**Semantic Segmentation (UNet)**

1. **UNet can be viewed as the composition of an encoder and a decoder. Explain the  
   functionality of each module. (10 points)**

The left part and foot of the UNet is called the contracting path, whereas the right part is called the expansive path. Combinedly, and with the help of skip connections, U-nets downsample an input image to learn about its salient features, then to reconstruct the input (like a segmentation mask) utilizing upsampling.

*Encoder (left side or contracting path):* It consists of the repeated application of two convolutions. Each conv is followed by a ReLU and batch normalization. Then a max pooling operation is applied to reduce the spatial dimensions. Again, at each downsampling step, we double the number of feature channels, while we cut in half the spatial dimensions.

*Decoder path (right side or expansive path):* Every step in the expansive path consists of an upsampling of the feature map followed by a transpose convolution, which halves the number of feature channels. We also have a concatenation with the corresponding feature map from the contracting path, and usually a convolutional (each followed by a ReLU). At the final layer, a convolution is used to map each 64-component feature vector to the desired number of classes

In total the network has 23 convolutional layers.

1. **What kind of pooling and upsampling methods are used in the encoder and decoder  
   respectively? (10 points)**

Max pooling operation is used to compute the pooling operation in encoder part of the network and transposed convolution is used to compute the up sampling operation in the decoder part of the network.

1. **In addition to the sequential path that goes through the encoder-decoder modules, there are some skip connections that connects different levels of the encoder and the decoder. Explain the importance of these skip connections. (5 point)**

The function of the skip connections in UNet is the same as that of the skip connections in ResNets and other architectures. It helps to resolve the vanishing gradients problem by efficiently passing the information through the network from the shallower layers to the deeper layers.

If there are no skip connections, then when we trying to upsample the image back to the original resolution it could be difficult to decode the image as it did before. Therefore, the skip connections are crucial to pass features from the encoder to the decoder part of the network to recover spatial information lost during downsampling. Otherwise, UNet degrade it’s overall performance.

1. **What is the loss function that is used to train the model? (5 points)**

They used cross entropy loss function to train the model.

1. **An evaluation metric to assess the performance of segmentation models is Intersection Over Union (IOU). Explain how this metric is computed and what does it show. (10 points)**

The Intersection over Union (IOU) metric is a method for calculating the percentage of overlap between the target mask and the prediction output. IoU is computed by dividing the number of shared pixels between the target and prediction masks by the total number of pixels present in both masks. IOU gives an overlap score core between target and predicted masks.

It can be presented visually to provide better understanding of this metric.

Diagram

Description automatically generated with medium confidence

The intersection () consists of all pixels that appear in both the prediction mask and the ground truth mask, whereas the union () consists of all pixels that appear in either the prediction mask or the target mask.

1. **Train the model on TGS salt segmentation dataset and report the train and test loss. (30 points)**

The train and test loss after training model on TGS salt segmentation dataset is as follows:

Chart, line chart

Description automatically generated

1. **Use the trained model from previous step to do segmentation of some of the test images. Report the results. (30 points)**

The segmentation results for some of the test images are as follows, where the first column represents the test image, and the second and third columns represent the original mask and predicted mask respectively:

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |