CAP5415 Computer Vision

Yogesh S Rawat

yogesh@ucf.edu

HEC-241

Final exam

When?

- 12/02/2024 from 1:00PM - 2:00PM

Where?

- Same as mid-term

Notes:

- The questions pattern will be same as mid-term.
- It will count towards 20% of your total grade.
- Syllabus - Lecture 9: SIFT - Lecture 16: Instance segmentation
- Zoom link: will be provided.
- Video should be on for all students.
- Please keep your mic on mute. You can use mic to ask questions.

Reminder

- Student evaluation of course instruction (SPI)
  - Now available on Webcourse
  - Will be available for the next ~15 days
- Online questionnaire includes
  - ONLY eleven (11) questions
  - 9 multiple choice and 2 open comment questions

Questions?

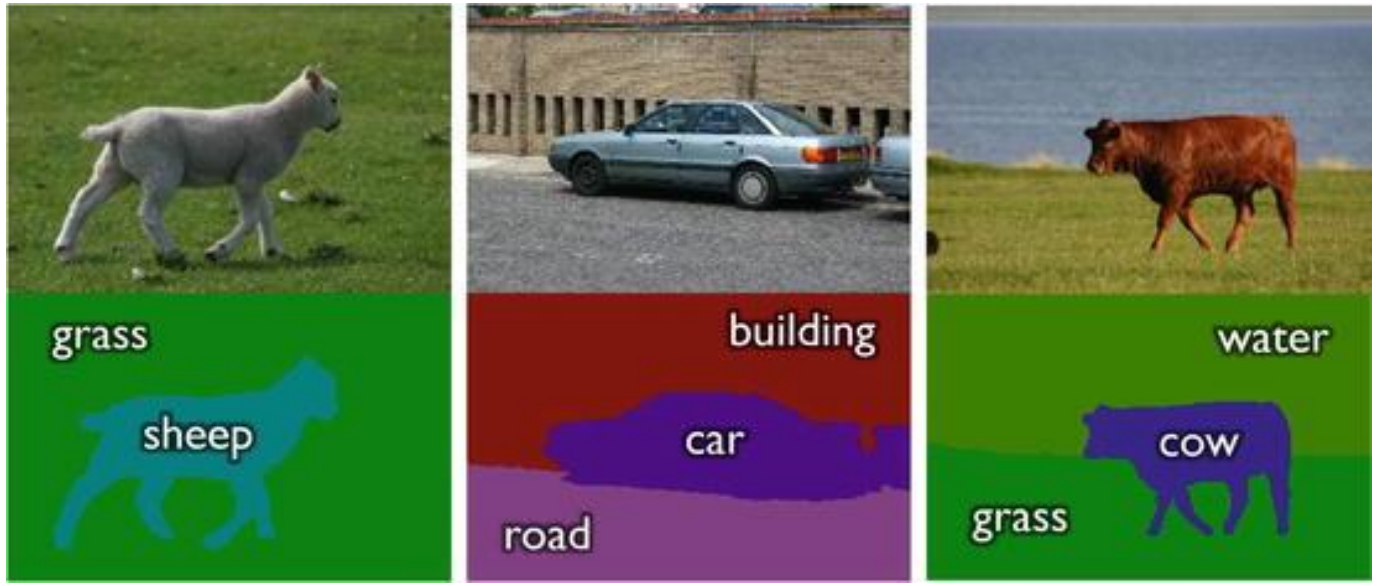
Semantic Segmentation

Lecture 15

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

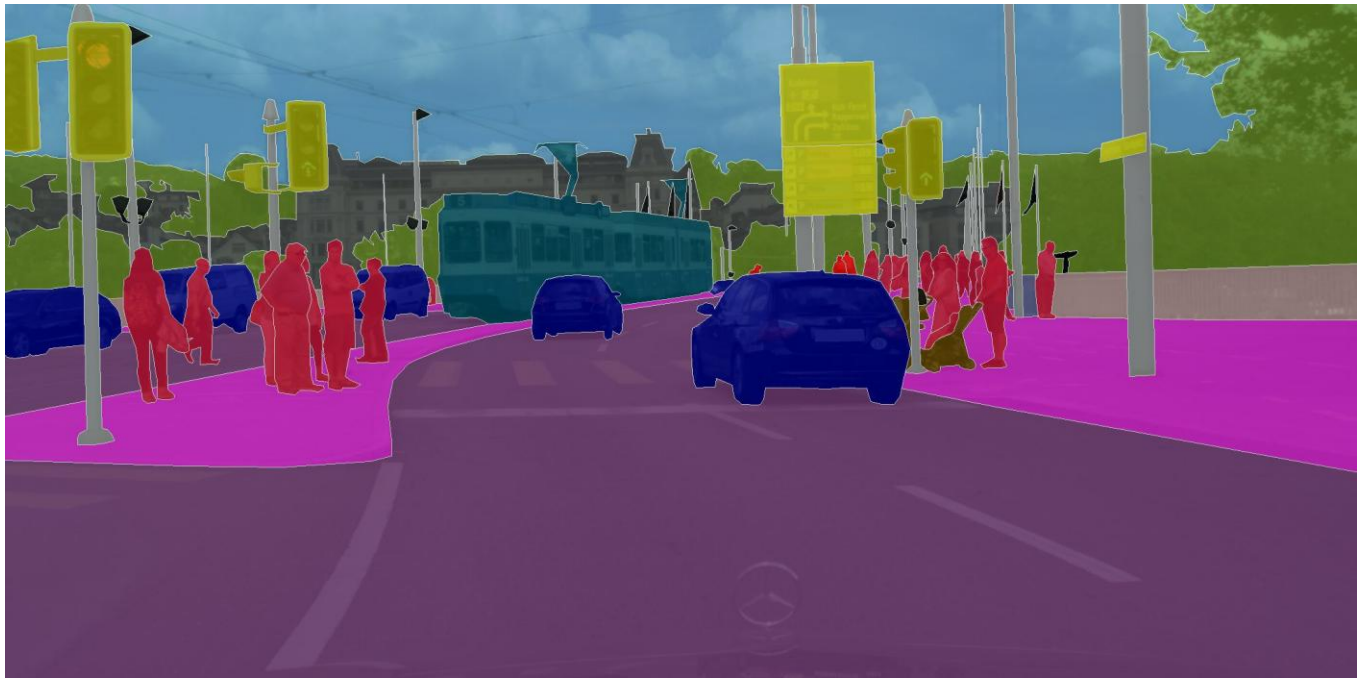
Semantic Segmentation

- A key part of Scene Understanding
- Applications
  - Autonomous navigation
  - Assisting the partially sighted
  - Medical diagnosis
  - Image editing



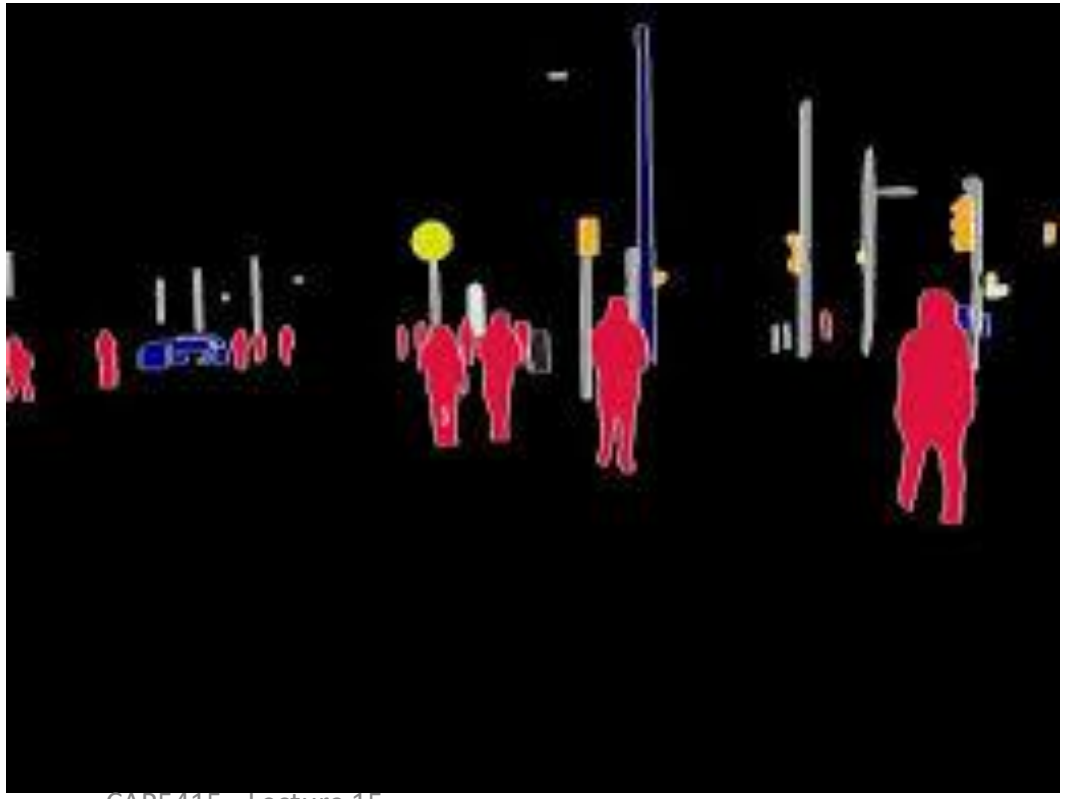
Semantic Segmentation

- Applications
  - Autonomous navigation



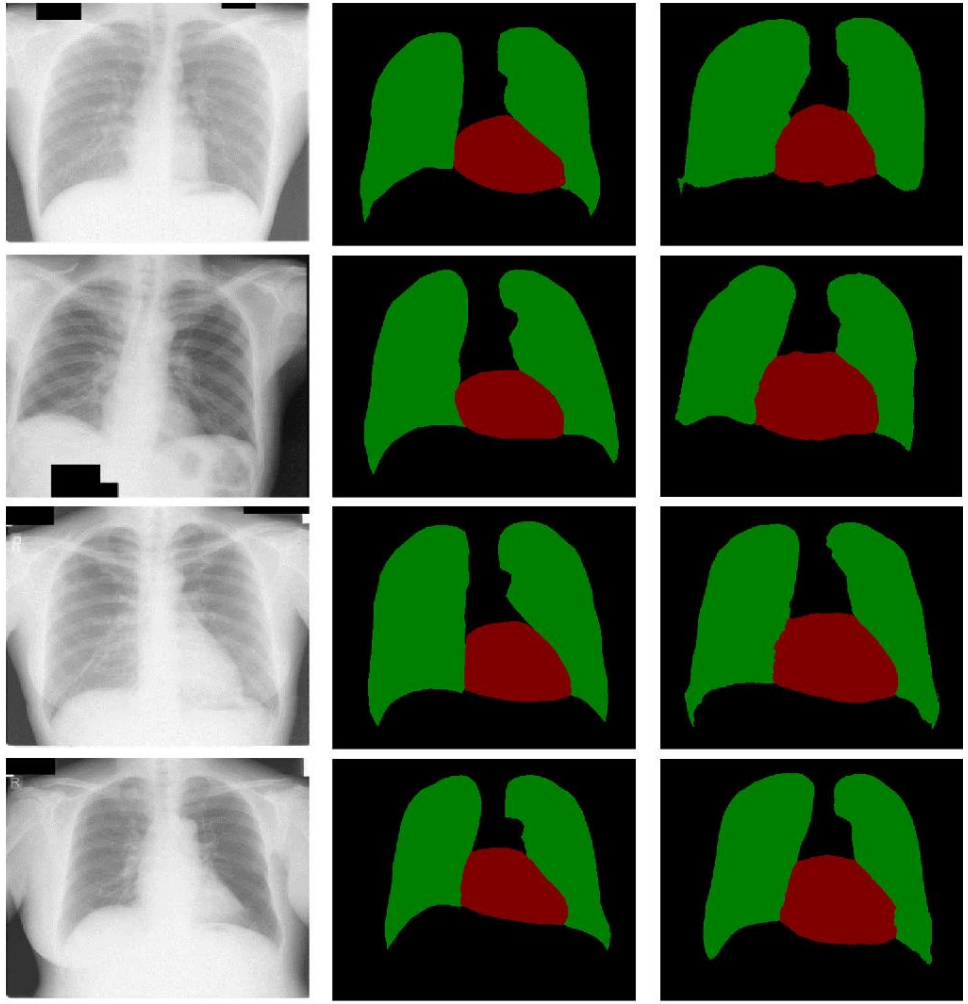
Semantic Segmentation

- Applications
  - Assisting the partially sighted



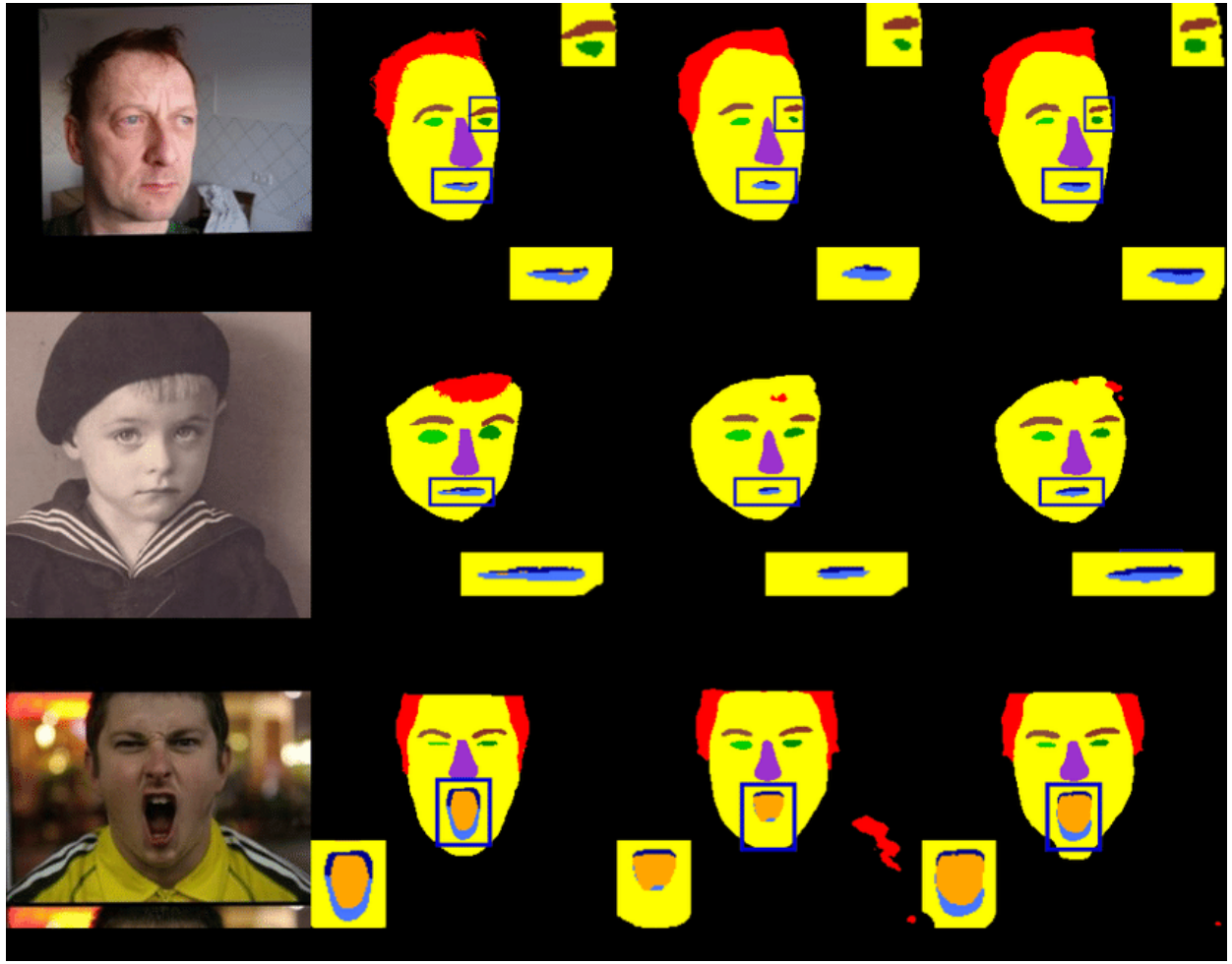
Semantic Segmentation

- 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?

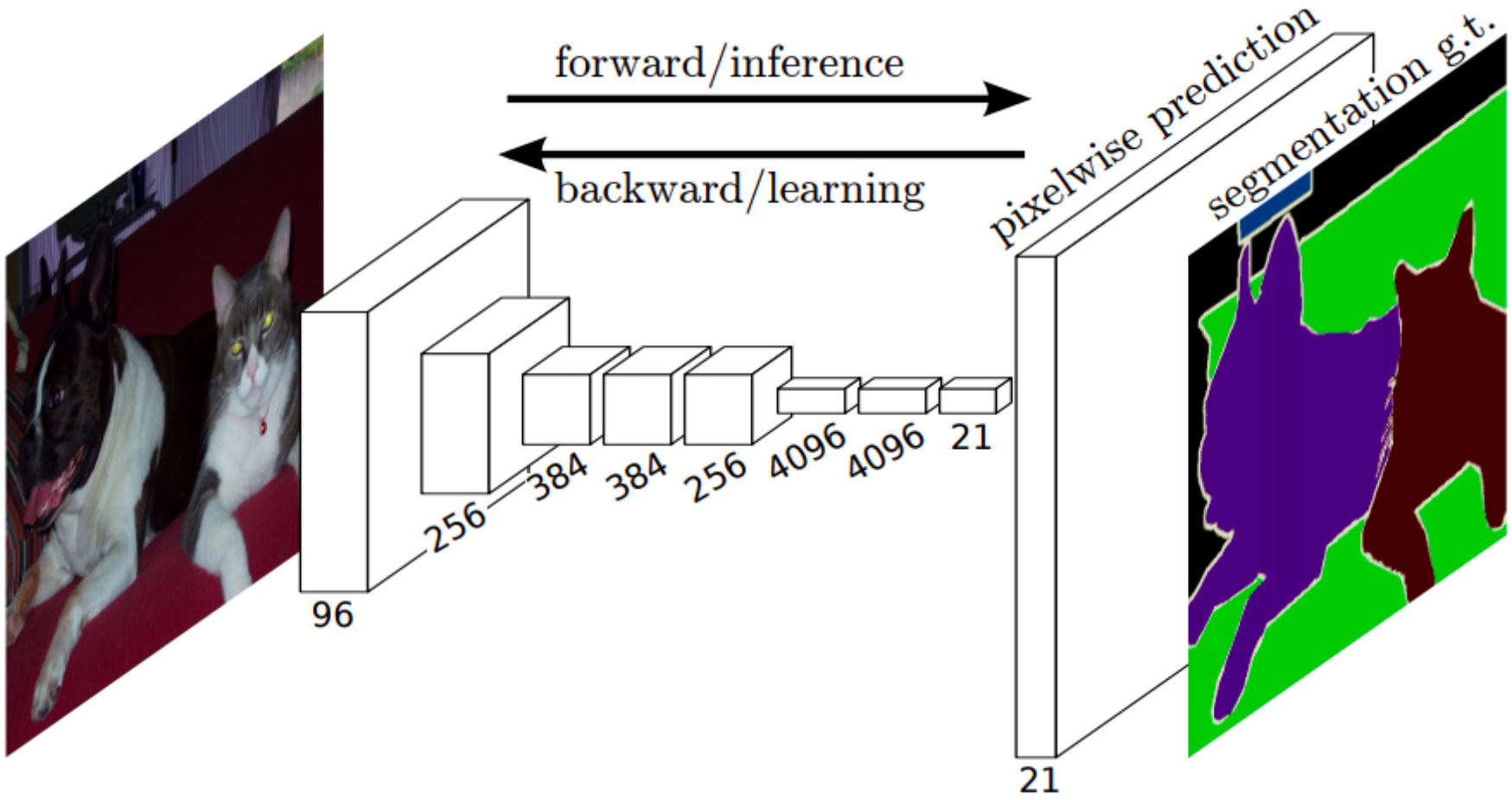
Image Source: http://host.robots.ox.ac.uk/pascal/VOC/voc2012/segexamples/index.html

Before deep learning

Many techniques… Such as:

- TextonBoost
  - J. Shotton, J. Winn, C. Rother, and A. Criminisi, TextonBoost: Joint Appearance, Shape And Context Modeling For Multi-class Object Recognition And Segmentation, ECCV 2006.
- TextonForest
  - "Semantic Texton Forests for Image Categorization and Segmentation", 2008
- Conditional Random Forest based approaches:
  - SuperParsing (J. Tighe and S. Lazebnik, SuperParsing: Scalable Nonparametric Image Parsing with Superpixels, ECCV 2010)

Fully Convolutional networks (FCN)

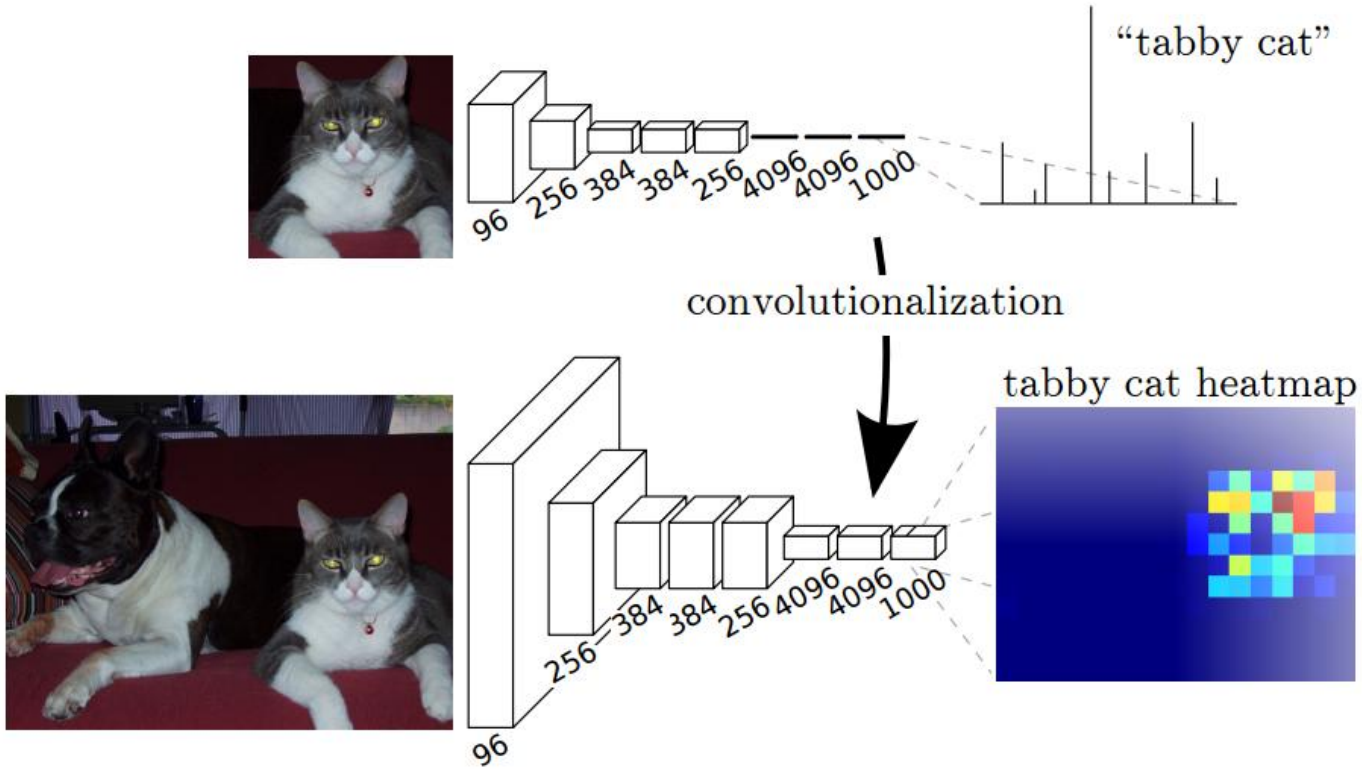


J. Long, E. Shelhamer, and T. Darrell, Fully Convolutional Networks for Semantic Segmentation, CVPR 2015

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.

