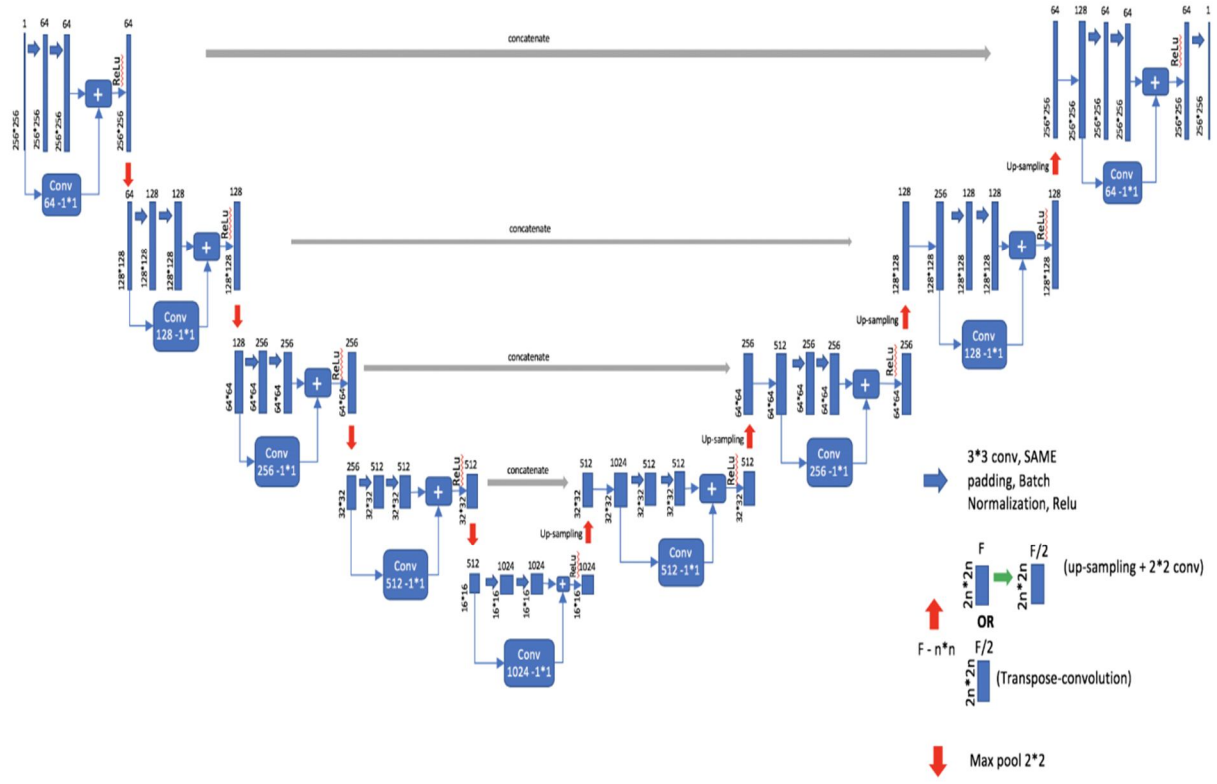


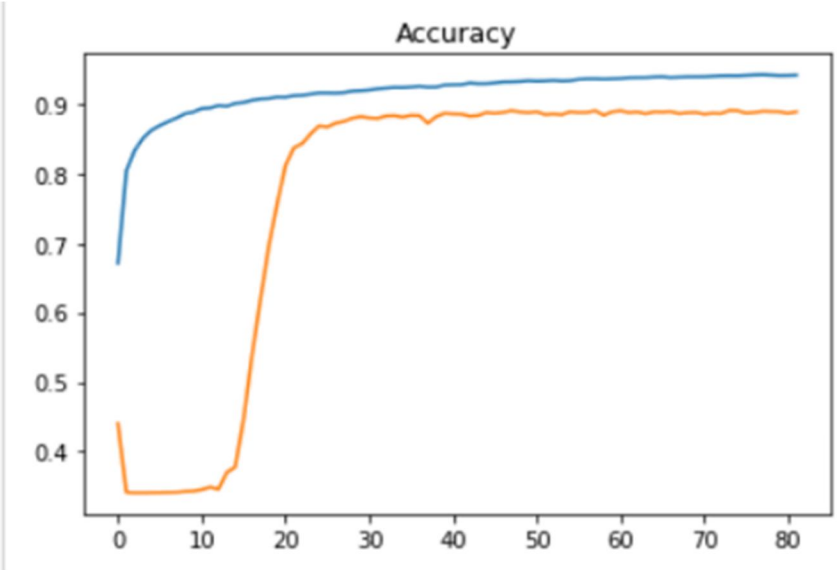
Figure Learning Curves where orange line plot shows validation results while blue line plot shows results on training set

