SEGMENTATION OF CARONAR

ARTERY USING SEGNET

A project report submitted for final year of

Bachelor of Technology in

Computer Science and Engineering

By

M.Ruthumma(N130417) T.Srikanya(N130659)

K.VeeraVanitha N130661

Under the Guidance of:

**Mr.M.V.Bhanu Prakash**

## Assistant Professor in Department of Computer Science & Engineering



Department of Computer Science and Engineering

Rajiv Gandhi University of Knowledge Technologies – Nuzvid Nuzvid, Krishna –521202

April, 2019

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES

(A.P. Government Act 18 of 2008) RGUKT-NUZVID, Krishna Dist - 521202 Tele Fax : 08656 – 235557/235150

----------------------------------------------------------------------------------------------------------------------------- ---------

# CERTIFICATE OF COMPLETION

This is to certify that the work entitled, “Shopee(E-commerce portal)” is the bonafied work of M.Ruthumma , ID No: N130417, T.Srikanya , ID No: N130659 and K.Veera Vanitha , ID No : N130661 carried out under my guidance and supervision for final year project of Bachelor of Technology in the department of Computer Science and Engineering under RGUKT IIIT Nuzvid. This work is done during the academic session August 2018 – May 2019, under your guidance.

------------------------------------ -------------------------------------

## Mr.M.V.Bhanu Prakash Mr.R.Upendar Rao

Project Supervisor Head of Department

Assistant professor in Dept. of CSE Faculty Dept. of CSE

RGUKT IIIT Nuzvid, Nuzvid RGUKT IIIT Nuzvid, Nuzvid

## RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES

(A.P. Government Act 18 of 2008) RGUKT-NUZVID, Krishna Dist - 521202 Tele Fax : 08656 – 235557/235150

----------------------------------------------------------------------------------------------------------------------------- ---------

# CERTIFICATE OF EXAMINATION

This is to certify that the work entitled, “Segmenation(Caronary Artery )using segnet” is the bonafide work of M.Ruthumma, ID No: N130417, T Srikanya, ID No: N130659, and K.Veera Vanitha ID No:N130661 and here by accord our approval of it as a study carried out and presented in a manner required for its acceptance in final year of Bachelor of Technology for which it has been submitted. This approval does not necessarily endorse or accept every statement made, opinion expressed or conclusion drawn, as a recorded in this thesis. It only signifies the acceptance of this thesis for the purpose for which it has been submitted.

---------------------------------- --------------------------------

# Mr.Bhanu Prakash Examiner

Project Supervisor Assistant Professor in Dept.of CSE

RGUKT IIIT Nuzvid, Nuzvid

## RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES

(A.P. Government Act 18 of 2008) RGUKT-NUZVID, Krishna Dist - 521202 Tele Fax : 08656 – 235557/235150

----------------------------------------------------------------------------------------------------------------------------- ---------

# DECLARATION

We, M.Ruthumma , ID No: N130417, T.Srikanya, ID No: N130659 and K.Veera Vanitha an hereby declare that the project report entitle “Segmentation of Caronary Artery using Segnet” done by us under the guidance of Mr. Bhanu Prakash is submitted for final year of Bachelor of Technology in Computer Science and Engineering the academic session August 2018 – April 2019 at RGUKT – Nuzvid.

We also declare that this project is a result of our own effort and has not been copied or imitated from any source. Citations from any websites are mentioned in the references.

The results embodied in this project report have not been submitted to any other university or institute for the award of any degree or diploma.

Date : Place :

M, Ruthumma [N130417] T. Srikanya [N130659]

K.Veera Vanitha [N130661]

# ACKNOWLEDGEMENT

We would like to express our profound gratitude and deep regards to my guide Mr. Bhanu Prakash for his exemplary guidance, monitoring and constant encouragement throughout the course of this thesis.

We are extremely grateful for the confidence bestowed in us and entrusting our project entitled “Segmentation using Segnet”.