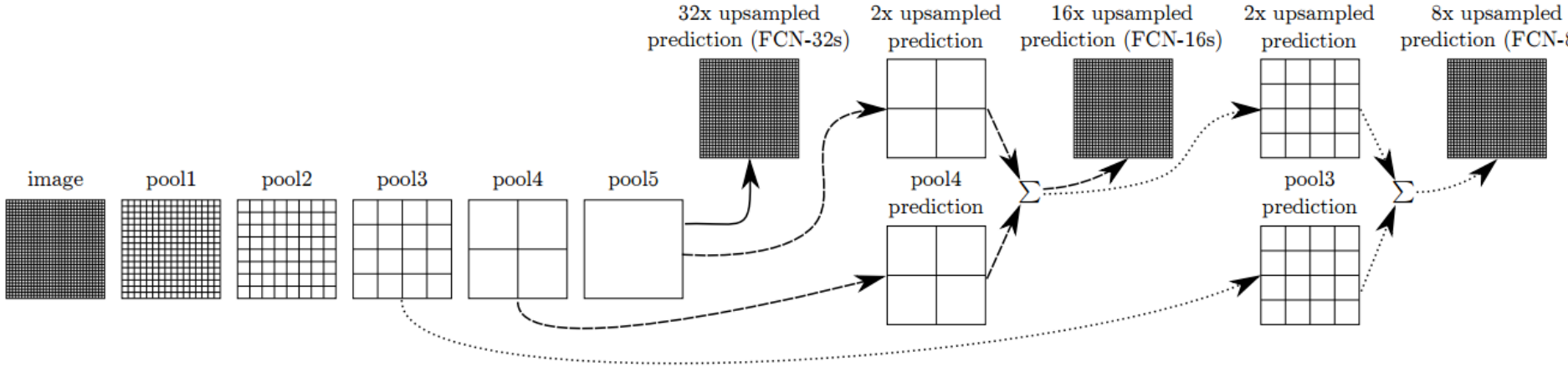
Fully Convolutional networks (FCN)

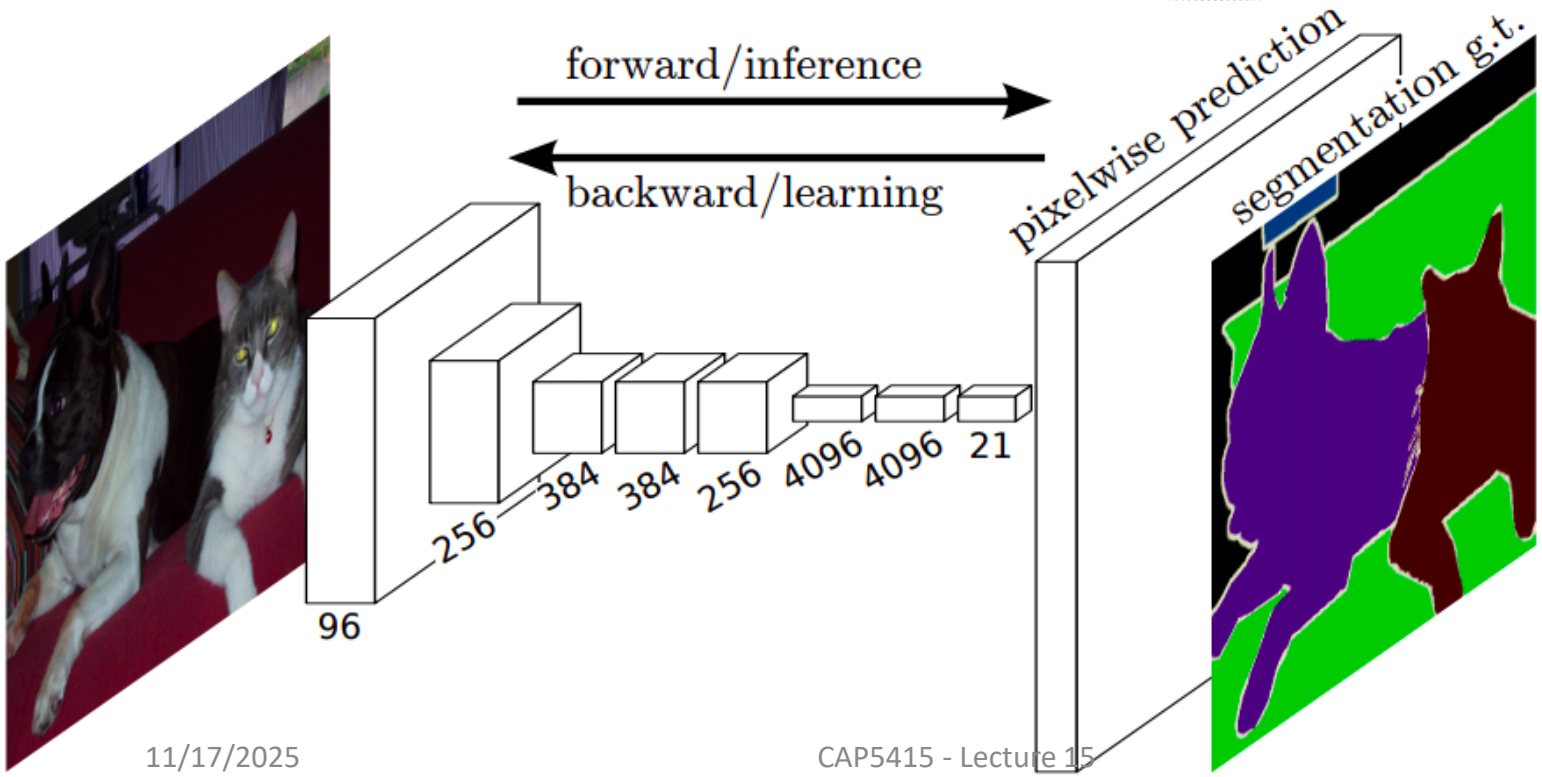


Interpret the FC layers as conv layers.

