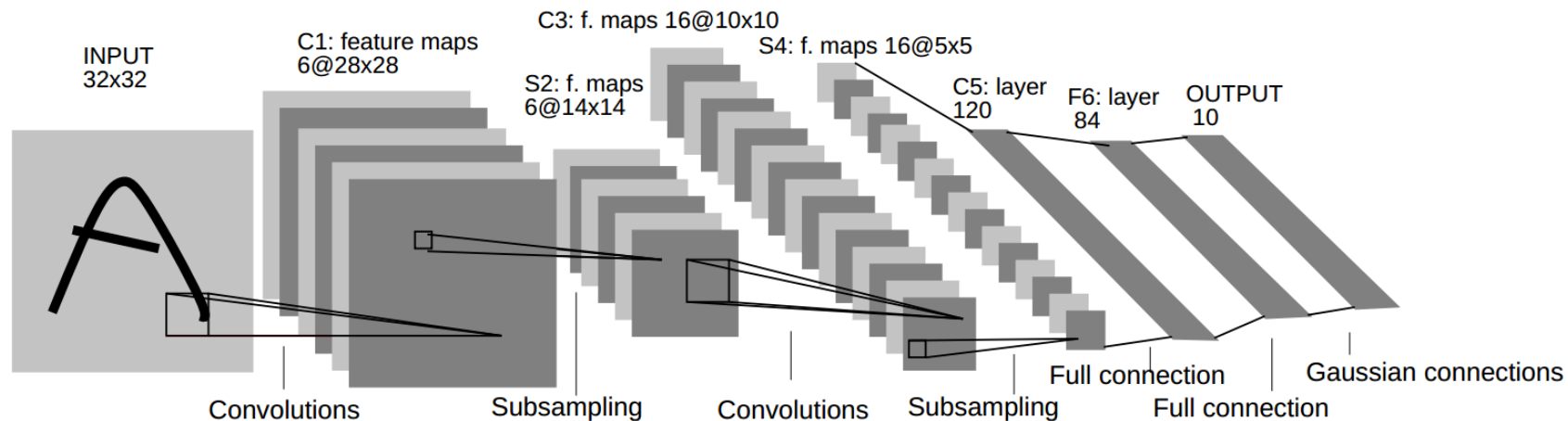


Deep Convolutional Neural Networks – Beyond Classification

Abhishek Sharma

CNNs so far

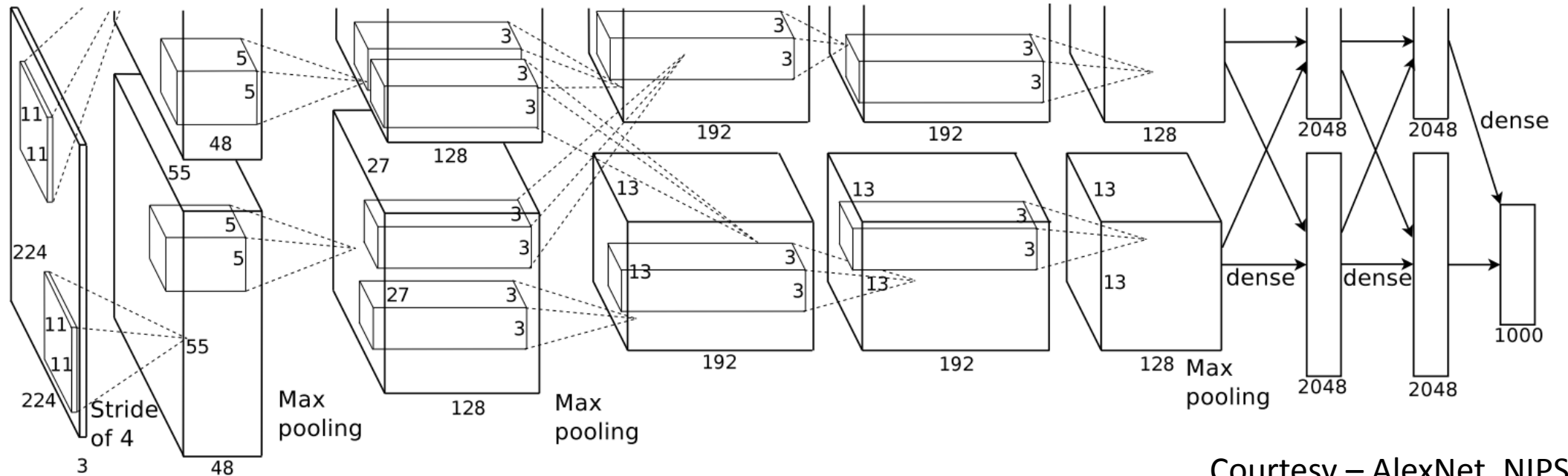
- Handwritten digit classification, 1995ish
 - State-of-the-art recognition accuracy for handwritten digits [0-9], used in automatic check deposit and postal applications
 - Plain CNN structure with 3 stages of CNN + 2 fully connected