E. U- Net (Resnet 50 backbone)



Resnet50 architecture won Imagenet competition in 2015. Residual Networks became the new way of training deeper networks and achieve better accuracy and performance of your deep neural network.[19]

U-Net's shape is similar to letter U, that's why the name. There is no fully connected layer, and each layer is complimented by a ReLu activation function non linear layer. This model rose the idea of increasing the output resolution of the segmented image for better results.[20].



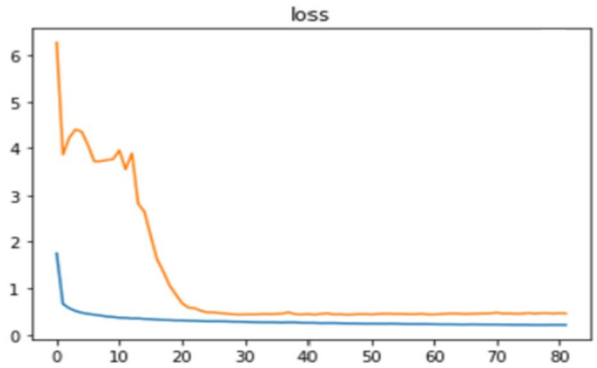
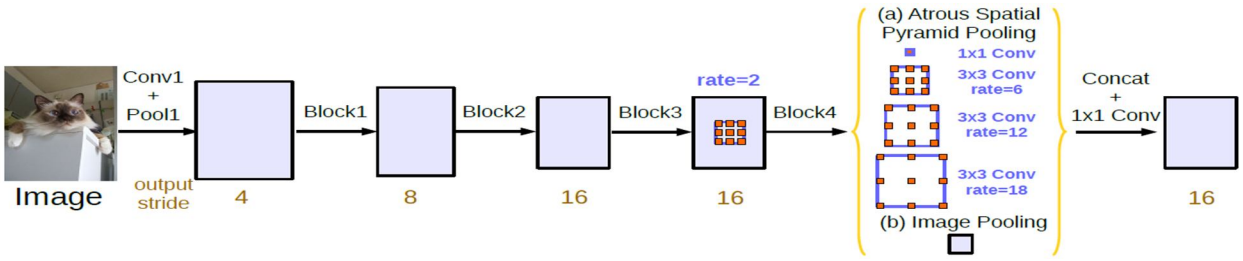


Figure Learning Curves where orange line plot shows validation results while blue line plot shows results on training set

F. Deeplab V3 (Resnet 50 backbone)



Residual Nets solved the problem of vanishing gradients using skip connections. This is done by allowing the gradients to flow through high level layers.[19]. While DeepLabV3 was the next improvement in DeeplabV2 architecture in order to solve the problem of resolution like in U-Net did, DeepLabV3 brought the idea of atrous spatial pyramid pooling layer, that would upsample the output image to get better resolution of image. Also the CNN is not restricted to accept same sized image.[21]

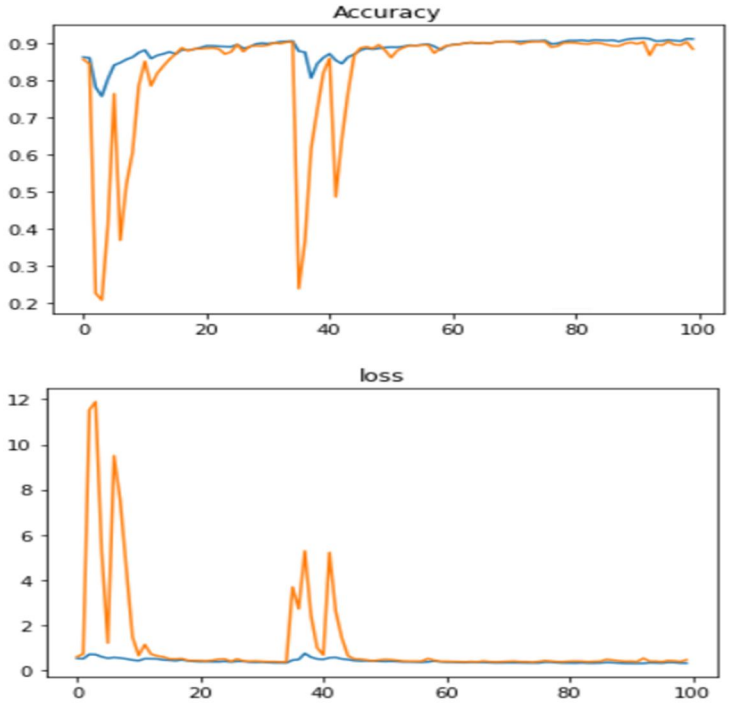


Figure Learning Curves where orange line plot shows validation results while blue line plot shows results on training set

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

IV. OBSERVATIONS

As per the training analysis that is mentioned in table 1, the training done on Google Collab's GPU turns out to reduce training time effectively. While in terms of metrics U- Net approaches lowest cross entropy loss in 82 epochs, while FCN and DeepLab V3 has 0.32 and 0.31 loss respectively in 100 epochs after applying early stop at validation loss with patience of 40. In terms of validation loss there is not much of a difference however FCN has the lowest validation loss.

Since we are treating image segmentation as a classification task we can observe the test analysis in table 2, which shows that on test set that was run on CPU, U- Net's performance is better than other models. While the accuracy of DeepLab V3 and U- Net is closer, the loss that we were treating as a metric to measure these algorithms is much better for U- Net.

Table 1 Training Analysis while training on GPU

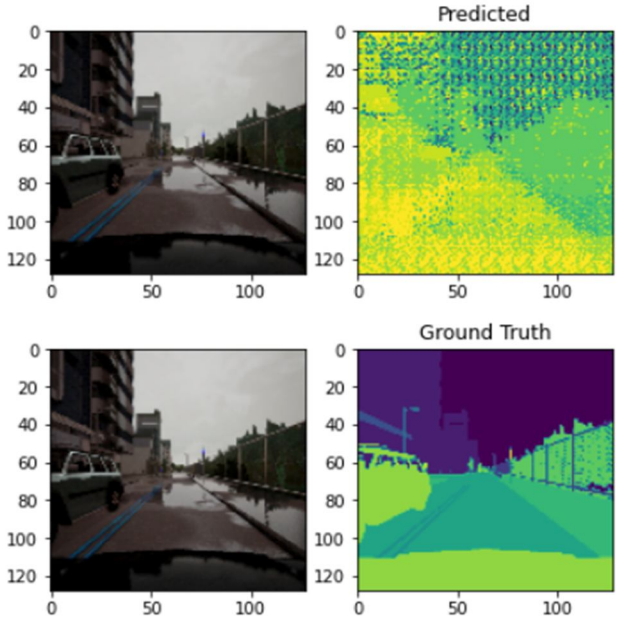
Model	Training Accuracy at final epoch	Training Cross Entropy Loss at final epoch	Training Time	Epochs trained	Validation Accuracy at final epoch	Validation Cross Entropy loss at final epoch
FCN	0.903772	0.325934	7 min 9 sec	100	0.880146	0.444771
U- Net	0.940938	0.208573	20 min 27 sec	82	0.888615	0.459098
DeepLabV3	0.910189	0.318460	36 min 33 sec	100	0.882955	0.463669

