CS x476 Project 2

Aaron Lopes
alopes7@gatech.edu
alopes7
903407727
Section 4476

Part 1.1: Image Segmentation

What is image segmentation? Why do we want to do image segmentation?

The goal of image segmentation is to separate an imagine into distinguishable “objects”. By grouping together pixels of similar value, future image processing computations are sped up.

What are some applications that use image segmentation? List at least 2.

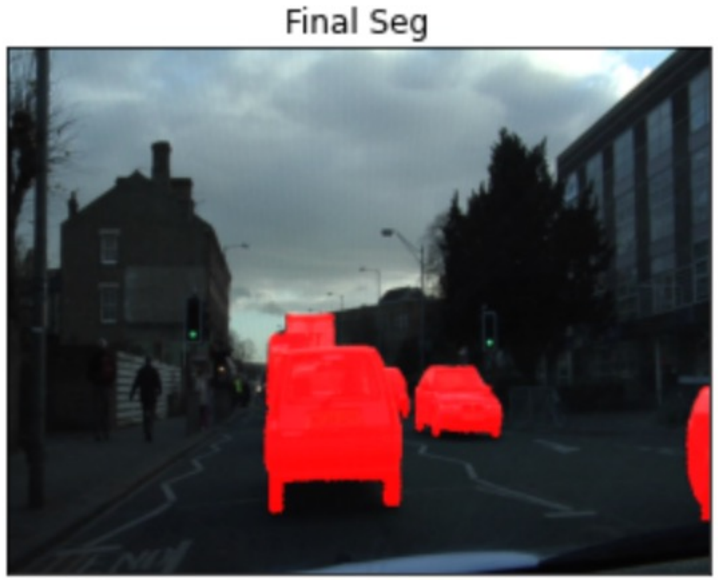
Object detection for self driving cars and medical imaging to locate tumours are two exciting applications that involve image segmentation.

Part 1.2: Sigmoid v. Softmax

What is the difference between sigmoid and softmax in terms of how they are used? What is the similarity in terms of their output values?

When using image segmentation to detect one type of class, the sigmoid function can be used to convert a broad spectrum of values to range between 0 and 1. When segmenting into multiple classes, softmax can be used to output probability values that also range between 0 and 1. The range of their output values is the same, however softmax will output C values for C classes.

Part 1.3: Apply Mask to Image



Part 2.1a: Pre-trained Models

!! Please see the link in the title to help you answer the following questions.

What are some other **available encoders** that are not used in the project 2? List 4.

From ResNet, the available encoders are ResNet18, ResNet34, ResNet101, ResNet152.

Part 2.1b: Pre-trained Models

!! Please see the link in the title to help you answer the following questions.

What is the architecture of one of the **segmentation models** that you are interested in that's not covered in the project 2? Provide some details of this architecture from its associated paper.

Unet++ contains an encoder and decoder that are connected by a series of nested dense convolutional blocks. This is to bridge the semantic gaps within feature maps before fusion. There are a different number of up-sampling, down-sampling, and skip-convolutions done based on the location of the input.

Part 2.1c: FCN Paper

!! Please see the link in the title to help you answer the following questions.

What is the result and reason of viewing fully connected layers as convolutions with kernels? (Hint: Look into Paper Section 3.1)

It's because the kernels cover the entire input region, casting them into fully convolutional networks that take input of any size and output classification maps.

What are the number of convolutional layers and parameters of the 3 models used for segmentations? (Hint: Look into Paper Table 1 and Section 4)

FCN-AlexNet has 8 convolution layers and 57M parameters, FCN-VGG16 has 16 convolution layers and 134M parameters, and FCN-GoogLeNet has 22 convolution layers and 6M parameters.

Part 2.2: VGG

!! Please see the link in the title to help you answer the following questions.

What is the total number of convolutional layers (Conv) in VGG-19? What is the total number of fully connected layers in VGG-19? (Hint: Look into Paper Figure 3)

There are 16 convolutional layers in VGG-19.
There are 3 fully connected layers.

What do you notice about the image height and image width as you go through the _encoder_ of the FPN+VGG-19? What about the _decoder_ of the FPN+VGG-19? (This is the question 1 is the Notebook)

The image height and width while going through the encoder FPN+VGG-19 decreases. The image heigh and width while going through the decoder increases.

Part 2.3: Resnet

!! Please see the link in the title to help you answer the following questions.

What is the total number of convolution layer (Conv) in ResNet-50? What is the total number of fully connected layers in ResNet 50? (Hint: Look into the Figure linked in notebook)

There are 48 convolution layers. There is one fully connected layer.

What do you notice about the size of the FPN+ResNet-50 network/model in comparison with the FPN+VGG-19 network/model? What are other major differences that you notice between the two model architectures? (List at least 2.) (This is the question 2 is the Notebook)

ResNet50 has 26 million parameters while VGG-19 has 19 million. Other major differences are the series of Sequential layers in FPN and the image size decreasing faster in FPN.

Part 2.4: Feature Map

What feature in the input image does the FCN-ResNet50 model appear to focus on:

- In the first layer of its encoder,
- In the last layer of its encoder
- In the last layer of its decoder?

The first layer of the encoder seems to focus on segmenting the sky. The last layer of the encoder seems to focus on the variety of edges that segment the distinct objects. The last layer of the decoder seems to focus on the various shadows cast by the objects.

What does this tell you about the learning process of the model?

The learning process of the model follows the order of distribution of light, from its source to the shadow.

Part 3.1: IoU

IoU encodes the shape properties of the object into the region property with normalized measure focusing on the area. What is the benefit of such property of IoU? (Hint: Check out the section 1 of paper linked in the title)

According to the paper, this property makes IoU invariant to the scale of the problem under consideration. Essentially stating that the images and bounding box can be of any size.

Which prediction result would have higher IoU score? Please Explain the reason. (This is the question 3 is the Notebook)

Pred Mask1 would have the higher IoU score, since the Union of the bounding box and image Pred Mask2 cover a larger area than that of Pred Mask2.

Part 3.2: Apply IoU

What is the IoU score for VGG-19 and ResNet-50? (Output from your Jupyter Notebook)

vgg19 IoU score is: [0.8623824, 0.85832924, 0.8974035]

ResNet50 IoU score is: [0.8536674, 0.8288056, 0.85655344]

Which FCN backbone has better performance? Based on your understanding, why does one FCN backbone perform better than the other?

VGG19 performs better than ResNet50, presumably because the bounding boxes generated by VGG19 are tighter.

Part 3.3: Performance

What is the relationship between the number of parameter and the performance?

More parameters can lead to overfitting the training data, resulting in worse performance.

Extra Credit 1: PSPNet

!! Please see the link in the title to help you answer the following questions.

What are some shortcoming of FCN mentioned in the PSPNet Paper? (Hint: Look into Paper Section 1)

A major issue with FCN based models is a “lack of suitable strategy to utilize global scene category clues”. In other words, FCN does not consider enough of the image context before classification.

What is the main difference between FCN and PSPNet? (Hint: Look into Paper Section 1)

The main difference between PSPNet and FCN is that in addition to traditional pixel prediction, global features are included pre-fusion. “The local and global cues together make the final prediction more reliable.”

Extra Credit 2: PSPNet

!! Please see the link in the title to help you answer the following questions.

What is the reason for using PPM based on the PSPNet Paper? (Hint: Look into Paper Section 3.2)

Using a PPM incorporates global context information, helpful with multi classification problems. By taking these features into account, the boundaries between multiple objects are more prominent.

What is your IoU score for PSPNet-ResNet50 and FPN-ResNet50?

PSPNet-ResNet50 IoU score is: [0.72816944, 0.7978466, 0.77717984]
FPN-ResNet50 IoU score is: [0.8536674, 0.8288056, 0.85655344]