At this juncture we feel deeply honored in expressing our sincere thanks to him for making the resources available at right time and providing valuable insights leading to the successful completion of our project.

We would like to thank RGUKT Nuzvid Director Mr.D.Suryachandra Rao, Head of Department of Computer Science and Engineering Mr.R.Upendar Rao, faculty and staff for their valuable suggestions and discussions.

Last but not least I thank almighty, and I place a deep sense of gratitude to my family members and my friends who have been constant source of information during the preparation of this project work.

# 

# ABSTRACT

There is large consent that successful training of deep networks requires many thousand annotated training samples. In this , we present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. We show that such a network can be trained end-to-end from very few images and outperforms the prior best method (a sliding-window convolutional network) on the ISBI challenge for segmentation of neustructures in electron microscopic stacks. Using the same network trained on transmitted light microscopy images (phase contrast and DIC) we won the ISBI cell tracking challenge 2015 in these categories by a large margin. Moreover,the network is fast. Segmentation of a 512x512 image takes less than a second on a recent GPU. The full implementation (based on Case) and the trained networks are available .

# at http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net.i

# TABLE OF CONTENTS

[Abstract 6](#_bookmark0)

1. Introduction 09

1.1.Purpose………………………………………………………………………11

1.2 Literature Review……………………………………………………………11

1.3 Project Scope…………………………………………………………………11

1. Definitions 12
   1. Data Sets……………………………………………………………………12

2.1.1 Train Set………………………………………………………………………12

2.1.2 Test Set………………………………………………………………………..12

2.2 IoU…………………………………………………………………………………...12

2.3 Ground Truth…………………………………………………………………………12

2.4 Prediction……………………………………………………………………………..12

1. Overall Description 13
   1. [Architecture 13](#_TOC_250010)

3.2 Decoder Variants……………………………………………………………..14

3.3 Analysis……………………………………………………………………….15

4 .External Interface Requirements 16

4.1[User Interfaces 17-19](#_TOC_250006)

4.2[Software Requirements 20](#_TOC_250005)

4.3[Hardware Requirements 20](#_TOC_250004)

5.Non-Functional Requirements 21

5.1 performance requirements ………………………………………………….21

6. Conclusion and Future Enhancement…………………………………………….22

7 References 23

LIST OF FIGURES

Figure 1.Training Images(Collection of data) 14

Figure 2.Testing Images(Input data) 14

Figure 3.Output 15

Figure 4.Accuracy 15

Figure 5.Flow Chart 17

# CHAPTER 1: INTRODUCTION

# In the last two years, deep convolutional networks have outperformed the state of

# the art in many visual recognition tasks, e.g. [7,3]. While convolutional networks

# have already existed for a long time [8], their success was limited due to the

# size of the available training sets and the size of the considered networks. The

# breakthrough by Krizhevsky et al. [7] was due to supervised training of a large

# network with 8 layers and millions of parameters on the ImageNet dataset with

# 1 million training images. Since then, even larger and deeper networks have been

# trained [12].

# The typical use of convolutional networks is on classication tasks, where

# the output to an image is a single class label. However, in many visual tasks,

# especially in biomedical image processing, the desired output should include

# localization, i.e., a class label is supposed to be assigned to each pixel. More-

# over, thousands of training images are usually beyond reach in biomedical tasks.

# Hence, Ciresan et al. [1] trained a network in a sliding-window setup to predict

# the class label of each pixel by providing a local region (patch) around that pixel

# as input. First, this network can localize. Secondly, the training data in terms

# of patches is much larger than the number of training images. The resulting

# network won the EM segmentation challenge at ISBI 2012 by a large margin.

**1.1 Purpose**

The purpose of this system is to segment the caronary images using segnet architecture a deep learning model in machine learning .

**1.2 Literature Review**

# Semantic pixel-wise segmentation is an active topic of research . Before the arrival of deep networks, the best performing methods mostly relied on hand engineered features classifying pixels independently.