J. Long, E. Shelhamer, and T. Darrell, Fully Convolutional Networks for Semantic Segmentation, CVPR 2015

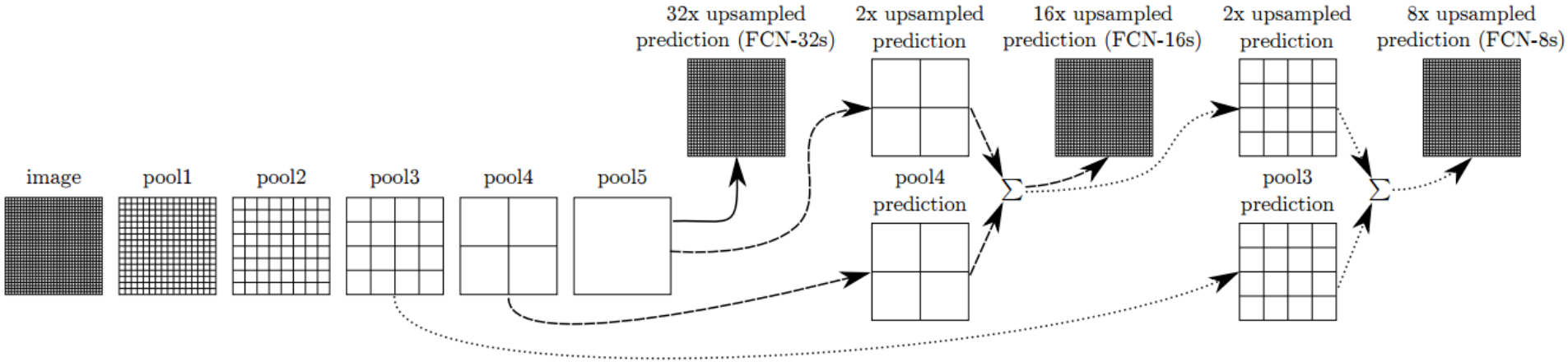
FCN

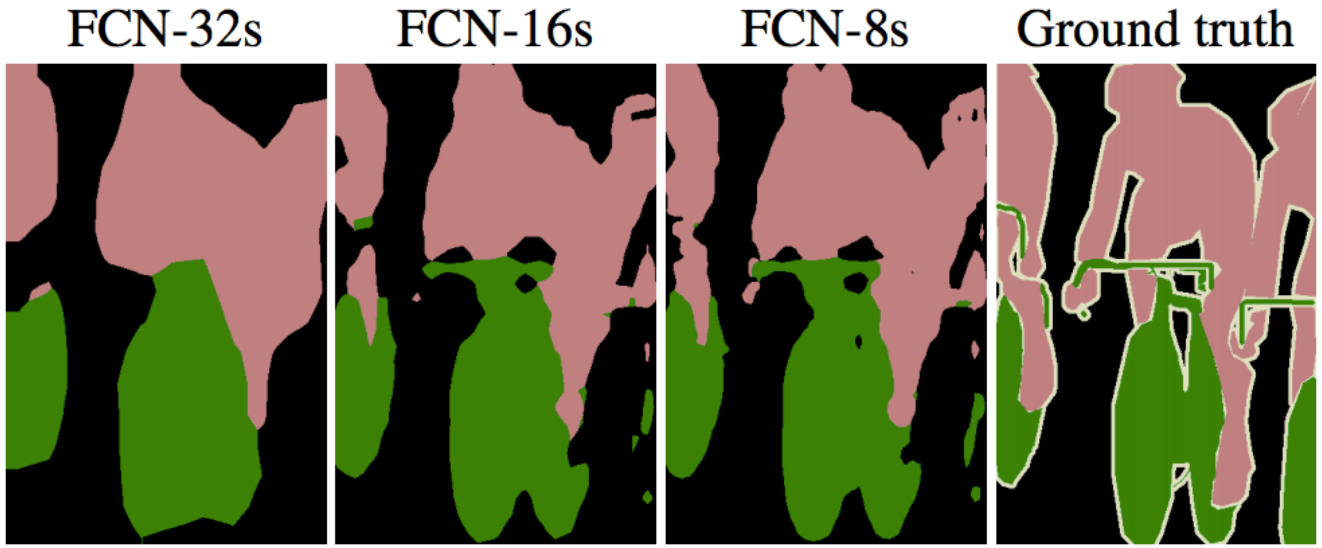




(21 is the number of classes here!)

FCN

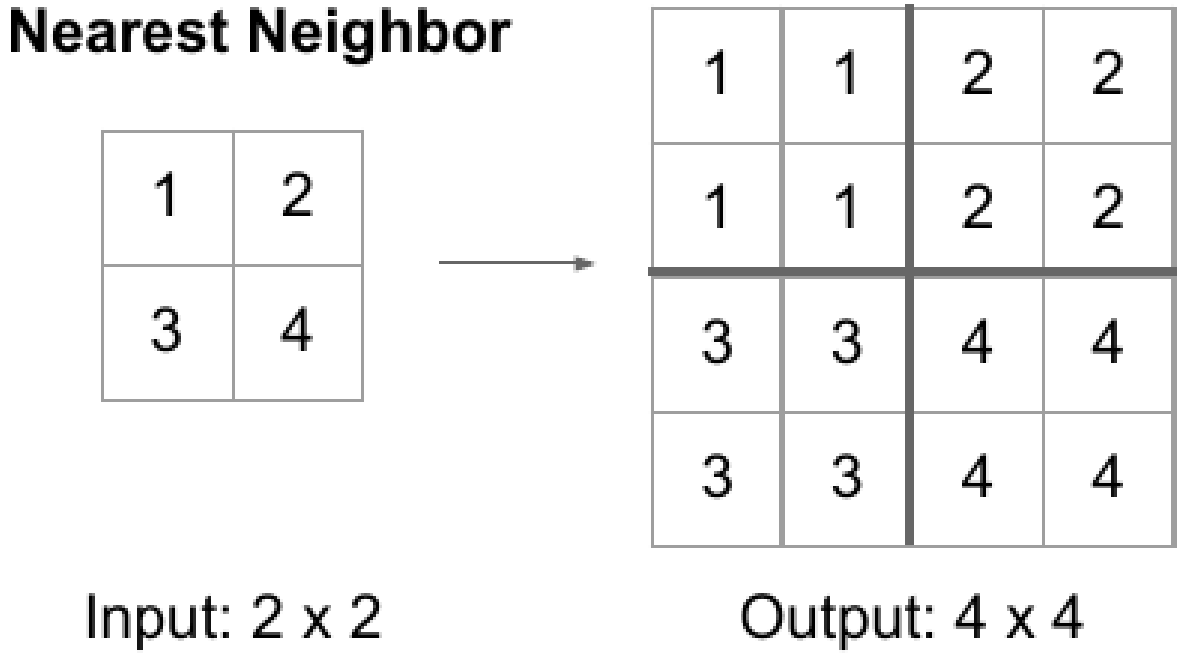




No skip connection | 1-skip connection | 2-skip connections

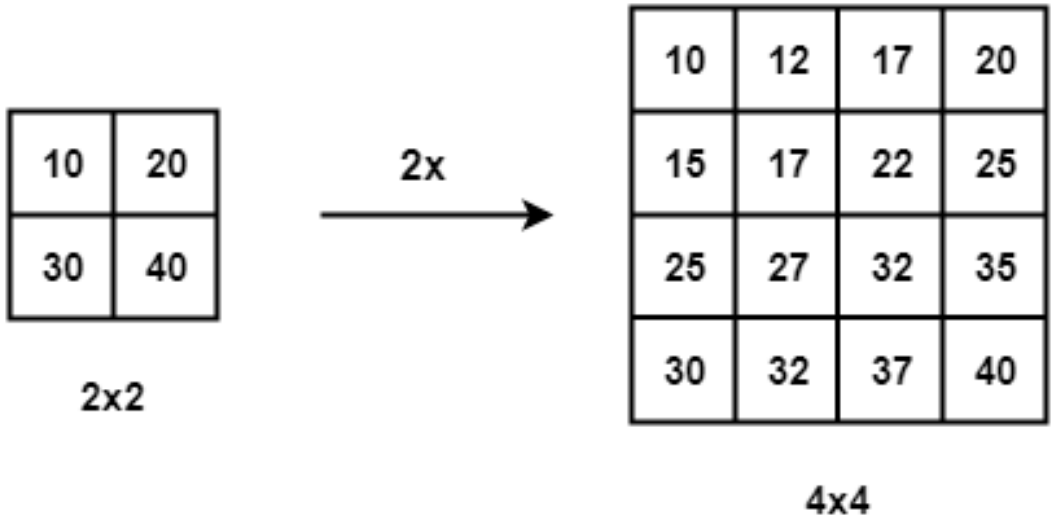
J. Long, E. Shelhamer, and T. Darrell, Fully Convolutional Networks for Semantic Segmentation, CVPR 2015

Upsampling


Nearest Neighbor
1 2
3 4
1 1 2 2
1 1 2 2
3 3 4 4
3 3 4 4
Input: 2 x 2
Output: 4 x 4

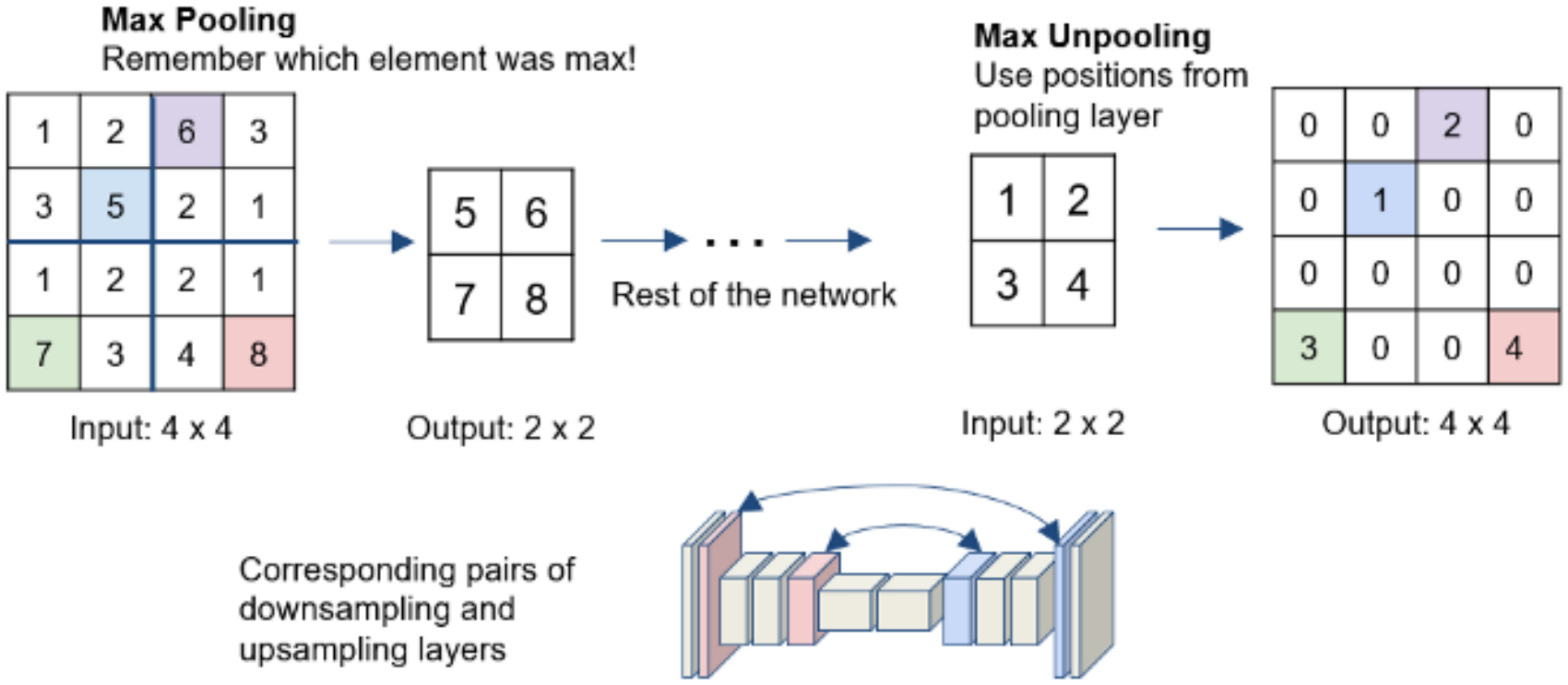
Upsampling

Bi-linear interpolation

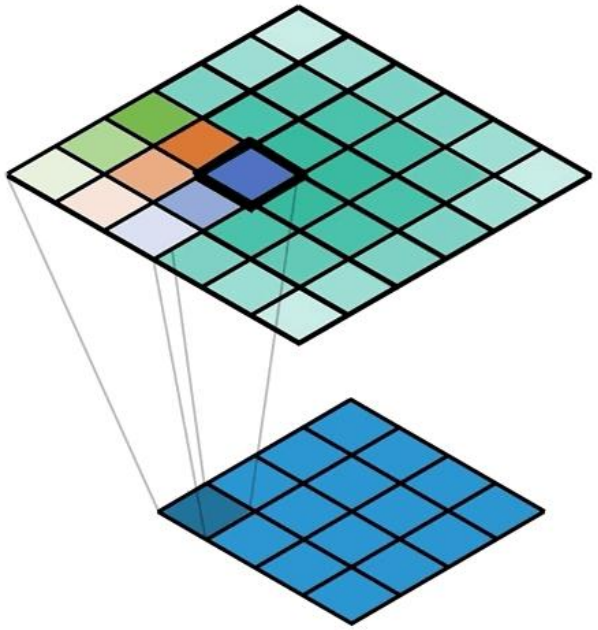


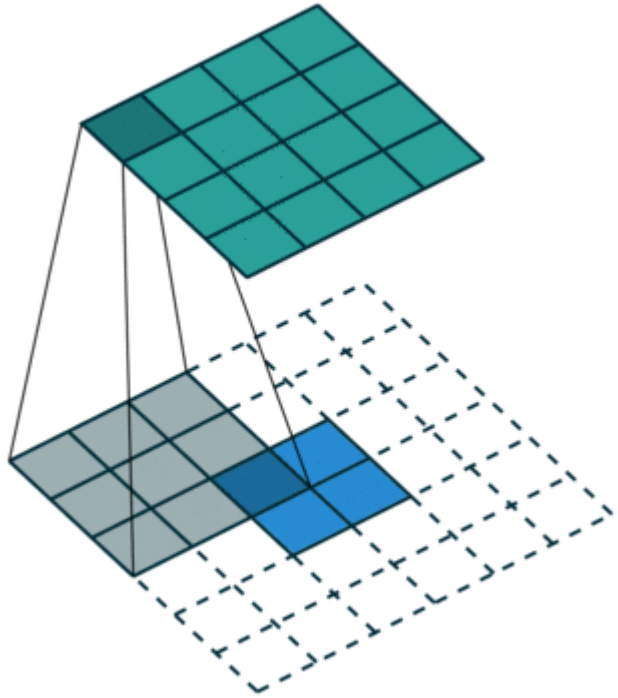
$$P \approx \frac{y_2 - y}{y_2 - y_1}\left(\frac{x_2 - x}{x_2 - x_1}Q_{11} + \frac{x - x_1}{x_2 - x_1}Q_{21}\right) + \frac{y - y_1}{y_2 - y_1}\left(\frac{x_2 - x}{x_2 - x_1}Q_{12} + \frac{x - x_1}{x_2 - x_1}Q_{22}\right)$$

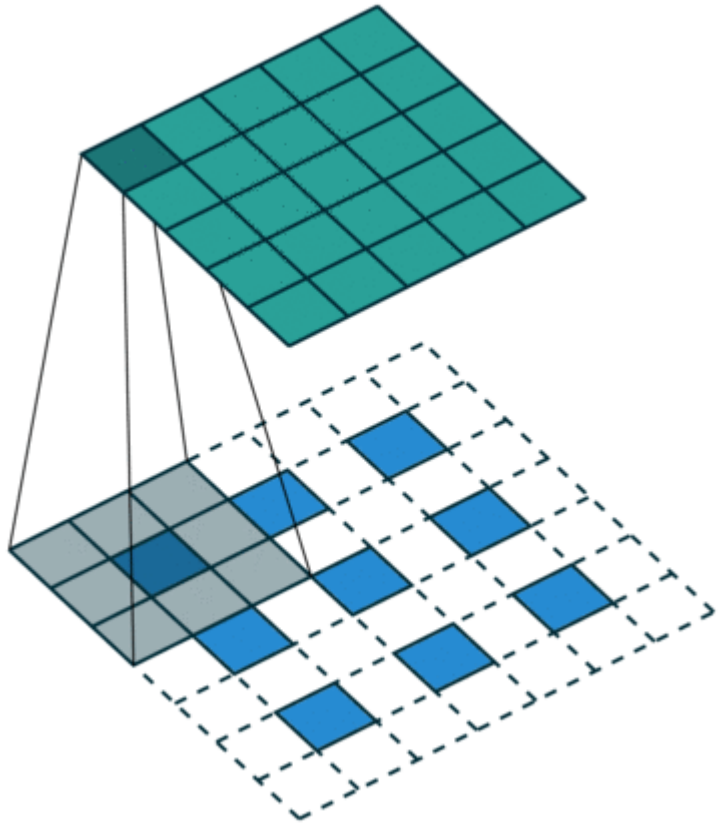
Upsampling


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
5 6
7 8
Output: 2 x 2
Rest of the network
Max Unpooling
Use positions from pooling layer
1 2
3 4
Input: 2 x 2
0 0 2 0
0 1 0 0
0 0 0 0
3 0 0 4
Output: 4 x 4
Corresponding pairs of downsampling and upsampling layers

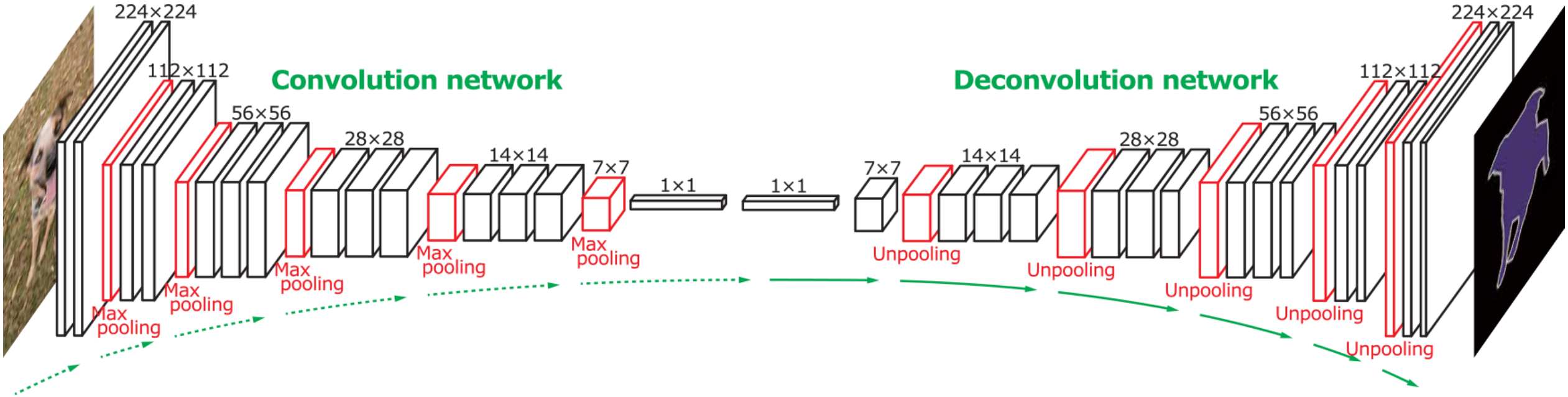
Deconvolution



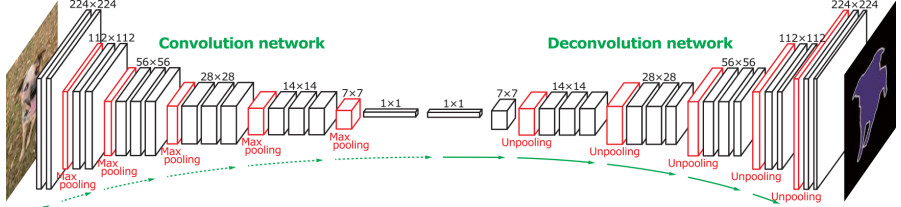




Deconvolution Network for Semantic Segmentation



H. Noh, S. Hong, and B. Han, Learning Deconvolution Network for Semantic Segmentation, ICCV 2015



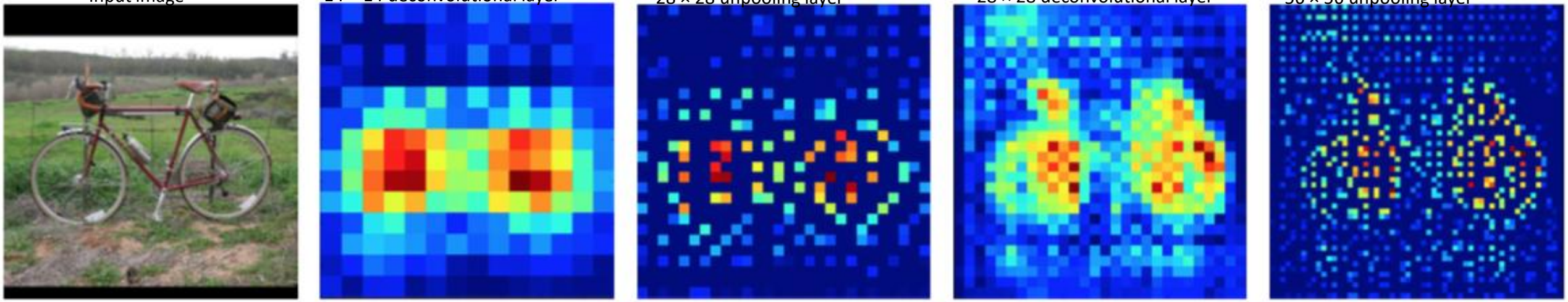
Input image

14 × 14 deconvolutional layer

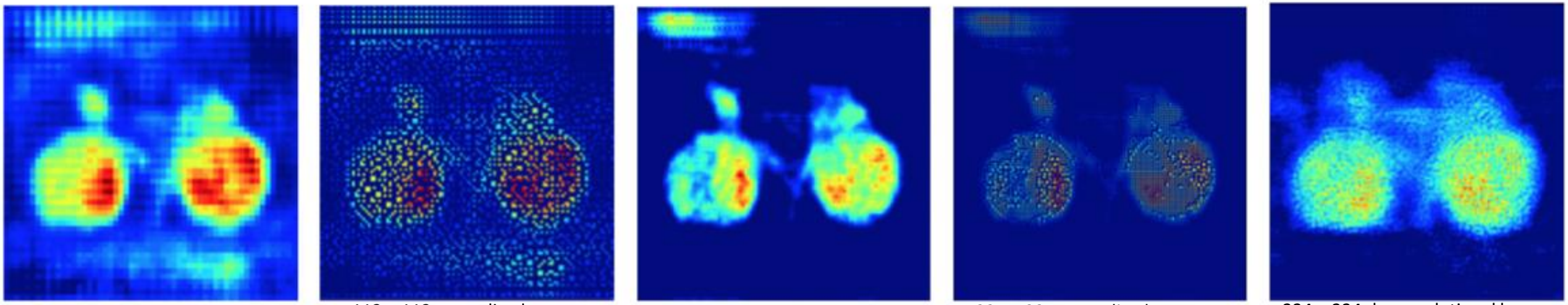
28 × 28 unpooling layer

28 × 28 deconvolutional layer

56 × 56 unpooling layer



(a) (b) (c) (d) (e)



56 × 56 deconvolutional layer

112 × 112 unpooling layer

112 × 112 deconvolutional layer

224 × 224 unpooling layer

224 × 224 deconvolutional layer

(f) (g) (h) (i) (j)

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

Learned upsampling architectures

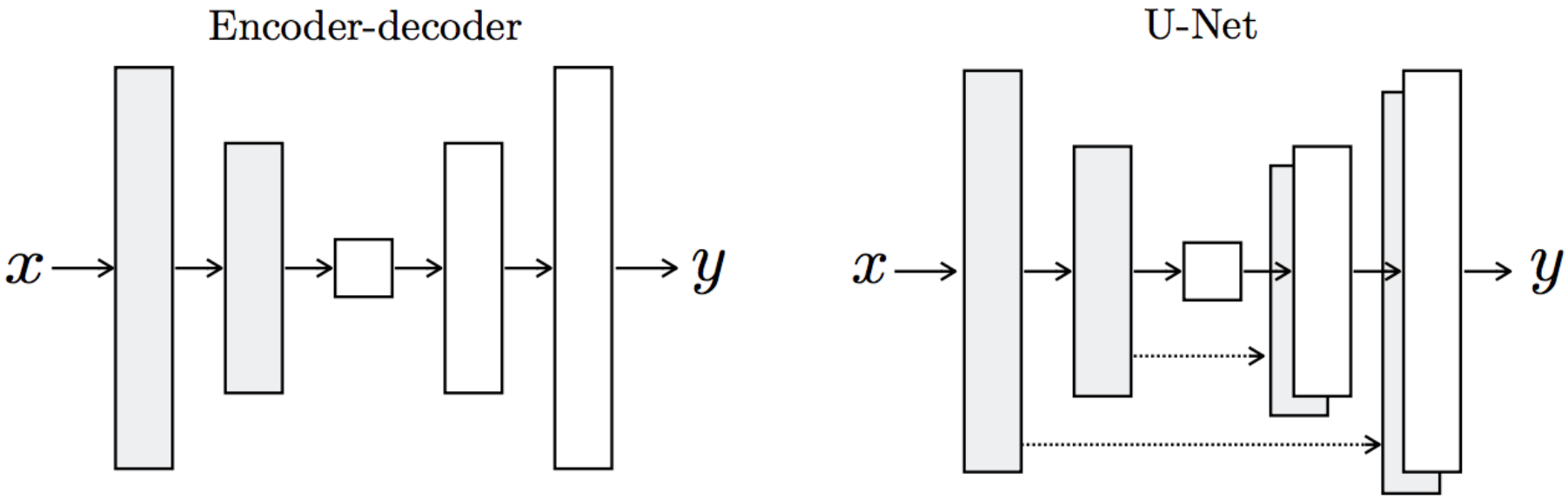
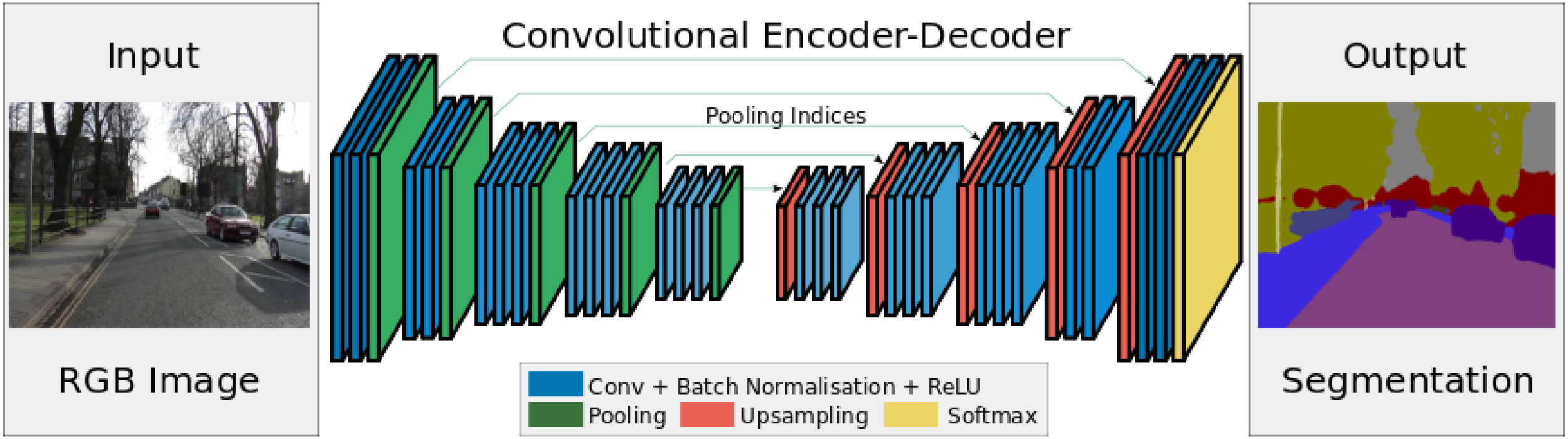


Figure source

SegNet



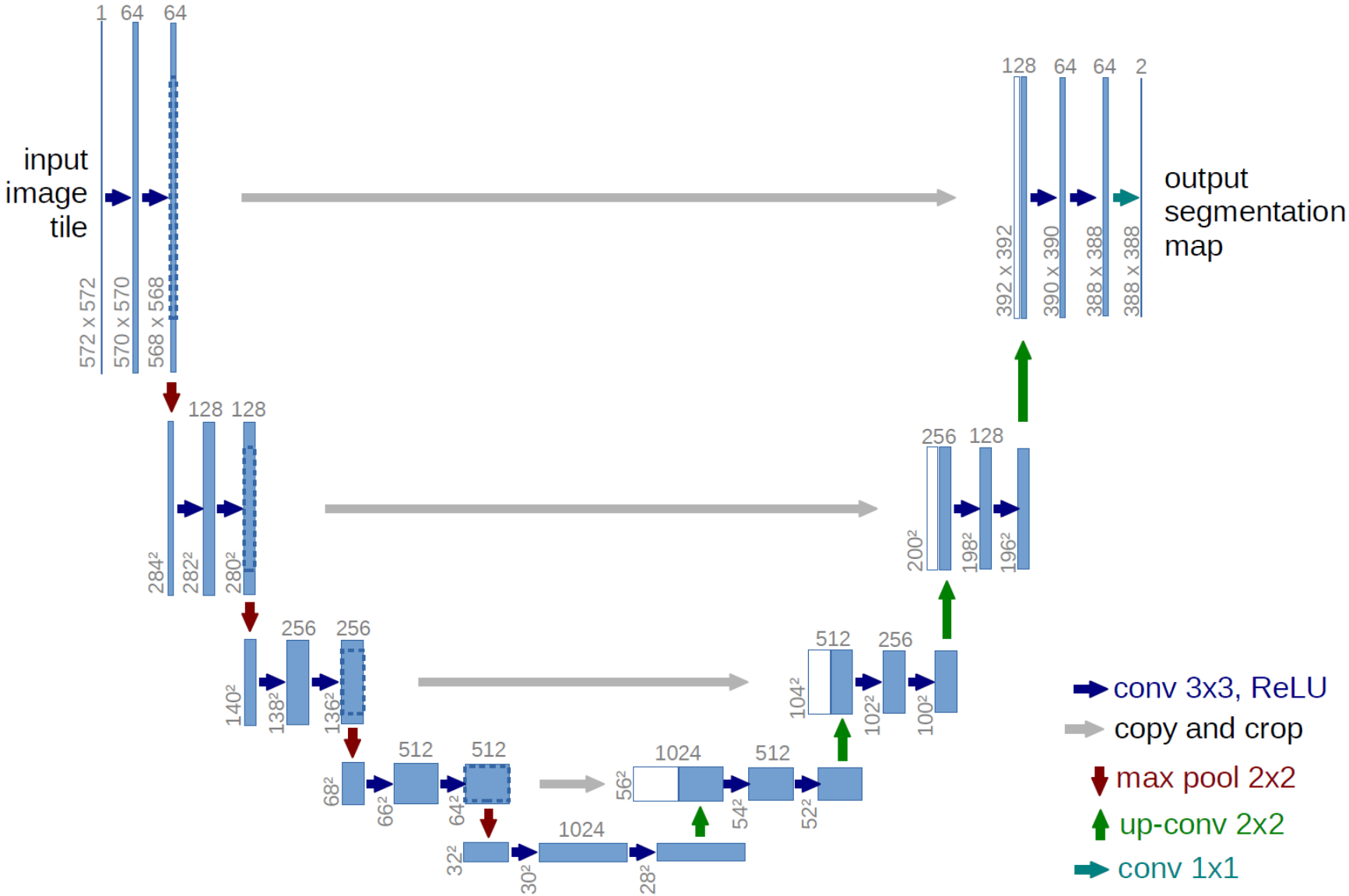
Uses VGG architecture!

No FC layer!

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

V Badrinarayanan, et al., A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling, 2015

U-Net



Source: Olaf Ronneberger, Philipp Fischer, Thomas Brox "U-Net: Convolutional Networks for Biomedical Image Segmentation", MICCAI, 2015

Questions?

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