B.Sc. Thesis Proposal

Bangladesh Army International
University of Science and technology (BAIUST) Date: 03/10/2019

Author: **Arik Md. Isthiaque**
E-mail: arikbncc@gmail.com
Phone: +8801521493029
Supervisor: Arifa Islam Champa

Proposed Topic:

Improvise SegNet architecture with Canny edge detector algorithm for more boundary accurate image segmentation.

Introduction

SegNet, is an encoder decoder architecture built with deep convolutional neural network for pixel-wise semantic segmentation for images and videos. It is generally built for roads, buildings, cars, pedestrians and for differentiated the context between roads and side-walks [1]. SegNet extracts low resolution pixel information from input image and classify them as per pixel. This extraction must generate some patterns which can be used in boundary localization.

SegNet is built with a stack of encoders with a similar number of decoders. The decoders made the output as the same size of the input. This is an important drawback for the architecture as the increasing number of layers increase the parameters of the model [2].

Our main contribution on this paper is to study and analysis the details of SegNet encoder decoder architecture and reduce the parameters of convolutional networks and perform Canny edge detector algorithm for more boundary localized feature maps, so that it can be more memory efficient and produce more robust sematic segmentation.

Motivation for Research

Image segmentation is a process where the objects of an image is classified by their pixel information. In an image certain pixel share some common characteristics. By identifying them we can label the pixels, so they can be detected from the image. We use image segmentation to find objects in images like text (OCR), cars (automated vehicle), cancer, tumor (medical image). In recent days, deep learning algorithms got a success in handwritten OCR, NLP [3], [4]. It has also a huge active interest for pixel-wise semantic segmentation. Recently there are some approaches are being carried out to adopt deep learning algorithm for pixel-wise semantic segmentation [5]. The results of these approaches are very encouraging [6]. But we still need better architecture to run the segmentation process. So, improvise the current architectures is needed for more robust segmentation.

Objective

- Reduce the number of parameters in the architecture
- Increase the number of classes
- More accurate feature extraction
- More boundary localized feature maps

Background Study

SegNet built with an encoder layer and a decoder for a pixelwise segmentation. The model is in Figure 1. The encoder layer (is) built with 13 convolutional layers which follow the VGG16 network [7]. The training process can be initialized from weights trained for classification with large datasets [8].

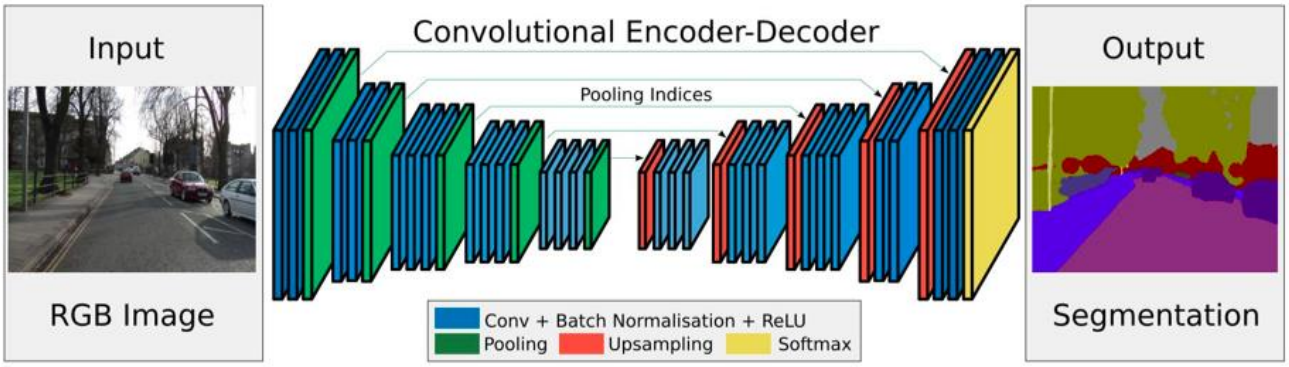


Figure 1: The complete SegNet encoder decoder architecture [1].

SegNet is based on VGG16 but without fully connected layers so the parameters in each convolutional layer as follows, shown in details for first 3 layers:

Input image: RGB (3 channels)

1st layer: 3x3@64 -> # params = (3x3x3+1) *64 = 1792

2nd layer: 3x3@64 -> # params = (3x3x64+1) *64 = 36,928

3rd layer: 3x3@128 -> #params = (3x3x64+1) *128 = 73,856

…….

The next layers parameters will be 147584, 295168, 590080, 590080, 1180160, 2359808, 2359808, 2359808, 2359808, 2359808. Total sum will be 14.7M.