# Typically, a patch is fed into a classifier e.g. Random Forest or Boosting to predict the class probabilities of the center pixel. Features based on appearance or SfM and appearance have been explored for the CamVid road scene understanding test . These per-pixel noisy predictions (often called unary terms) from the classifiers are then smoothed by using a pair-wise or higher order CRF , to improve the accuracy.

## 

## 1.3 Project Scope

Segmentation is the process of segmenting the different parts of caronary images called ground truth and prediction masks .Includes the process of prediction , IoU (Intersection over Union) and displaying the different images of original image. Therefore it consists of classification and prediction . This project supports all the MRI scanned images to work on them.

# **CHAPTER 2 : DEFINITIONS**

## 2.1 Data Sets

2.1.1 Training Set

The data which we gathered takes as an input means huge collection of data (bio-medical images .We use the CamVid road scenes dataset to benchmark the performance of the decoder variants.

This dataset is small, consisting of 367 training and 233 testing RGB images (day and dusk scenes) at 360480 resolution. The challenge is to segment 11 classes such as road, building, cars, pedestrians, signs, poles, side-walk etc. We perform local contrast normalization [54] to the RGB input .

2.1.2 Test Set

The derived data from training set which is used as input data, on which we have to perform segmentation task with the help of diffferent classification techniques and prediction techniques .

2.2 IoU (Intersection over Union)

Returns Intersection over Union score for ground truth and predicted masks.

Intersection over Union is an evaluation metric used to measure the accuracy of an object detector on a particular dataset.

2.3 Ground Truth

It refers to actual image. For Example if we consider segmentation ground truth image is the part which we have to focus to segment of the whole image .

2.4 Prediction

Predictive analytics is the process of using data analytics to make predictions based on data.

This process uses data along with analysis, statistics, and machine learning techniques to create a predictive model for forecasting future events.

# **CHAPTER 3: OVERALL DESCRIPTION**

# 3.1 ARCHITECTURE

SegNet has an encoder network and a corresponding decoder network, followed by a final pixelwise classification layer. This architecture is illustrated in Fig. 3. The encoder network consists of 13 convolutional layers which correspond to the first 13 convolutional layers in the VGG16 network [1] designed for object classification. We can therefore initialize the training process from weights trained for classification on large datasets [41]. We can also discard the fully connected layers in favour of retaining higher resolution feature maps at the deepest encoder output. This also reduces the number of parameters in the SegNet encoder network significantly (from 134M to 14.7M) as compared to otherrecent architectures [2], [4] (see. Table 6). Each encoder layer has a corresponding decoder layer and hence the decoder network has 13 layers. The final decoder output is fed to a multi-class soft-max classifier to produce class probabilities for each pixel independently.

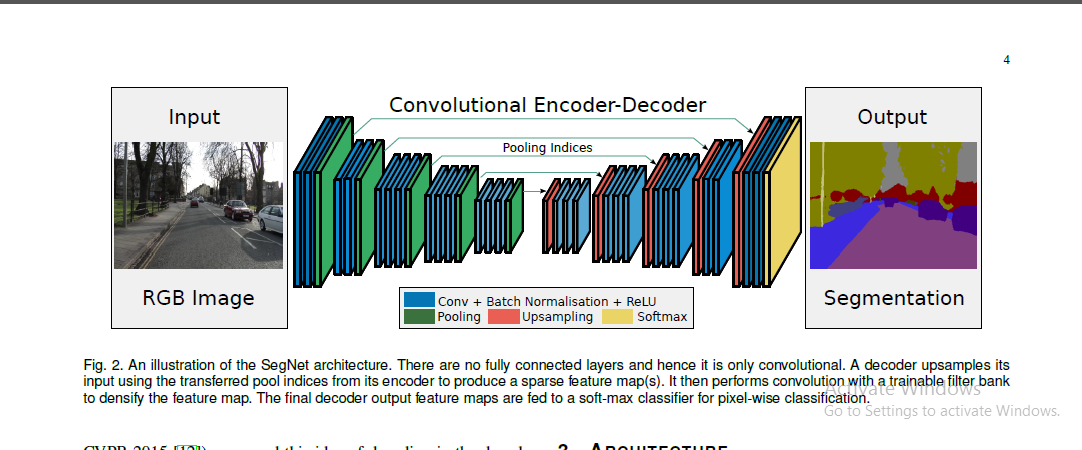
Each encoder in the encoder network performs convolution with a filter bank to produce a set of feature maps. These are then batch normalized [51], [52]). Then an element-wise rectifiedlinear non-linearity (ReLU) max(0; x) is applied. Following that,

