- There is a trade-off between localization accuracy and the use of context. Larger patches require more max-pooling layers (because larger the image, more impact of translational invariance) that reduce the localization accuracy, while small patches allow the network to see only little context.
- Cropped output features of each convolutional block are concatenated with upsampled output features of next convolutional block, thereby increasing resolution of the output.
- In the upsampling part we have also a large number of feature channels, which allow the network to propagate context information to higher resolution layers.
- To predict the pixels in the border region of the image, the missing context is extrapolated by mirroring the input image.
- The network does not have any fully connected layers and only uses the valid part of each convolution, i.e., the segmentation map only contains the pixels, for which the full context is available in the input image.(It classifies only the pixels which are in the centre portion)
- To compensate for limited training data in biomedical field, excessive elastic deformations which are common in that field.
- To tackle touching objects, higher weight is given to loss of background labels surrounding the touching objects.
- Last layer is 1x1 conv to reduce no. of channels to no. of classes.
- Loss is weighted average of cross entropy of log probabilities which are softmax outputs.
- The weight map is formed to tackle class imbalance and to learn hard examples of border backgrounds.
- Ideally the initial weights should be adapted such that each feature map in the network has approximately unit variance.