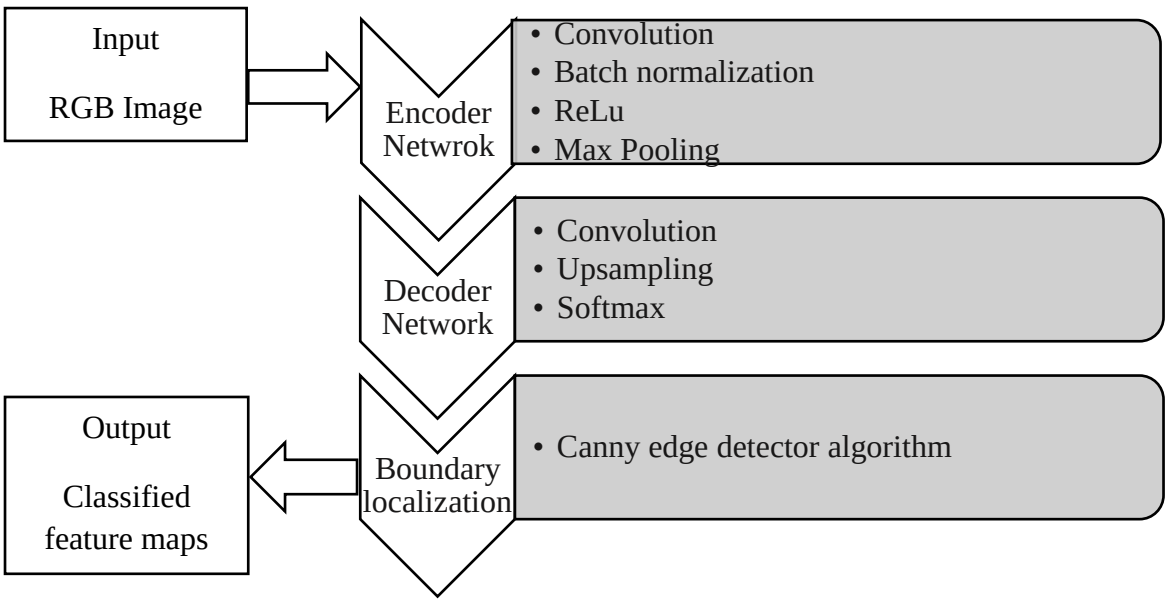
Working Hypothesis

The architecture will take RGB image as input and the input will be directly fed into the encoder network. The encoder network is topologically identical to the 13 convolutional layers in the VGG16 network [1]. The network will perform convolution with a filter bank to extract a set of feature maps. Then they will go for a batch normalization process [9]. Then an

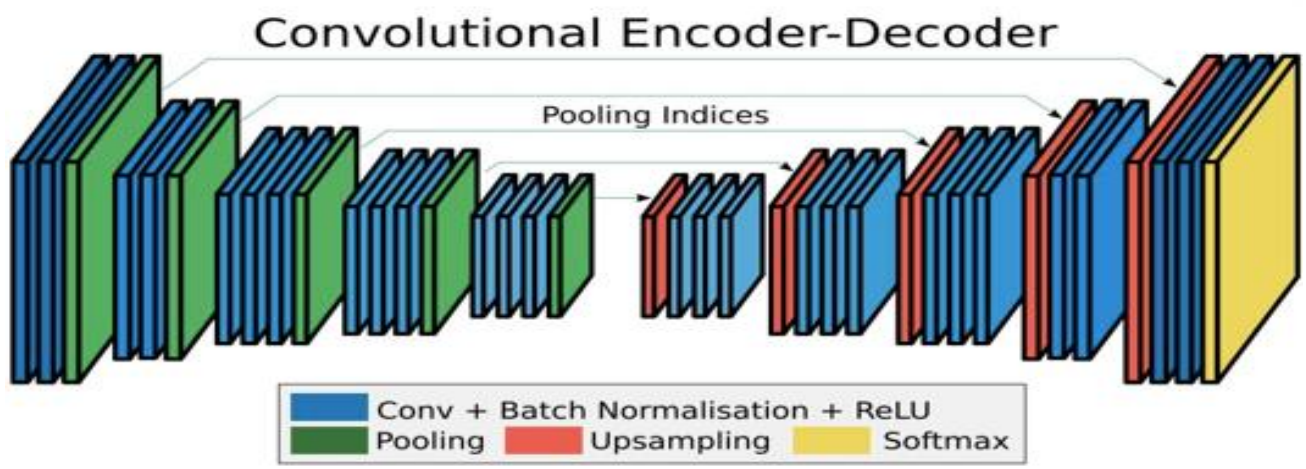
element-wise ReLU will be applied. After each convolutional operation a max-pooling will be performed. Max-pooling is used to achieve translation invariance over small spatial shifts in the input image.