max-pooling with a window and stride 2 (non-overlapping window) is performed and the resulting output is sub-sampled by a factor of 2. Max-pooling is used to achieve translation invariance over small spatial shifts in the input image. Sub-sampling results in a large input image context (spatial window) for each pixel in the feature map. While several layers of max-pooling and sub-sampling can achieve more translation invariance for robust

classification correspondingly there is a loss of spatial resolution of the feature maps. The increasingly lossy (boundary detail) image representation is not beneficial for segmentation where boundary delineation is vital. Therefore, it is necessary to capture and store boundary information in the encoder feature maps before sub-sampling is performed. If memory during inference is not constrained, then all the encoder feature maps (after subsampling)

can be stored. This is usually not the case in practical applications and hence we propose a more efficient way to store this information. It involves storing only the max-pooling indices,

i.e, the locations of the maximum feature value in each pooling window is memorized for each encoder feature map. In principle, this can be done using 2 bits for each 2 2 pooling window and is thus much more efficient to store as compared to memorizing feature map(s) in float precision.



The appropriate decoder in the decoder network upsamples its input feature map(s) using the memorized max-pooling indices from the corresponding encoder feature map(s). This step produces sparse feature map(s). This SegNet decoding technique is illustrated in Fig. 3. These feature maps are then convolved with a trainable decoder filter bank to produce dense feature maps.

A batch normalization step is then applied to each of these maps. Note that the decoder corresponding to the first encoder (closest to the input image) produces a multi-channel featuremap, although its encoder input has 3 channels (RGB).

3.2 DECODER VARIANTS

Many segmentation architectures [2], [3], [4] share the same encoder network and they only vary in the form of their decoder network. Of these we choose to compare the SegNet decoding technique with the widely used Fully Convolutional Network

(FCN) decoding technique [2], [10]. In order to analyse SegNet and compare its performance with FCN (decoder variants) we use a smaller version of SegNet, termed SegNet-Basic 1, which has 4 encoders and 4 decoders.

All the encoders in SegNet-Basic perform max-pooling and subsampling and the corresponding decoders upsample its input using the received max-pooling indices. Batch normalization is used after each convolutional layer in both the encoder and decoder

network. No biases are used after convolutions and no ReLU nonlinearityis present in the decoder network. Further, a constant kernel size of 7 7 over all the encoder and decoder layers is chosen to provide a wide context for smooth labelling i.e. a pixel in the deepest layer feature map (layer 4) can be traced back to a context window in the input image of 106106 pixels. This small size of SegNet-Basic allows us to explore many different variants

(decoders) and train them in reasonable time. Similarly we create FCN-Basic, a comparable version of FCN for our analysis which shares the same encoder network as SegNet-Basic but with the FCN decoding technique (see Fig. 3) used in all its decoders.

On the left in Fig. 3 is the decoding technique used by SegNet(also SegNet-Basic), where there is no learning involved in theupsampling step. However, the upsampled maps are convolvedwith trainable multi-channel decoder filters to densify its sparse inputs. Each decoder filter has the same number of channels as the number of upsampled feature maps. A smaller variant is one where the decoder filters are single channel, i.e they only convolve

their corresponding upsampled feature map. Thi variant(SegNet-Basic-SingleChannelDecoder)

reduces the number of trainable parameters and inference time significantly.