Courtesy – Yann Lecun et al.

CNNs so far

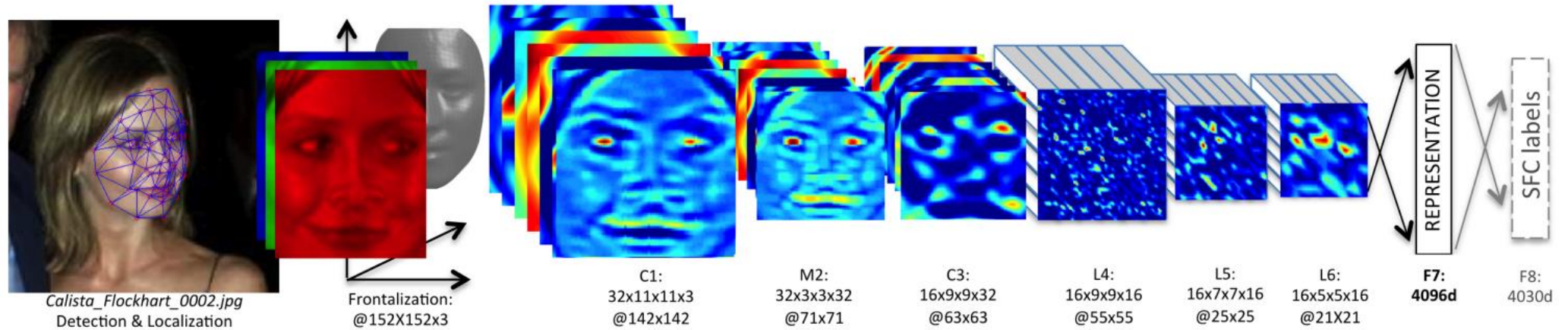
- ImageNet challenge, 2012
 - State-of-the-art accuracy on 1K object categories near 95%
 - AlexNet – 5 layers of CNN + 3 fully-connected layers
 - VGG – 16 layer network with decomposed filters



Courtesy – AlexNet, NIPS 2012

CNNs so far

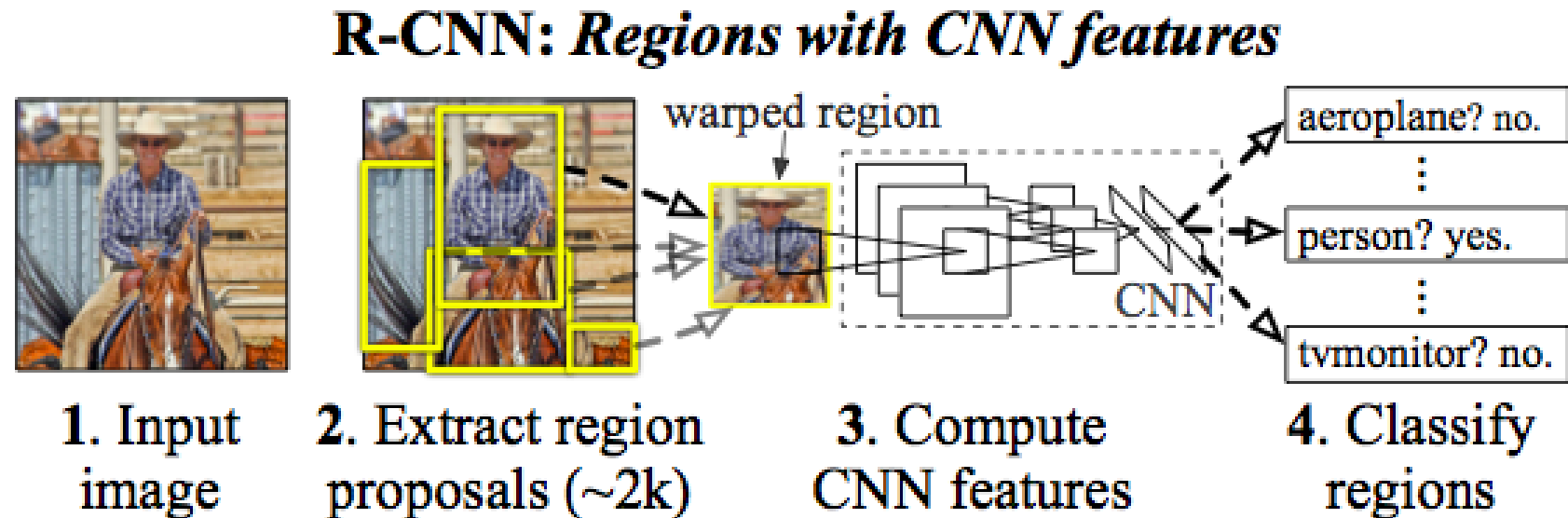
- Deep face 2014
 - Close to human accuracy on face verification (same or different)



