The paper presents a fully connected convolutional net for image segmentation. The inference is made on pixel-to-pixel on segmentation. This method enables to localize each pixel to a object from supervised training. Local information will be captured and integrated into global information.

A pooling is applied to find the local high responsive neuron and subsample the input. Tradition CNN produces non spatial outputs and reduces the dimensions of the output. To make FCN, the coarsen output needs to be upsampled to correspond to each pixel in the input.

As the coarsen output is stitched together for upsampling, filter needs to be enlarged to reproduce the corresponding inputs. The implementation is done by using deconvolution with output stride. High layer information combines lower layer information to get more detailed information, which is called skip net. It makes more sense intuitionally as human may infer a object by its contour or shape, which is captured by coarse layer, while the details of object are captured by finer layer. The method is also not complicated in mathematic.

Fine-tuning is done by back propagation to train the network. Training patches are selected randomly to achieve faster convergence. The method has striation in receptive fields size and filter sizer. Both of them can be data driven. The idea is to balance between the global and local information. Also the size of kernel should not be too large so that over interpolation can be avoid in deconvolution.