3.3 ANALYSIS

To compare the quantitative performance of the different decoder variants, we use three commonly used performance measures:global accuracy (G) which measures the percentage of pixels the mean of the predictive accuracy over all classes and mean intersection over union (mIoU) over all classes as used in the Pascal VOC12 challenge. The mIoU metric is a more stringent metric than class average accuracy since it penalizes false positive predictions. However, mIoU metric is not optimized for directly through the class balanced cross-entropy loss.

The mIoU metric otherwise known as the Jacard Index is most commonly used in benchmarking. However, Csurka et al note that this metric does not always correspond to human qualitative judgements (ranks) of good quality segmentation. They show with

examples that mIoU favours region smoothness and does not evaluate boundary accuracy, a point also alluded to recently by the authors of FCN. Hence they propose to complement the mIoU metric with a boundary measure based on the Berkeley contour matching score commonly used to evaluate unsupervised image segmentation quality [59]. Csurka et simply extend this to semantic segmentation and show that the measure of semantic contour accuracy used in conjunction with the mIoU metric agrees more with human ranking of segmentation outputs.

figure 1: Registration Page

# CHAPTER 4: EXTERNAL INTERFACE REQUIREMENTS

## User Interfaces

Figure 5: Home page

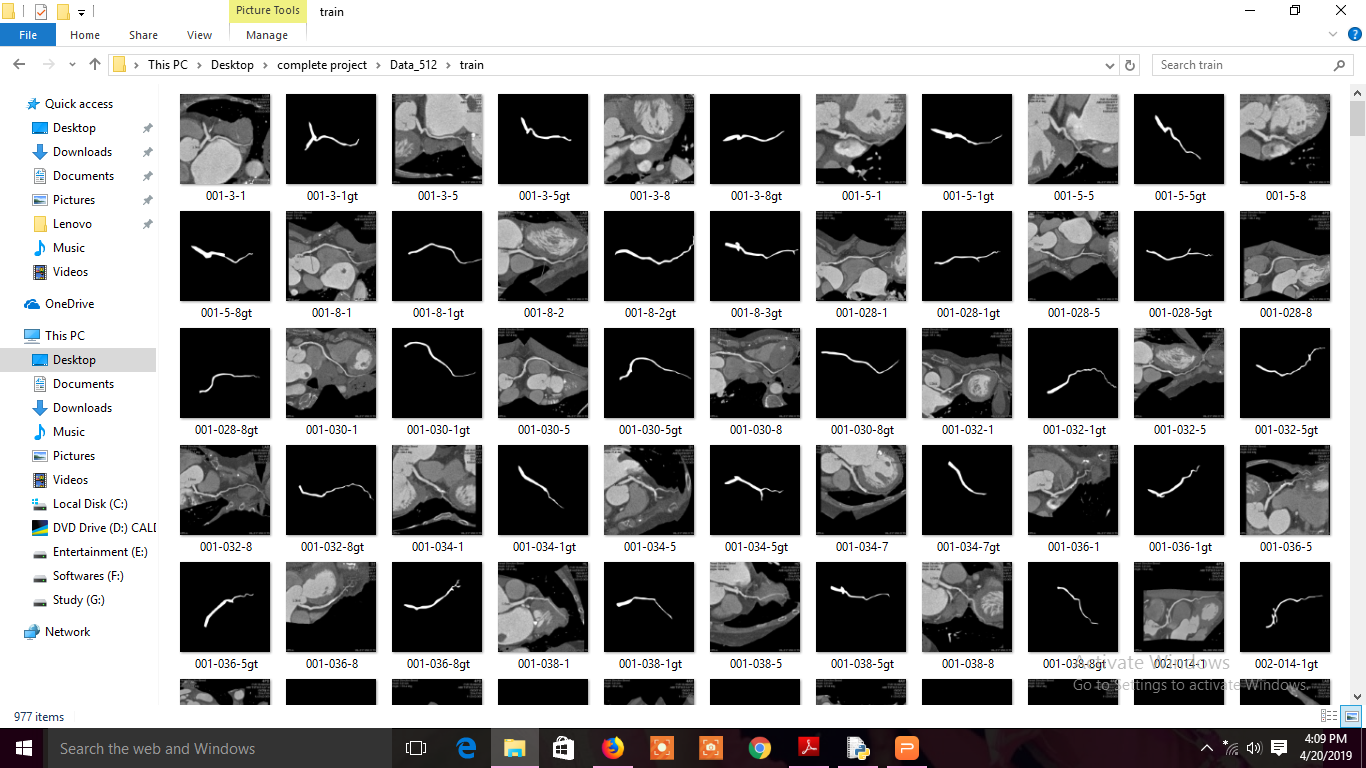


Fig 1 Train Set Data

Figure 6:Product view

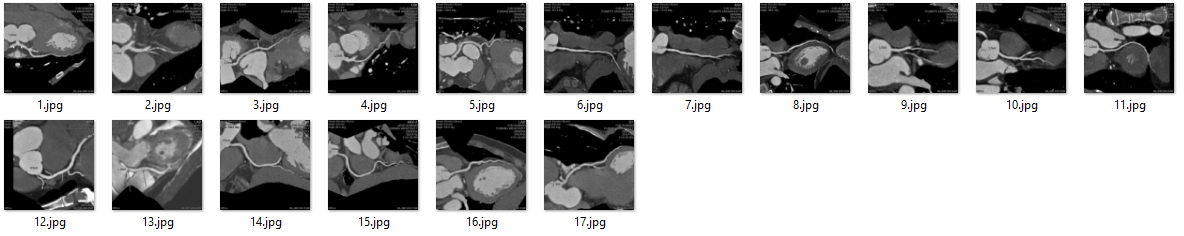


Fig: Test set Data

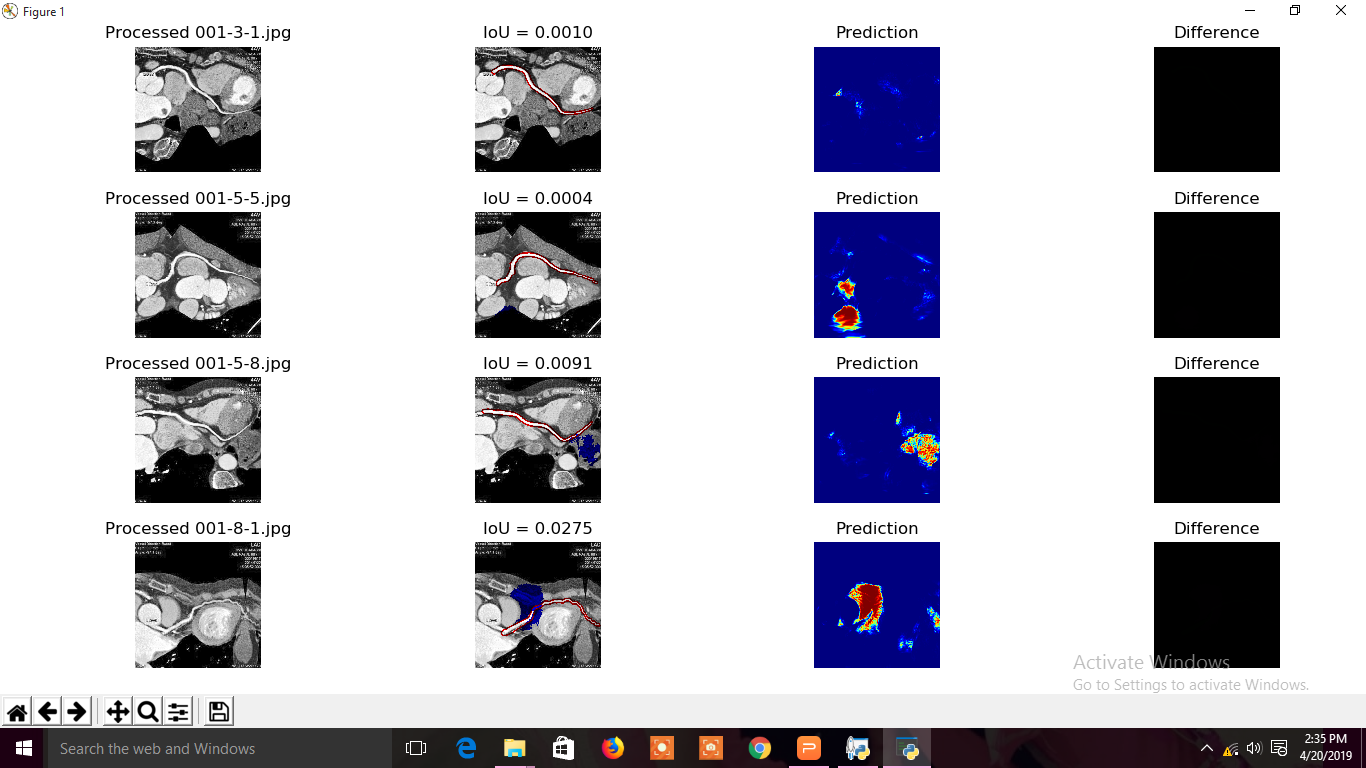


Figure 3: Output Image

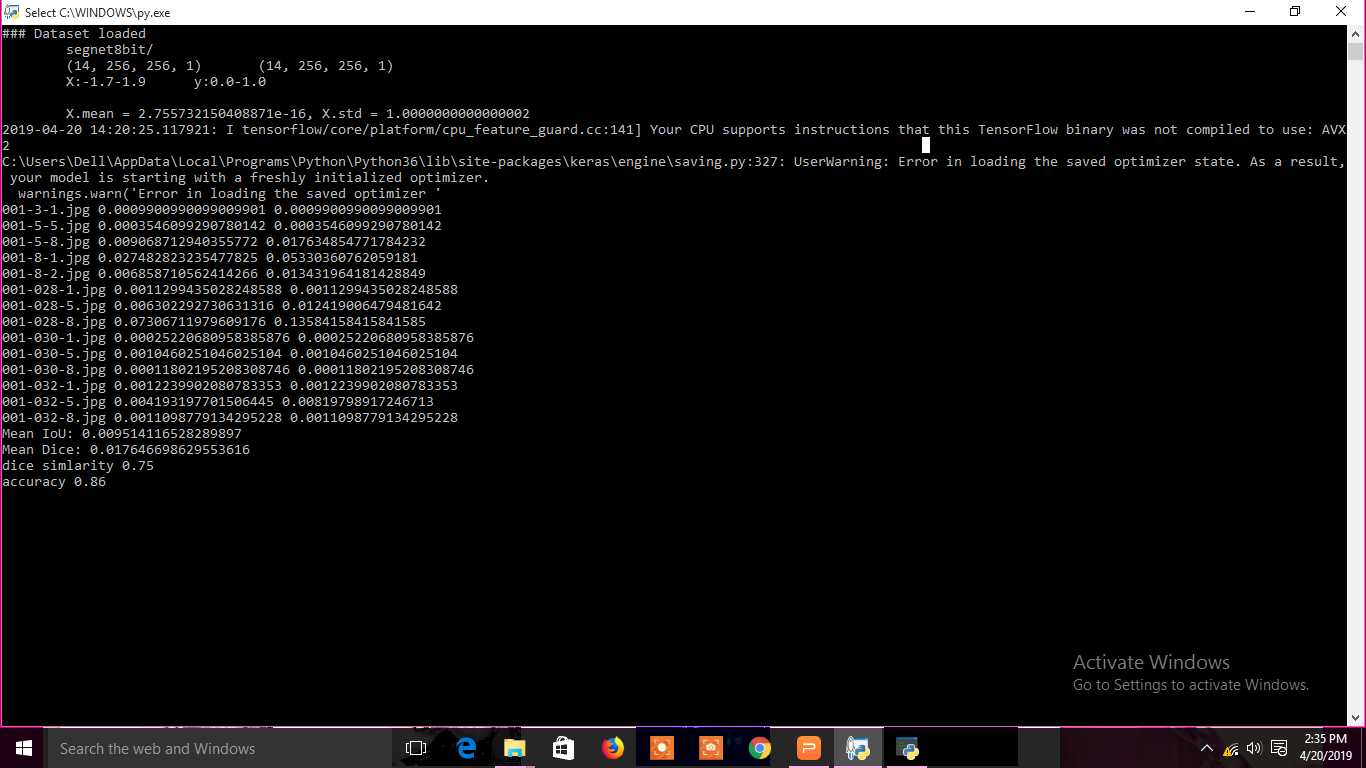


Figure 4 : Output image

4.2 Software Requirements

1. Python IDLE 3.6
2. Packages

2.1 Numpy

2.2 Tensor flow

2.3 Matplot lib

2.4 Pandas

2.5 Keras

2.6 Skimage

4.3 Hardware Requirements

1. RAM :4GB

2. HARD DISK 500 GB

3 . Intel Core i3

1. NON-FUNCTIONAL REQUIREMENTS

5.1 Performance Requirements



\

1. CONCLUSION AND FUTURE ENHANCEMENT
   1. Conclusion

We presented SegNet, a deep convolutional network architecture for semantic segmentation. The main motivation behind SegNet was the need to design an efficient architecture for road and indoor scene understanding which is efficient both in terms of memory and computational time. We analysed SegNet and compared it with other important variants to reveal the practical