## **Research paper-4: U-Net Convolutional Networks for Biomedical Image Segmentation**

Source: <https://arxiv.org/pdf/1505.04597.pdf>

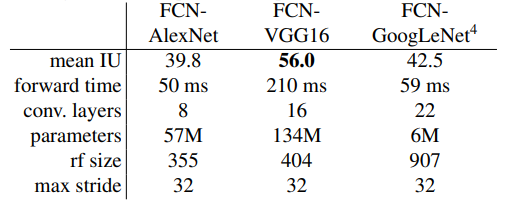
## **Summary**

Introduction:

* Deep convolutional networks have outperformed the state of the art in many visual recognition tasks however, their success has been limited to the availability of large training sets and the size of the considered networks.
* The author proposes a U-Net network and training strategy that relies on the strong use of data augmentation to use the available few samples more efficiently.
* The architecture consists of a contracting path to capture context and an expanding path that enables precise localization that can be trained end-to-end.
* The proposed network is fast in segmentation tasks and has won the ISBI cell tracking challenge by outperforming with a large margin.
* Unet is basically designed for biomedical imaging where output should include localization, i.e., a class label is supposed to be assigned to each pixel.

Related Work:

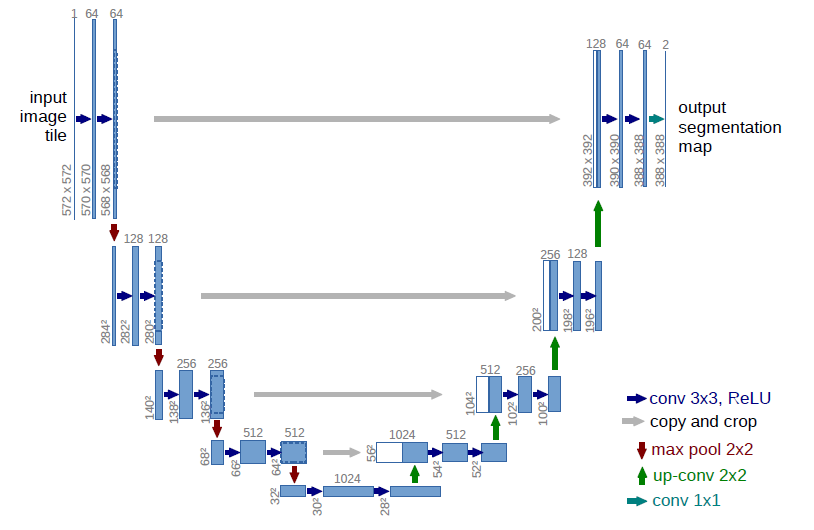
* [Deep neural networks segment neuronal membranes in electron microscopy images](https://papers.nips.cc/paper/4741-deep-neural-networks-segment-neuronal-membranes-in-electron-microscopy-images.pdf) by Ciresan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber where a network trained to predict the class label of each pixel by providing a local region (patch) around that pixel as input and which won the EM segmentation challenge at ISBI 2012 by a large margin.
* [Fully convolutional networks for semantic segmentation](https://arxiv.org/pdf/1411.4038.pdf) by Long, J., Shelhamer, E., Darrell, T (2014) where the state of the art Fully convolutional networks are trained for Image segmentation with different datasets.



Results for the validation set of PASCAL VOC 2011

**U-Net Architecture**

The below image represents the U-Net architecture which is built based on a fully convolutional network where It consists of a contracting path (left side) and an expansive path (right side) with various operations applied throughout both paths.



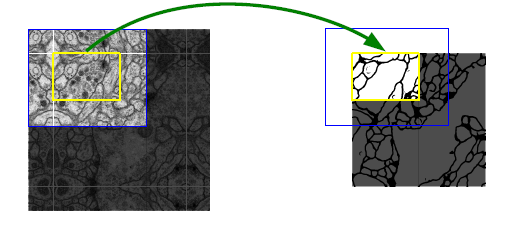
**Implementation Detail:**

The Blue box represents the feature maps that are obtained after an operation is applied and all operations are represented by different color arrow marks, The numbers of filters or depth of the feature maps are represented at the top of the blue box, resolution of the feature map is represented at the bottom left of each blue box, the White box is the feature maps that are copied from the adjacent block for concatenation for further operation.

* The implementation consists of a contracting path and an expansive path.
* The contracting path (Encoder) follows the typical architecture of a convolutional network, It consists of the repeated two 3x3 unpadded convolutions followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling.
* The Encoder is for downsampling where the number of feature map channels is doubled for each successive block.
* The expansive path (Decoder) consists of 2x2 up-convolution that is applied to feature maps that have a cropped feature concatenated with it from contraction path followed by two 3x3 convolutions, each with ReLU.
* The Decoder is for upsampling where the number of feature map channels is reduced by a factor of 2 for each successive block.
* The Final layer has a 1x1 convolution that is used to map each 64 component feature vector to the desired number of classes and network has a total of 23 convolutional layers.

Training:

* Use Overlap-tile strategy to train where large images are divided into patches for seamless segmentation



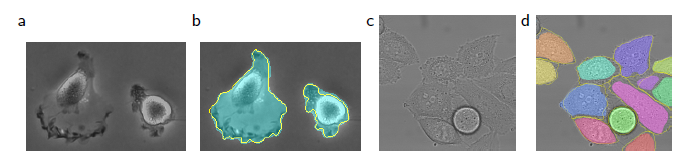
* Prediction of the segmentation in the yellow area requires image data within the blue area as input. Missing input data is extrapolated by mirroring.
* The input images and their corresponding segmentation maps are with the stochastic gradient descent implementation of Caffe and favor large input tiles over a large batch size and hence reduce the batch to a single image.
* The energy function is computed by a pixel-wise soft-max over the final feature map combined with the cross-entropy loss function.



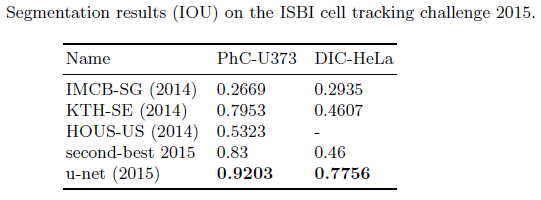
* Weight initialization is done from Gaussian distribution with a standard deviation of p square\_root(2/N), where N denotes the number of incoming nodes of one neuron.
* Data augmentation is used which essential to teach the network the desired invariance and robustness properties when only a few training samples are available.
* Data augmentation primarily involves shift and rotation invariance as well as random elastic deformations and gray value variations to train a segmentation network with very few annotated images.

Experiments:

This segmentation task is part of the ISBI cell tracking challenge and the experiment contains U-Net applied to a cell segmentation task in light microscopic images.



(a) part of an input image of the ‘PhC-U373" data set. (b) Segmentation result with manual ground truth (yellow border) (c) input image of the ‘DIC-HeLa" data set. (d) Segmentation result (random colored masks) with manual ground truth (yellow border).



The results of two different datasets with IOU Score shown above where Unet out-perform every other model by a great margin.

**Conclusion:**

* Unet is basically designed for biomedical imaging which is built to work with a small number of data samples with data augmentation.
* The U-Net combines the information from the downsampling path with the contextual information in the upsampling path to finally obtain general information.
* This tiling strategy helps to apply the U-Net network to large resolution images where GPU memory is limited for segmentation.
* The U-net architecture achieves very good performance on various different biomedical segmentation and suitable for biomedical applications.