Table 2 Testing Analysis on CPU

Model	Cross Entropy Loss	Accuracy
FCN	0.652957506058458	0.7893147506058216
U- Net	0.513983026838307	0.8658383028740287
Deep Lab V3	0.589138570046443	0.8556396443364644

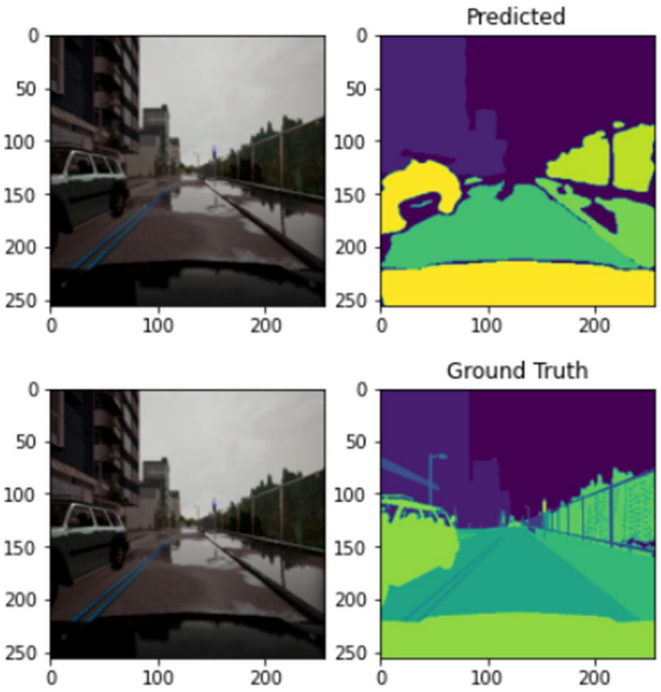
A. Image Prediction Vs Ground Truth

1) FCN



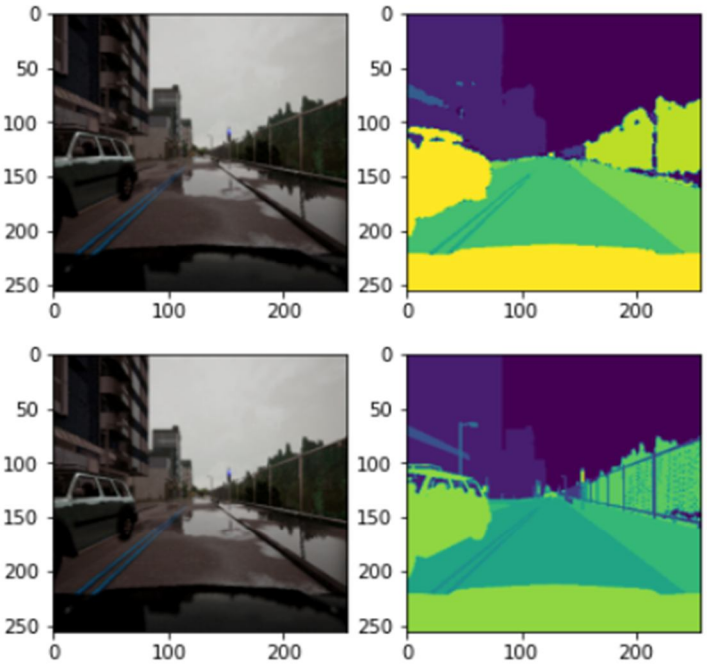
CPU times: user 1.16 s, sys: 19 ms, total: 1.18 s Wall time: 1.18 s

