Image Segmentation

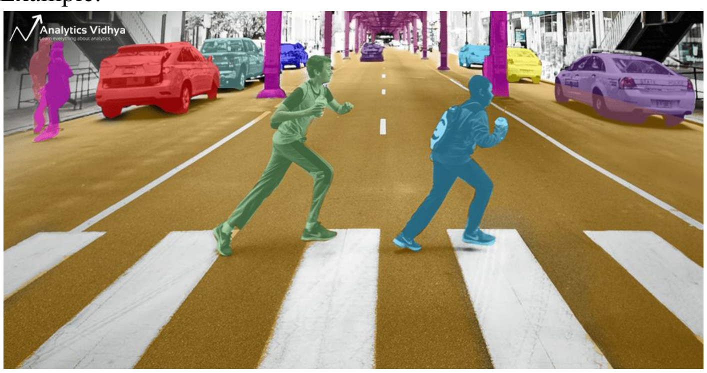
Aditya Tiwari (2016csb1029), Sonu Kedia (2016meb1142)

Date : 22th November, 2019

Abstract- Image Segmentation is a field of Computer vision where an image is grouped in segments for different purposes. Grouping similar pixels together on the basis of some attribute similarity is called image segmentation.

"The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze."~ Wikipedia

Example:



As you can see in the above figure, roads' pixels are grouped together and similarly for the human figures.

Input: Image
Output: Image with segmentation where similar pixels are of similar color.
Example:

More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics

Motivation: The motivation of the project comes from industrial requirements where segmentation is a preprocessing step and some data needs to be analyzed. Example Content Based Image Retrieval, once the segmentation is done, different tags can be assigned to segments and retrieving similar images on the basis of similar tags can be done. Medical Purposes require Computer Vision support for doctors for analyzing the medical images of the patient. Object Recognition can also be done after segmentation. In most cases, image segmentation is a preprocessing or pre requisite step.

LITERATURE REVIEW

There are many algorithms used for image segmentation, and some of them segmented an image based on the object while some can segment automatically. Nowadays, no one can point out which the optimal solution is due to different constraints.

Some of the common neural network:

1. **U-net Convolutional Neural Network:** U-Net was created by Olaf Ronneberger, Philipp Fischer, Thomas Brox in 2015 at the paper "UNet: Convolutional Networks for Biomedical Image Segmentation". It's an improvement and development of FCN.
2. **AlexNet:** AlexNet competed in the ImageNet Large Scale Visual Recognition Challenge[5] on September 30, 2012. The network achieved a top-5 error of 15.3%, more than 10.8 percentage points lower than that of the runner up. The original paper's primary result was that the depth of the model was essential for its high performance, which was computationally expensive, but made feasible due to the utilization of graphics processing units (GPUs) during training
3. **ResNet:** ResNet makes it possible to train up to hundreds or even thousands of layers and still achieves compelling performance. Taking advantage of its powerful representational ability, the performance of many computer vision applications other than image classification have been boosted, such as object detection and face recognition.
4. **Fully Connected Convolutional Neural Network:** Fully Convolutional Networks (FCNs) owe their name to their architecture, which is built only from locally connected layers, such as convolution, pooling and upsampling.

INTRODUCTION

Project is based on our work on Fully Connected convolutional neural network and U-net convolutional neural networks. The objective is to employ some traditional techniques for image segmentation and then employ Deep neural networks following some research papers and experiment to showcase the results.

METHOD

FCN: Fully Convolutional Networks (FCNs) owe their name to their architecture, which is built only from locally connected layers, such as convolution, pooling and upsampling.
With no dense layer helps to reduce the number of parameters and computation time and have the ability to predict for arbitrary input size image.

Key features of FCC Architecture:

- FCN transfers knowledge from VGG16 to perform semantic segmentation. VGG16 is a pre trained model which is used in FCN
- The fully connected layers of VGG16 is converted to fully convolutional layers, using 1x1 convolution.
- The upsampling of these low resolution semantic feature maps is done using transposed convolutions
- At each stage, the upsampling process is further refined by adding features from coarser but higher resolution feature maps from lower layers in VGG16.
- Skip connection is introduced after each convolution block to enable the subsequent block to extract more abstract, class-salient features from the previously pooled features.
- There are 3 versions of FCN (FCN-32, FCN-16, FCN-8)

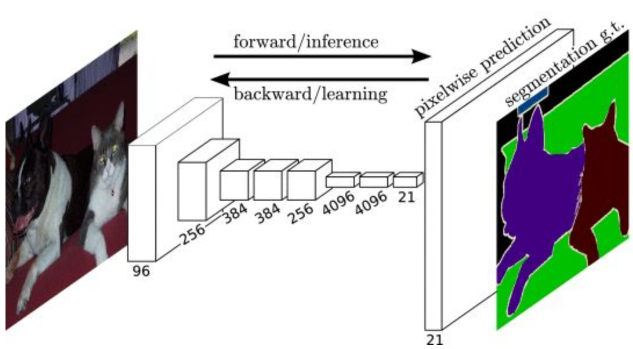


Fig 4: Fully Convolutional Networks Architecture

U-Net Convolutional Network:

- Developed for biomedical image segmentation
- Downsample to reducing the number of trainable parameters
- The main idea is to supplement a usual contracting network by successive layers
- Pooling operations are replaced by upsampling operators to increase the resolution of images
- The network only uses the valid part of each convolution without any fully connected layers.
- To predict the pixels in the border region of the image, the missing context is extrapolated padding same pixels in input

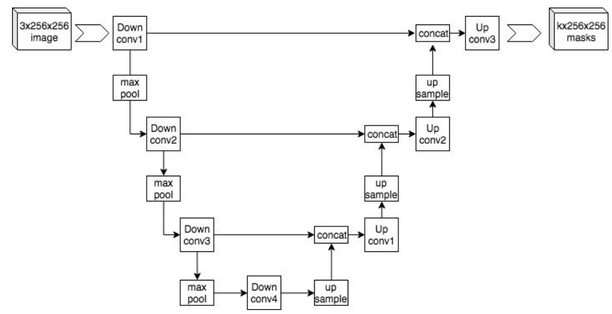


Fig 5: U-net CNN Architecture

The network consists of a contracting path and an expansive path, which gives it the u-shaped architecture.