trade-offs involved in designing architectures for segmentation, particularly training time, memory versus accuracy. Those architectures which store the encoder network feature maps in full perform best but consume more memory during inference time. SegNet on the other hand is more efficient since it only stores the max-pooling indices of the feature maps and uses them in its decoder network to achieve good performance .

6.2 Future Enhancement

Deep learning models have often achieved increasing success due to the availability of massive datasets and expanding model depth and parameterisation. However, in practice factors like memory and computational time during training and testing are important

factors to consider when choosing a model from a large bank of models.

This was the primary motivation behind the proposal of SegNet, which is significantly smaller and faster than other competing architectures, but which we have shown to be efficient for taskssuch as road scene understanding etc.

7 REFERENCES

1. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[2] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in CVPR, pp. 3431–3440, 2015.

[3] C. Liang-Chieh, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected crfs,” in ICLR, 2015.

[4] H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation,” in ICCV, pp. 1520–1528, 2015.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in CVPR, pp. 1–9, 2015.

[6] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR, vol. abs/1409.1556, 2014.

[7] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning hierarchical features for scene labeling,” IEEE PAMI, vol. 35, no. 8, pp. 1915–1929,

2013.

[8] N. Hft, H. Schulz, and S. Behnke, “Fast semantic segmentation of rgb-d scenes with gpu-accelerated deep neural networks,” in KI 2014: Advances in Artificial Intelligence (C. Lutz and M. Thielscher, eds.), vol. 8736 of Lecture Notes in Computer Science, pp. 80–85, Springer International Publishing, 2014.

[9] R. Socher, C. C. Lin, C. Manning, and A. Y. Ng, “Parsing natural scenes and natural language with recursive neural networks,” in ICML, pp. 129–136, 2011.

10] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, “Conditional random fields as recurrent neural networks,” in Proceedings of the IEEE International Conference on

Computer Vision, pp. 1529–1537, 2015. W. Liu, A. Rabinovich, and A. C. Berg, “Parsenet: Looking wider to see better,” CoRR, vol. abs/1506.04579, 2015.