The most important phase of this architecture is the decoder network. Here, there is no learning involved in this network. First upsampling will be performed in the feature maps. Here the upsampling process will be performed by inverse convolution using a fixed or trainable multi-channel upsampling kernel. The final decoder output feature maps will be fed to a soft-max classifier for pixel-wise classification. After that we will run Canny edge detector for more localized boundary detection.

Methodology



(a)

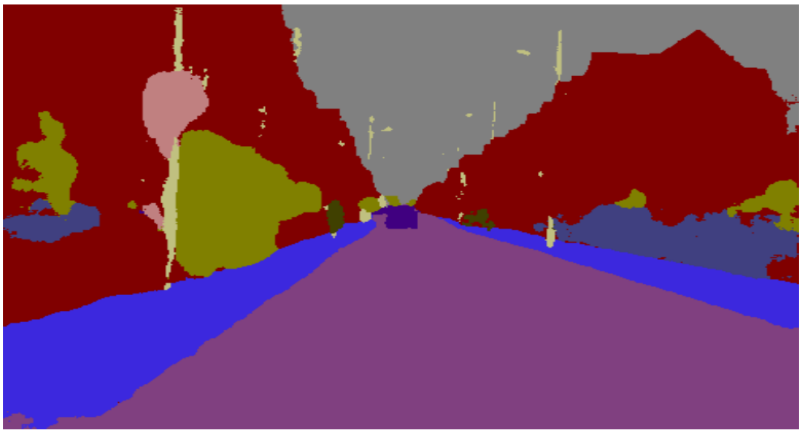


(b)

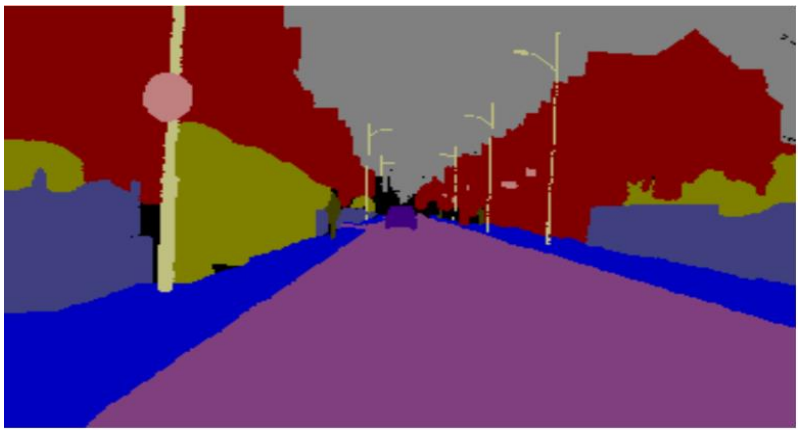
Figure 2: (a) The proposed methodology, (b) Convolutional Encoder- Decoder [1].

Outline

(a)



(b)



(c)

Figure 3: (a) Input RGB image, (b) Current SegNet output image, (c) Expected improvised output [1].

Conclusion

In this implementation one major challenge is to build the DCNN properly, because the accuracy of the architecture depends on the DCNN model. It is quite challenging task to interface the Canny edge detector algorithm with the DCNN model and without analyzing the real time output it's tough to say how much it will fit to the DCNN model. We will analyze the architecture with "Street-View Change Detection with Deconvolutional Networks Dataset" which contains aligned image pairs from street-view imagery with structural, lighting, weather and seasonal changes.

References

1. V. Badrinarayanan, A. Kendall, R. Cipolla, Senior Member, IEEE, "*SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation* "[2017 TPAMI] https://arxiv.org/abs/1511.00561. IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 39 , Issue: 12 , Dec. 1 2017)
2. V. Badrinarayanan, A. Kendall, R. Cipolla, Senior Member, IEEE, "*SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling*", https://arxiv.org/abs/1505.07293. Submitted on 27 May 2015
3. 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.
4. K. Simonyan and A. Zisserman, "*Very deep convolutional networks for large-scale image recognition*," CoRR, vol. abs/1409.1556, 2014.
5. 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.
6. 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.
7. K. Simonyan and A. Zisserman, "*Very deep convolutional networks for large-scale image recognition*," arXiv preprint arXiv:1409.1556, 2014.
8. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "*ImageNet Large Scale Visual Recognition Challenge*," International Journal of Computer Vision (IJCV), pp. 1–42, April 2015.
9. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deepnetwork training by reducing internal covariate shift," CoRR, vol. abs/1502.03167, 2015.

Supervisor Signature