

# Notes - Fully Convolutional Network (FCN)

Wednesday, December 12, 2018 10:32 PM

## What is FCN?

It is a machine learning algorithm used for the computer vision application especially semantic segmentation.

## What is the purpose?

The purpose of FCN is to segment the images and assign pixels belonging to an object to a specific class label.

## What is the advantage of FCN?

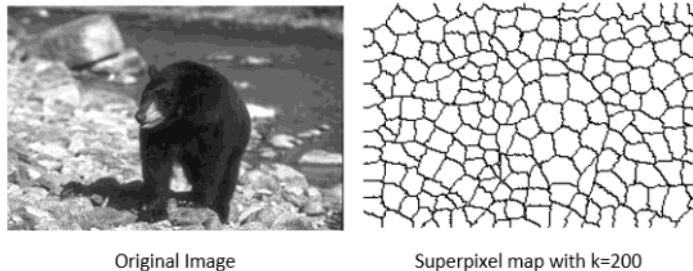
When compared to the approaches that were proposed prior to FCN, FCN exceeds the performance of the state of the art algorithm and can be trained end to end.

## What is the common pipeline of previous approaches?

Previous approaches have used pre-processing and post-processing. Pre-processing includes superpixels, proposals and post-processing includes refinement by random-fields or local classifiers.

## What is superpixel? (One of the techniques used by previous approach)

Superpixel is a perceptually meaningful entity that can be obtained from a low-level grouping process. Normalized cuts is a classical region segmentation algorithm which used spectral clustering to exploit pairwise brightness, color and texture affinities between pixels. This algorithm can be applied to oversegment images to obtain superpixels. Following example shows such a superpixel map with the number of clusters 200.



## How Fully Convolutional network uses the CNN architecture for semantic segmentation?

Deep feature hierarchies created by Convolutional Neural Network encode location and semantics in a nonlinear local-to-global pyramid i.e. the deep layers (feature layers near the end) in the CNN can produce the coarse semantic information and the shallow layers (initial feature layers) provide the fine, appearance related information. These two layers are combined using the skip architecture.

## What are the previous approaches related to FCN?

The approach used by FCN combines the successes of deep nets for image classification and transfer learning.

### Fully convolution approach:

This approach is initially used for the extension of convnets to arbitrary sized inputs, where the classic LeNet is used to recognize the string of digits. Here the input is considered as one dimensional strings and Viterbi decoding is used to obtain the results. The approach is also used in a lot of coarse multi-class segmentation in medical applications, sliding window detection, image restoration etc.,

### Dense predictions with convnets:

Several approaches have used the convnets to do the semantic segmentation. Common elements used by them include the following:

1. Having small models and restricting the capacity and receptive fields
2. Patch wise training
3. Post-processing using the super-pixel projection, random field regularization, filtering or local classification
4. Input shifting and output interlacing for dense output
5. Multi-scale pyramid processing
6. Saturating tanh nonlinearities
7. Ensembles

But the FCN does not use any of these machineries. FCN uses patch wise training and shift-and-stitch dense output, in-network up-sampling.

FCN uses fusion architecture that fuses the features across layers to define a nonlinear local-to-global representation which is then tuned end to end.

#### References:

Fully Convolutional Network

Link: [https://people.eecs.berkeley.edu/~jonlong/long\\_shelhamer\\_fcn.pdf](https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf)

Superpixel

Link: <http://ttic.uchicago.edu/~xren/research/superpixel/>