2) Deeplab



CPU times: user 2.88 s, sys: 32.5 ms, total: 2.91 s Wall time: 2.92 s

3) U-NET



CPU times: user 1.89 s, sys: 32.9 ms, total: 1.98 s Wall time: 1.98 s

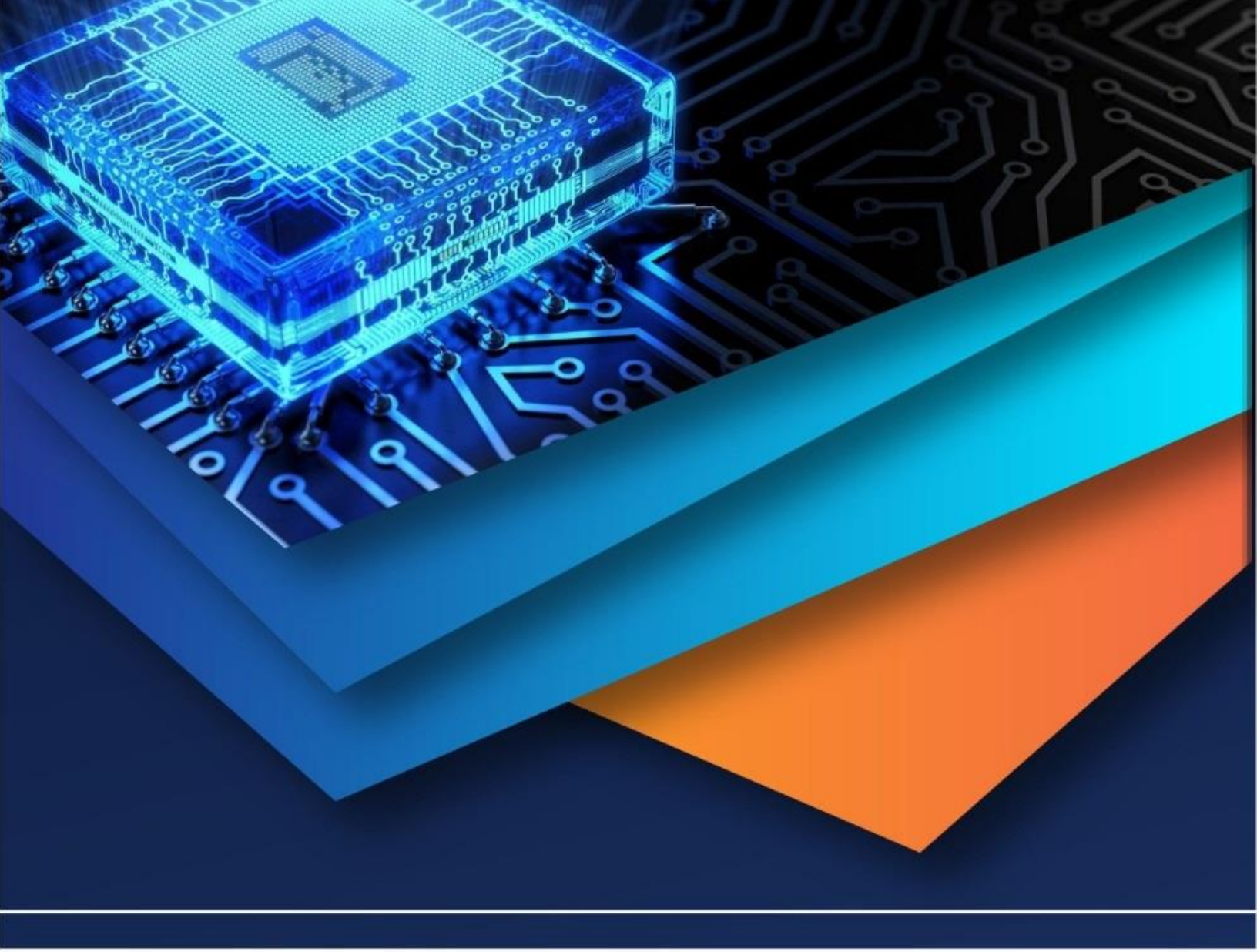
V. CONCLUSION

Autonomous Vehicles are sensitive to accurate performance of model and prediction time as for the real- time system. We used Cross Entropy loss as the primary and Accuracy as the secondary metric to measure the performance of three popular algorithms for the task of image segmentation for Autonomous Vehicles. In the analysis our finding led to the conclusion that the older algorithm FCN is not accurate and fast enough to compete with the other two models. While U- Net and DeepLab V3 have quite competitive results. U- Net's skip connections and encoder- decoder structure do lead to faster training and better results, while DeepLab takes longer time to achieve the same result. Therefore, although having a close score DeepLabV3 is computationally expensive as compared to U- Net. However, there is still a need for higher accuracy to achieve real- time supremacy.