The contracting path is a typical convolutional network that consists of repeated application of convolution, each followed by a rectified linear unit (ReLU) and a max pooling operation

EXPERIMENT

FCN Model has been proven to outperform U-net model with Fully connected networks. But has 329 Million Parameters and U-net has 9 Million Parameters.

Dataset:

- “Setting an attention region for convolutional neural networks using region selective features, for recognition of materials within glass vessels”
- The Paper presented a new Image Segmentation Dataset of Chemical Vessels and Annotated Vessels.
- 601 Images with annotated expected output

Modified U-Net CNN:

.

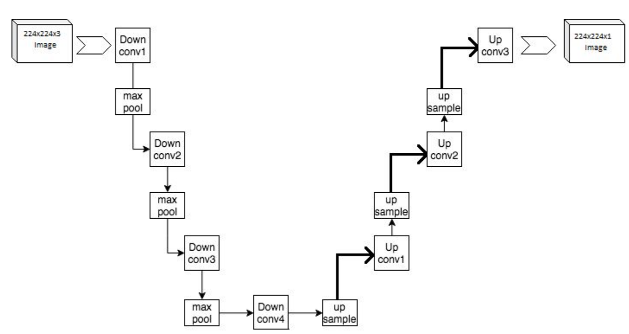


Fig 6: Modified Architecture of U-Net

We have done few modifications in the U-Net:

1. Removed Skip connections
2. Used transposed convolutions(bilinear interpolation) for up sampling (Like in FCN)
3. This reduced parameters from 9 million to 1.4 Million

Our Modified U-net has 1.4 Million Parameters. Loss Function Used is Binary Cross Entropy Output is probability for each class to belong to vessel. Modified U-net is trained for 20 epochs, batch size = 32 with adam optimizer.(Time ~ 13 Hours, 40 min/epoch)

Binary Cross Entropy Function

$$-(y \log(p) + (1 - y) \log(1 - p))$$

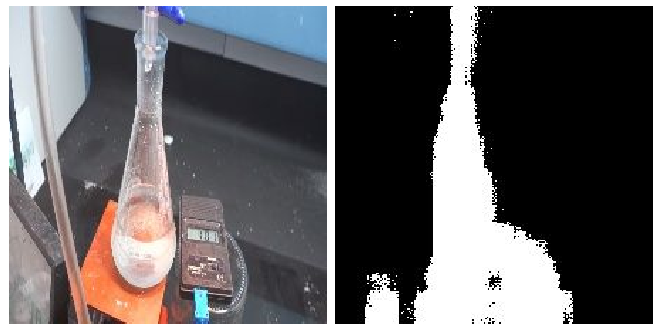
Y ~ true output, P ~ predicted output, Threshold = 0.5
Since output is probability for each pixel, we threshold the pixel values.

RESULTS

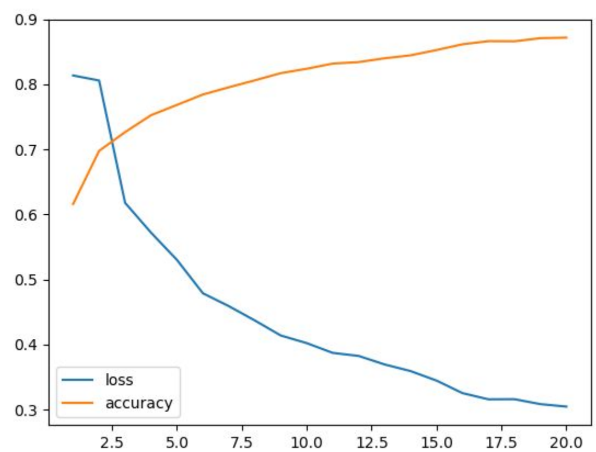
Input & Output Image

Input & Output Image

Input & Output Image



Input & Output Image



Graph 1: Loss & Accuracy Graph

INFERENCE

- Skip Connections definitely Improve results but were removed to reduce parameters
- The Model was trained on 20.
- The Graph shows loss's gradient is decreasing, hence for the modified U-net model maybe 40 epochs could give better results.
- But to improve the results a lot, skip connections are required like in original U-net

REFERENCES

1. Scene Parsing through ADE20K Dataset. Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso and Antonio Torralba. Computer Vision and Pattern Recognition (CVPR), 2017.
2. Microsoft COCO: Common Objects in Context
3. Semantic Image Segmentation using Deep Convolutional Neural Networks and Super-Pixels Santosh Jangid, P S Bhatnagar
4. https://en.wikipedia.org/wiki/U-Net
5. https://www.youtube.com/watch?v=0Lc1SsuP8U8
6. https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf
7. https://towardsdatascience.com/review-fcn-semantic-segmentation-eb8c9b50d2d1
8. https://nanonets.com/blog/how-to-do-semantic-segmentation-using-deep-learning/
9. Original paper of Faster R-CNN : https://arxiv.org/abs/1703.06870
10. Setting an attention region for convolutional neural networks using region selective features, for recognition of materials within glass vessels
11. Tensorflow
12. Adam: A Method for Stochastic Optimization