Courtesy –
DeepFace, CVPR
2014

What is really important ?

- The convolutional layers are the most important part
- A pre-trained network for ImageNet classification can be used for many different vision tasks.
- Detection



What next ?

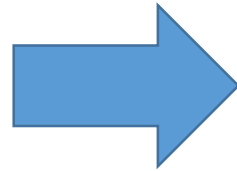
- Segmentation
- Depth Image
- Image Quality Estimation
- Stereo ?
- 3-D reconstruction ?
- Imagine and grab the glory

Semantic segmentation

I



Image



Y

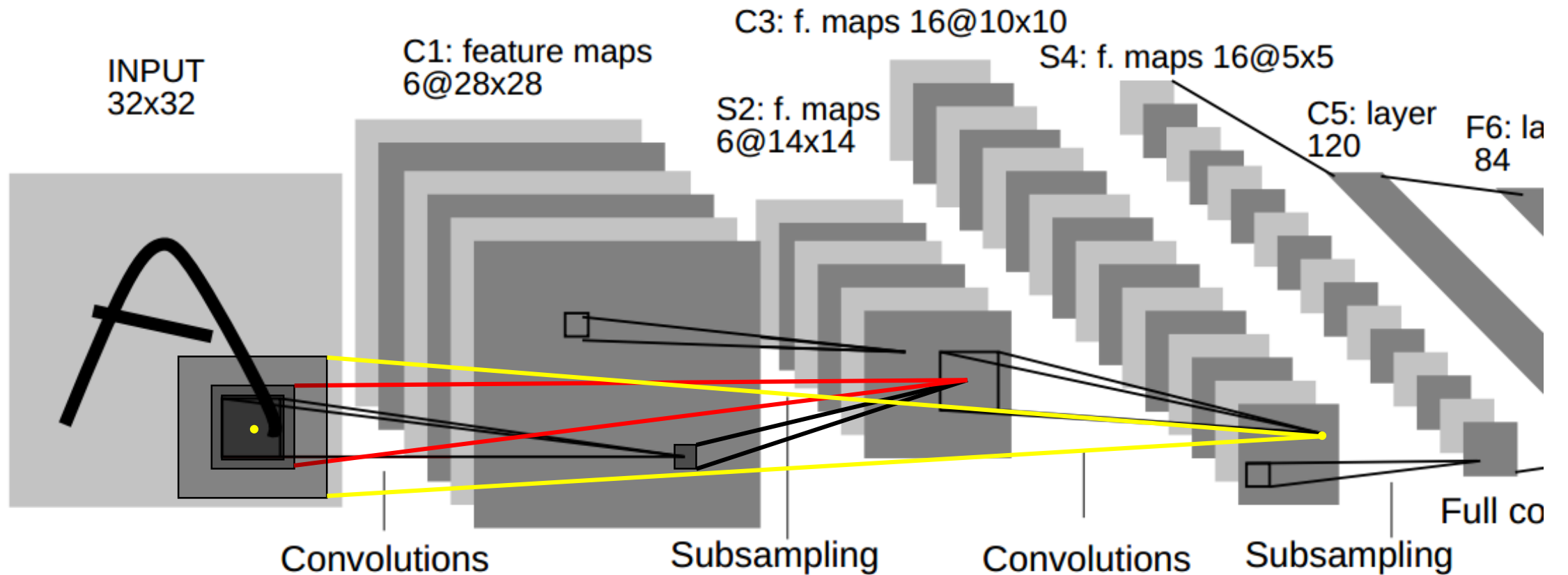


Per-pixel Label

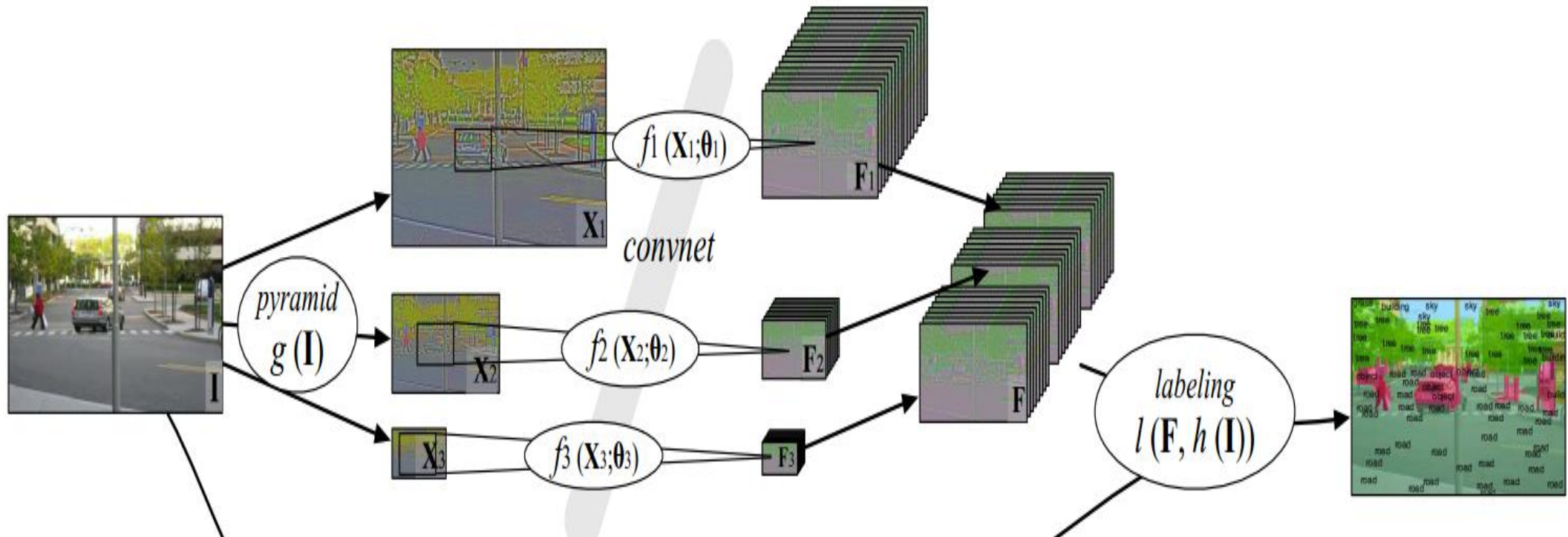
Can CNN be used ?

- Yes
- The state-of-the-art is indeed a CNN based model.
- How ?
 - Use CNN and extract per-pixel features.
 - Treat each output location (i,j) in a CNN cube as a feature for a patch.
 - Match correspondence of output CNN locations to pixels.
 - Use pixel-features to classify each pixel.
- But really how ??

Understanding CNN cubes



Multi-scale CNN for better accuracy



Courtesy – Farabet and Lecun, TPAMI

Pre-training is still better

- Using VGG pre-trained model and simply up-sampling per-pixel classification gives better accuracy than all previous models.

Discussion and informal thoughts

- CNN is actually a very powerful feature learning paradigm and has the potential to be used almost everywhere owing to its trainability and cross-application generalization.
- Think of it as extracting features from a patch and then doing something useful depending on the application.