Future work entails to compare the algorithms performance in modular structure not as a whole to find out the best components that can tied together to produce better results. Secondly to combine the high-resolution power of DeepLabV3 with the U- Net's better performing architecture to propose an algorithm that is capable of achieving the goal of semantic segmentation in real- time system.

REFERENCES

[1] Jonathan Long, Evan Shelhamer, and Trevor Darrell,"Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015, pp. 3431–3440.
[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," arXiv preprint arXiv:1606.00915, 2016.
[3] Fisher Yu and Vladlen Koltun, "Multi-scale context aggregation by dilated convolutions," arXiv preprint arXiv:1511.07122, 2015.
[4] Guosheng Lin, Chunhua Shen, Anton van den Hengel, and Ian Reid, "Exploring context with deep structured models for semantic segmentation," arXiv preprint arXiv:1603.03183, 2016.
[5] Shuai Zheng, Sadeep Jayasumana, Bernardino Romera- Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr, "Conditional random fields as recurrent neural networks," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1529–1537.
[6] Evan Shelhamer, Kate Rakelly, Judy Hoffman, and Trevor Darrell, "Clockwork convnets for video semantic segmentation," in Computer Vision–ECCV 2016 Workshops. Springer, 2016, pp. 852–868.
[7] Mark Everingham, Luc Van Gool, Christopher KIWilliams, John Winn, and Andrew Zisserman, "The pascal visual object classes (voc) challenge," International journal of computer vision, vol. 88, no. 2, pp.303–338, 2010.
[8] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, andRob Fergus, "Indoor segmentation and support inference from rgbd images," Computer Vision–ECCV 2012, pp. 746–760, 2012.
[9] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele, "The cityscapes dataset for semantic urban scene understanding," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3213–3223.
[10] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bul`o, and Peter Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in Proceedings of the International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 22–29.
[11] Hengshuang Zhao, Xiaojuan Qi, Xiaoyong Shen, Jianping Shi, and Jiaya Jia, "Icnet for real-time semantic segmentation on high-resolution images," arXiv preprint arXiv:1704.08545, 2017.
[12] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," arXiv preprint arXiv:1606.02147, 2016.
[13] Anderson, J. M., Nidhi, K., Stanley, K. D., Sorensen, P., Samaras, C., & Oluwatola, O. A. (2014). Autonomous vehicle technology. Santa Monica, CA: Rand Corp
[14] Altche, F., Qian, X., & De La Fortelle, A. (2017). An algorithm for supervised driving of cooperative semi-autonomous vehicles. IEEE Transactions on Intelligent Transportation Systems, 18: 3527–3539.
[15] Asif Faisal, Tan Yigitcalnar, Md Kamruzzaman, Graham Currie (2019) Understanding autonomous vehicles: A systematic literature review on capability, impact, planning and policy, The Journal of Transport and Land Vol. 12 No. 1: 45- 72
[16] Generalized Cross Entropy Loss for Training DeepNeural Networks with Noisy Labels Zhilu ZhangMert R. Sabuncu 2010 pp.10-19
[17] Very Deep Convolutional Networks by K. Simonyan and A. Zisserman ILSVRC -2014
[18] Fully Convolutional Networks for Semantic Segmentation TPAMI 2017 pp 147-148
[19] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385,2015.pp 30-39
[20] *U-Net: Convolutional Networks for Biomedical Image Segmentation* by Olaf Ronneberger, Phillip Fischer, and Thomas Brox [2015] pp 12-16
[21] Rethinking Atrous Convolution for Semantic Image Segmentation 2017 arXiv pp 7-9



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY