# Fully Convolutional Networks for Semantic Segmentation

Jonathan Long
Evan Shelhamer
Trevor Darrell

# PROBLEM TO BE SOLVED -SEMANTIC SEGMENTATION

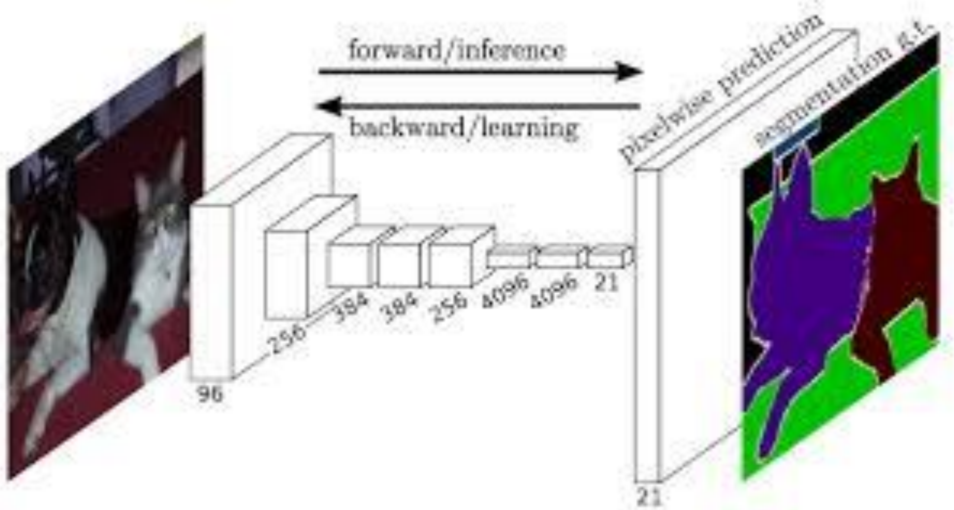
- Segmentation is the partition of an image into several "coherent" parts, but without any attempt at understanding what these parts represent.
- Coherence can be defined in terms of low-level cues such as color, texture and smoothness of boundary.
- Semantic segmentation attempts to partition the image into semantically meaningful parts, and to classify each part into one of the pre-determined classes.
- Semantic segmentation is understanding an image at the pixel level i.e we want to assign each pixel in the image to an object class.

Apart from recognizing the bike and the person riding it, we also have to delineate the boundaries of each object. Therefore, unlike classification we need dense pixel-wise predictions from our models.

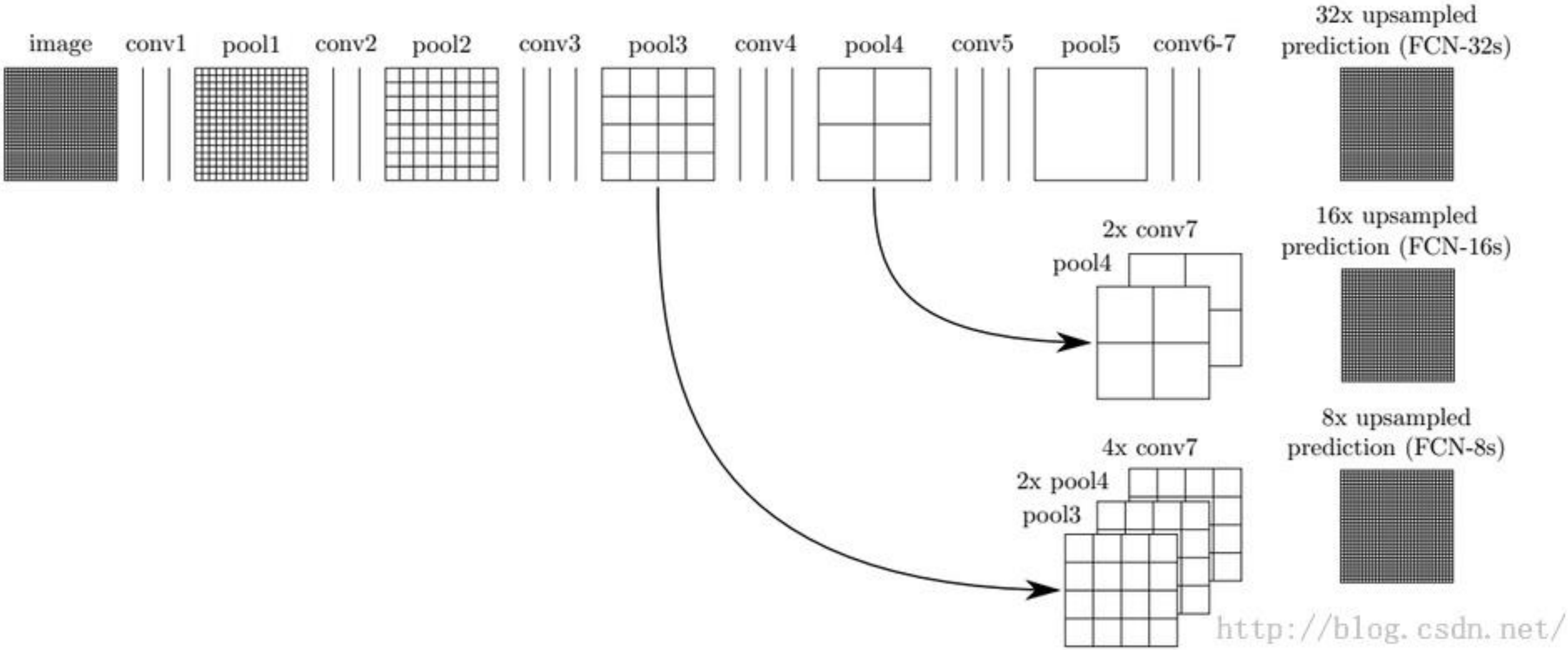
# Overview

- Reinterpret standard classification convnets as “Fully convolutional” networks (FCN) for semantic segmentation and replaces the fully connected layers with convolutional layers.
- Uses a paradigm of transfer learning by using a network trained on ImageNet for feature learning.
- Deep learning approach : combine information from different layers for segmentation
- State-of-the-art segmentation for PASCAL VOC 2011/2012, NYUDv2, and SIFT Flow at the time
- simple and efficient both asymptotically and absolutely. No pre and post processing techniques are needed for this method.
- Inference less than one fifth of a second for a typical image

# FCNN Workflow



# FCNN Architecture



# Architecture Explained

- Firstly ImageNet trained model like AlexNet is taken and the fully connected layers are replaced with convolution layers.
- Then a series of deconvolution layers are used to up sample the image as we need to classify every pixel.
- Skip layers which are indicated by arrows in the above image are used to transfer the information captured in the initial layers and was required for re construction during the upsampling.
- Due to absence of fully connected layers it can take input of any arbitrary size and produce the corresponding output.

# IMPLEMENTATION

- Will take an ImageNet trained model to apply the paradigm of transfer learning. As the model is trained on large data it will save computation and also serves as a good feature identifier.
- Then the fully connected layers are removed and replaced with convolution layers which will be followed by a series of deconvolution layers to upsample the image.
- Skip connections will be applied to model learns about the low level layers while up sampling.
- Then the model will be tuned with the help of PASCAL VOC dataset.

# Tentative Plan

- First focus will be on implementation of the entire architecture and making sure that there is no flaw in any connections or kernel parameters.
- Less no of categories will be considered initially and the model performance will be noted and if it’s not good necessary changes will be made to the model.
- After getting reasonable output, various models like AlexNet, Googlenet will be tried as a pre trained Network to see the performance gain.
- After that model will be trained on large number of categories. Exact number will be based on computation feasibility.

# Conclusion

- Instead of segmenting images in form of bounding boxes, with the help of fully convolution networks we can segment image at pixel level, i.e we gave label to each and every pixel which is called semantic segmentation.

- Use of transfer learning paradigm helps us save a lot of computational cost and improves the efficiency.

- It popularized the use of end to end convolution networks for semantic segmentation. It is now used for contour detection, optical flow and weakly supervised semantic segmentationm, pixel labelling of microscopy images. It’s efficiency In semantic segmentation is greatly improved by using it along with CRFS.