Semantic Segmentation

Lecture 17
Semantic Segmentation

Assign a label to each pixel in an image:

- Pixel-level image annotation/analysis (vs. object-level analysis)

Common datasets: PASCAL VOC (2012) and MSCOCO
Semantic Segmentation

- A key part of Scene Understanding

- Applications
  - Autonomous navigation
  - Assisting the partially sighted
  - Medical diagnosis
  - Image editing
Semantic Segmentation

- Applications
  - Autonomous navigation
Semantic Segmentation

- Applications
  - Assisting the partially sighted
Semantic Segmentation

- Applications
  - Medical diagnosis
Semantic Segmentation

- Applications
  - Image editing
Segmentation tasks

A sample image from the PASCAL VOC2011 dataset

Original (input) Image  Instance (object) Segmentation  Semantic (class) Segmentation

**Question:** How about the frame, the fruit (tomato?) and the papers on the table?
Before deep learning

Many techniques... Such as:

• TextonBoost

• TextonForest
  • “Semantic Texton Forests for Image Categorization and Segmentation”, 2008

• Conditional Random Forest based approaches:
Fully Convolutional networks (FCN)

Fully “CONVOLUTIONAL” Networks (FCN)

• Use pre-trained networks for classification for segmentation! (VGG, AlexNet, etc.)

• Re-interpret the fully-connected layers as fully convolutional networks.

• Utilize skip-layer concept to improve the segmentation accuracy.
Fully Convolutional networks (FCN)

Interpret the FC layers as conv layers.
FCN

(21 is the number of classes here!)
FCN

Upsampling

Nearest Neighbor

Input: 2 x 2

Output: 4 x 4
Upsampling

Bi-linear interpolation

\[ P \approx \frac{y_2 - y}{y_2 - y_1} \left( \frac{x_2 - x}{x_2 - x_1} Q_{11} + \frac{x - x_1}{x_2 - x_1} Q_{21} \right) + \frac{y - y_1}{y_2 - y_1} \left( \frac{x_2 - x}{x_2 - x_1} Q_{12} + \frac{x - x_1}{x_2 - x_1} Q_{22} \right) \]
Upsampling

Max Pooling
Remember which element was max!

1 2 6 3
3 5 2 1
1 2 2 1
7 3 4 8

Input: 4 x 4

Output: 2 x 2

Max Unpooling
Use positions from pooling layer

0 0 2 0
0 1 0 0
0 0 0 0
3 0 0 4

Input: 2 x 2

Output: 4 x 4

Rest of the network

Corresponding pairs of downampling and upsampling layers
Deconvolution
Deconvolution Network for Semantic Segmentation

Image source: H. Noh, S. Hong, and B. Han, Learning Deconvolution Network for Semantic Segmentation, ICCV 2015
Learned upsampling architectures

Encoder-decoder

U-Net
SegNet

Uses VGG architecture!
No FC layer!

Image source: http://mi.eng.cam.ac.uk/projects/segnet/
Questions?

Sources for this lecture include materials from works by Sedat Ozer, Ulas Bagci, and Svetlana Lazebnik