Improved Unet

Improved unet is an improvised form of the unet. The network follows a similar architecture to unet. It starts with a context aggregation path encoding abstract representations of the input as we go deeper into the network. This is followed by a localization pathway that recombines these representations with shallower features to precisely localize the structures of interest. Then gradient signals are injected further deep in the network through deep supervision.

Context pathway:

Context modules are used as activations in the context pathway. Each context module consists of – a 3x3 convolution layer, followed by a dropout layer with 0.3 which followed by another 3x3 convolution layer. Each convolution layer uses Leaky ReLU as the activation function for the layer. In the context pathway, the context modules are connected by 3X3 convolution layers with stride 2 hence reducing the resolution of the feature map.

Localization pathway:

Localiaztion pathway takes the ouput of the context pathway whi9ch encode contextual information from low spatial resolution to higher spatial resolution. This is done by using a set of upsampling modules followed by localization modules. An upsample module consists of an upsampling layer with size 2 followed by a 3x3 convolution layer with Leaky ReLU as the activation function. The output of each upsampling module is concatenated with the output of the corresponding level of context aggregation pathway. This is then fed to a localization module which consists of a 3x3 convolution layer followed by a 1x1 convolution layer with Leaky ReLU activation function used for both the layers. This combination of an upsampling module, concatenation of corresponding ouput of the context pathway and localization module is repeated 4 times but the last set contains a 3x3 convolution layer in the place of a localization module.

Deep supervision:

Deep supervision is used in the localization pathway by using segmentation layers. These segmentation layers are combined using element wise summation of the outputs of localization modules and finally this is followed by an output layer with 1 filter as we have binary classification of pixels.

Therefore in an Improved Unet there is a context pathway that encodes abstract features followed by a localization pathway that recombines the outputs of context pathway at various levels and finally deep supervision with segmentation layers.

Visual Architecture with modules labelled:

