THE UNIVERSITY OF TEXAS AT DALLAS

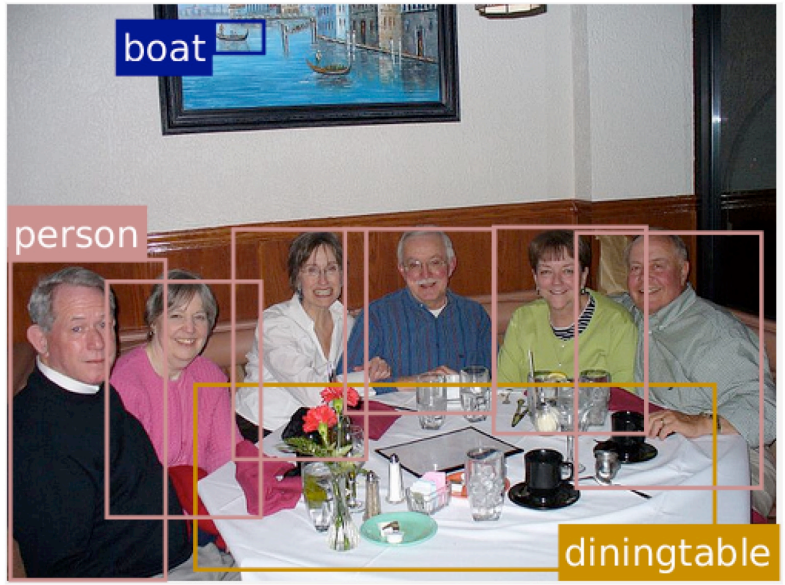
Semantic Segmentation

CS 4391 Introduction to Computer Vision

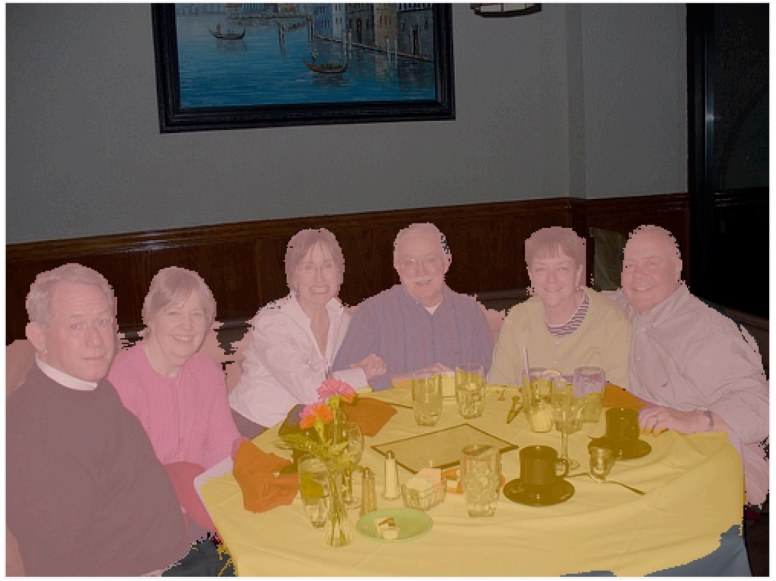
Professor Yapeng Tian

Department of Computer Science

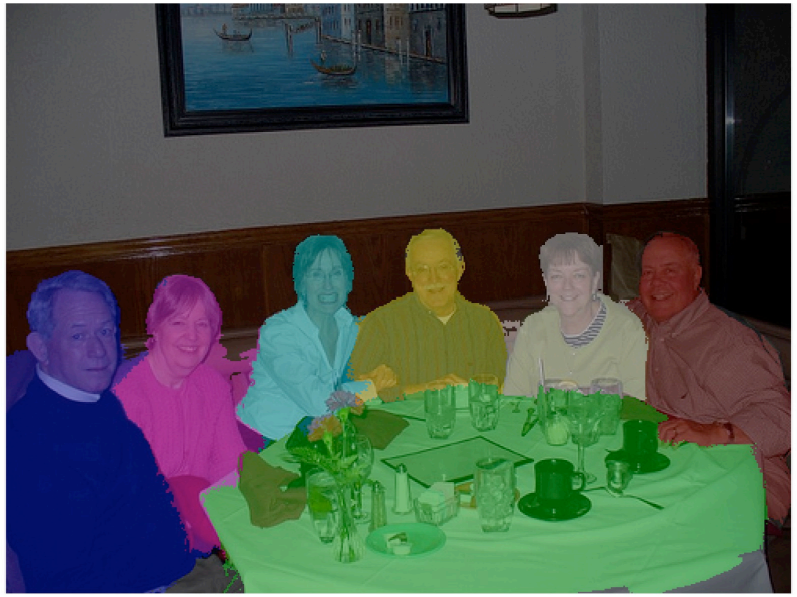
Semantic Scene Understanding



Object Detection



Semantic Segmentation



Instance Segmentation

Semantic Segmentation

Semantic image segmentation is the task of **classifying each pixel in an image from a predefined set of classes**

The pixels belonging to the bed are classified in the class “bed”, the pixels corresponding to the walls are labeled as “wall”, etc.

Problem Formulation

Input

segmented

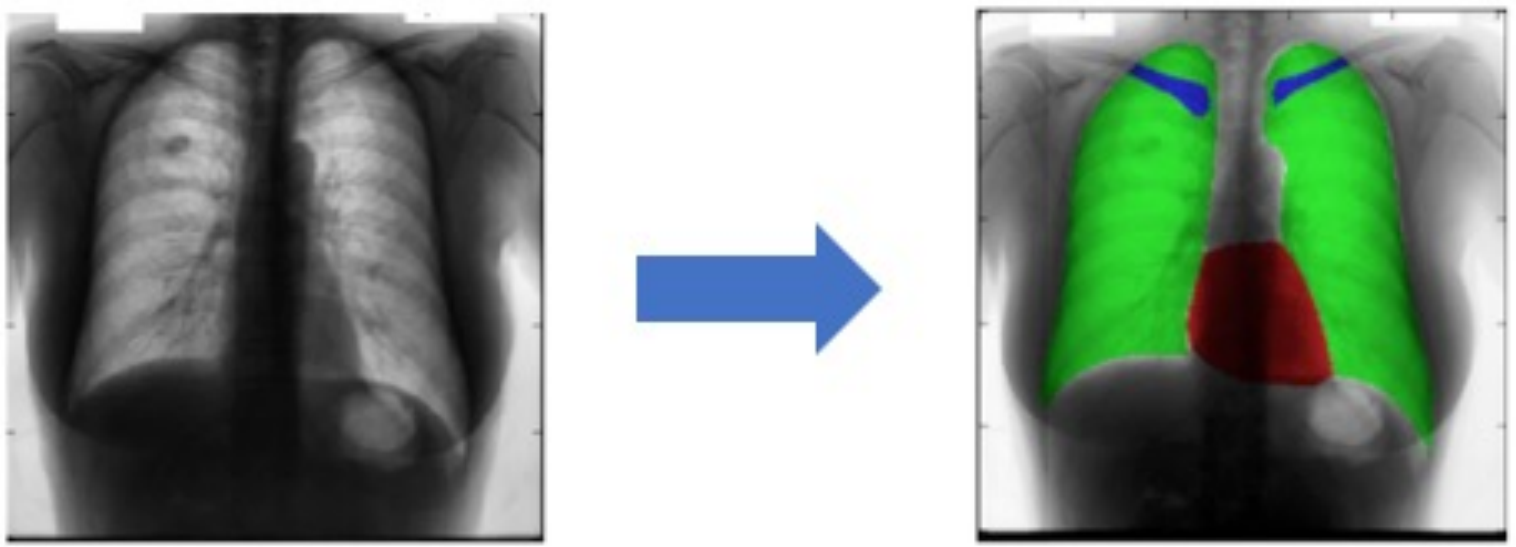
1: Person
2: Purse
3: Plants/Grass
4: Sidewalk
5: Building/Structures

3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	1	1	1	1	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	5	5	5	5	5	5	5
5	5	3	3	3	3	1	1	3	3	5	5	5	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	4	4	5	5	5
4	4	4	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	1	4	4	4	4	4	4	4

Semantic Labels

Given an image of size W x H x 3, we aim to generate a W x H matrix containing the predicted class labels corresponding to all the pixels.

Applications: Medical images



A chest x-ray with the heart (red), lungs (green), and clavicles (blue) are segmented.

Novikov et al. Fully Convolutional Architectures for Multi-Class Segmentation in Chest Radiographs, 2018

Applications: Autonomous Vehicles

A real-time segmented road scene for autonomous driving

https://www.youtube.com/watch?v=ATlcEDSPWXY

Semantic Segmentation

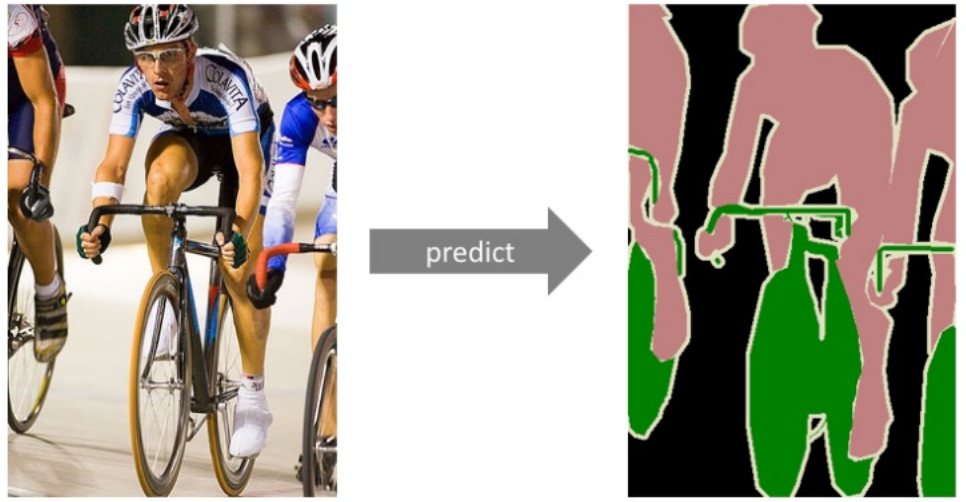
Label pixels into semantic classes

Naïve method

- Classify each pixel independently

Better idea

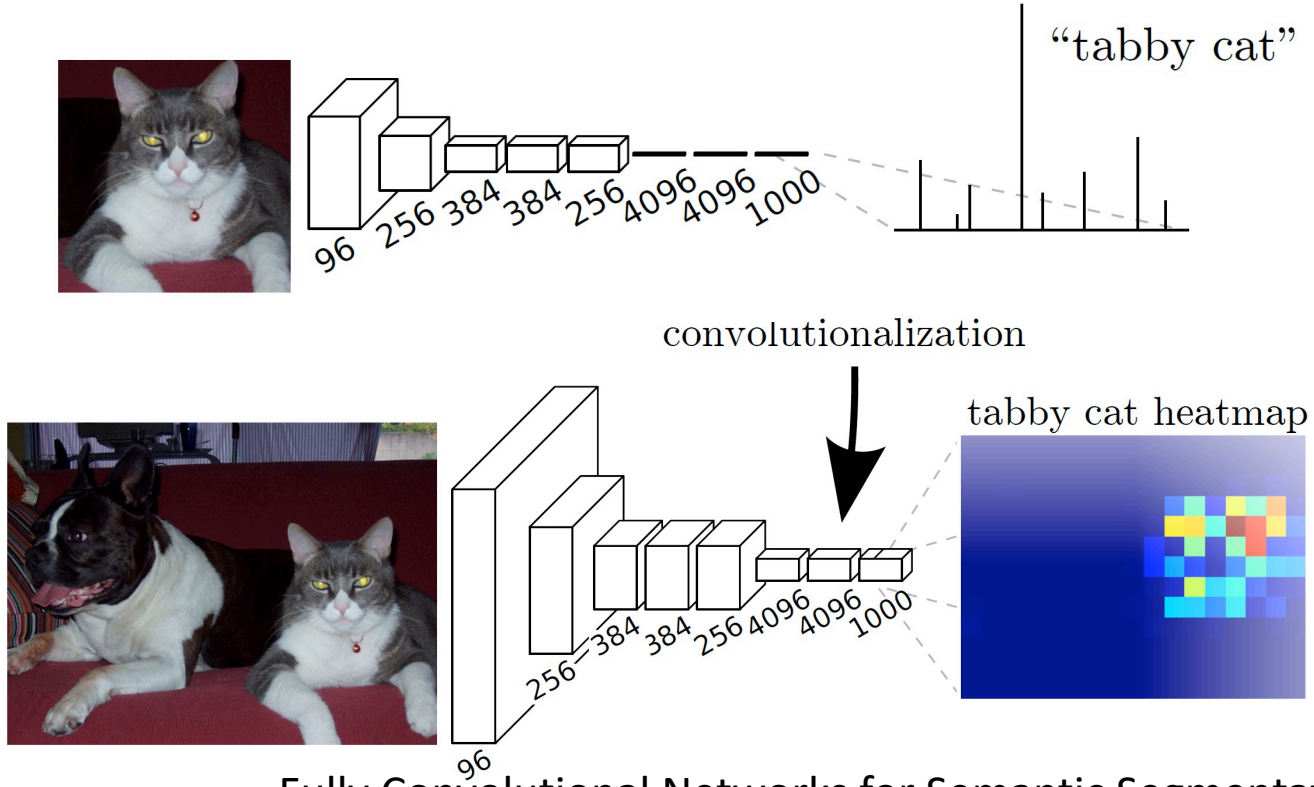
- Using context of pixels



Pixel-wise image classification

Fully Convolutional Networks

Adapt classification networks for dense prediction



- These FC layers can also be viewed as convolutions with kernels that cover their entire input regions
- Transforming FC layers into Conv layers enables a classification net to output a heatmap

Fully Convolutional Networks for Semantic Segmentation. Long et al., CVPR, 2015

Fully Convolutional Networks

Convert AlexNet

[224x224x3] INPUT
[55x55x96] CONV1: 96 11x11 filters at stride 4, pad 0
[27x27x96] MAX POOL1: 3x3 filters at stride 2
[27x27x96] NORM1: Normalization layer
[27x27x256] CONV2: 256 5x5 filters at stride 1, pad 2
[13x13x256] MAX POOL2: 3x3 filters at stride 2
[13x13x256] NORM2: Normalization layer
[13x13x384] CONV3: 384 3x3 filters at stride 1, pad 1
[13x13x384] CONV4: 384 3x3 filters at stride 1, pad 1
[13x13x256] CONV5: 256 3x3 filters at stride 1, pad 1
[6x6x256] MAX POOL3: 3x3 filters at stride 2
[4096] FC6: 4096 neurons
[4096] FC7: 4096 neurons
[1000] FC8: 1000 neurons (class scores)

```
layer {
  name: "fc6"
  type: "Convolution"
  bottom: "pool5"
  top: "fc6"
  convolution_param {
    num_output: 4096
    pad: 0
    kernel_size: 6
    group: 1
    stride: 1
  }
}
```

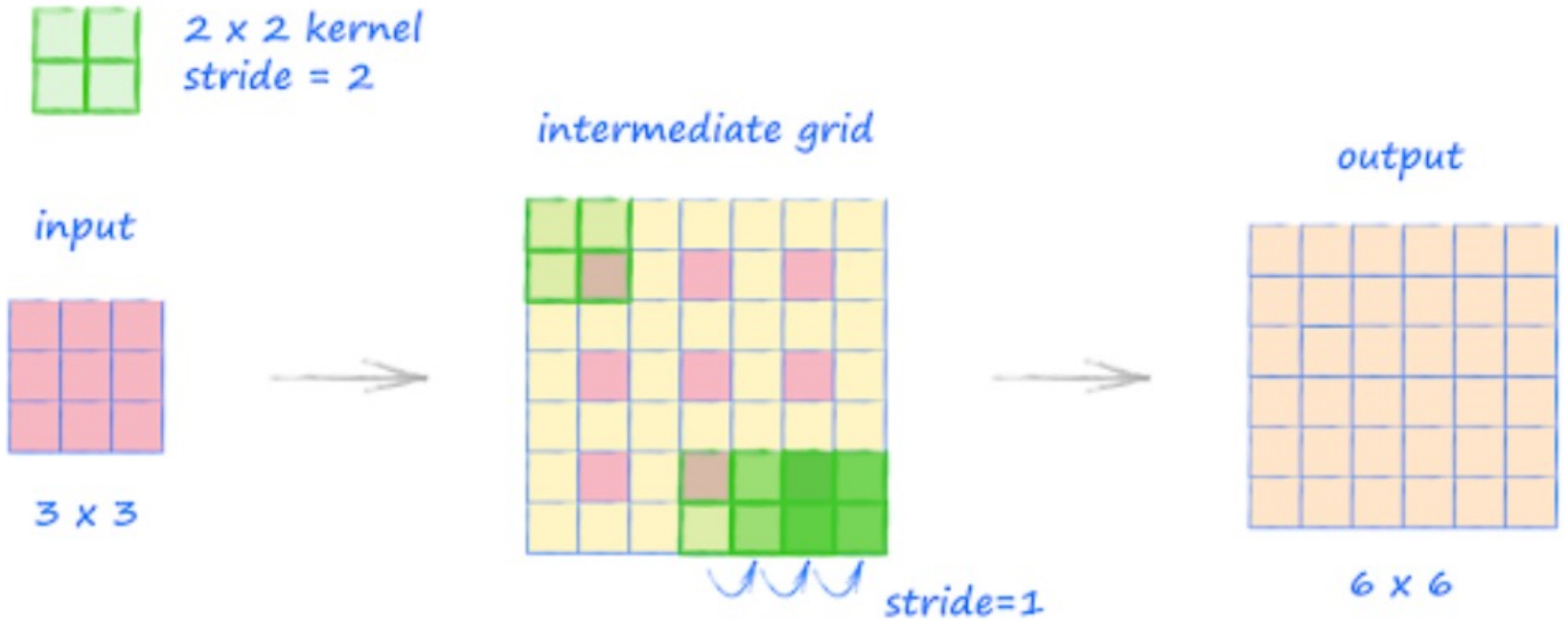
```
layer {
  name: "fc7"
  type: "Convolution"
  bottom: "fc6"
  top: "fc7"
  convolution_param {
    num_output: 4096
    pad: 0
    kernel_size: 1
    group: 1
    stride: 1
  }
}
```

```
layer {
  name: "score_fr"
  type: "Convolution"
  bottom: "fc7"
  top: "score_fr"
  param {
    lr_mult: 1
    decay_mult: 1
  }
  param {
    lr_mult: 2
    decay_mult: 0
  }
  convolution_param {
    num_output: 21
    pad: 0
    kernel_size: 1
  }
}
```

Fully Convolutional Networks for Semantic Segmentation. Long et al., CVPR, 2015

Fully Convolutional Networks

Deconvolution for up-sampling




```
layer {
  name: "upscore"
  type: "Deconvolution"
  bottom: "score_fr"
  top: "upscore"
  param {
    lr_mult: 0
  }
  convolution_param {
    num_output: 21
    bias_term: false
    kernel_size: 63
    stride: 32
  }
}
```

```
Pytorch: nn.ConvTranspose2d(in_channels, out_channels, kernel_size=2, stride=2)
```

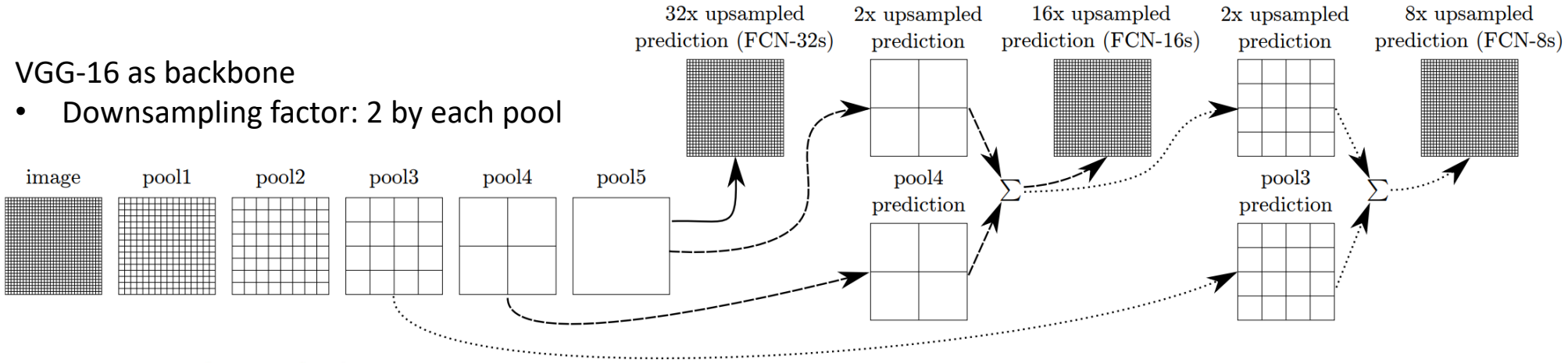
source

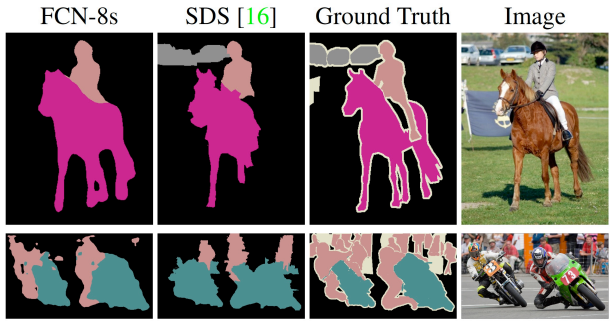
Fully Convolutional Networks

Combine predictions with different resolutions

VGG-16 as backbone

- Downsampling factor: 2 by each pool

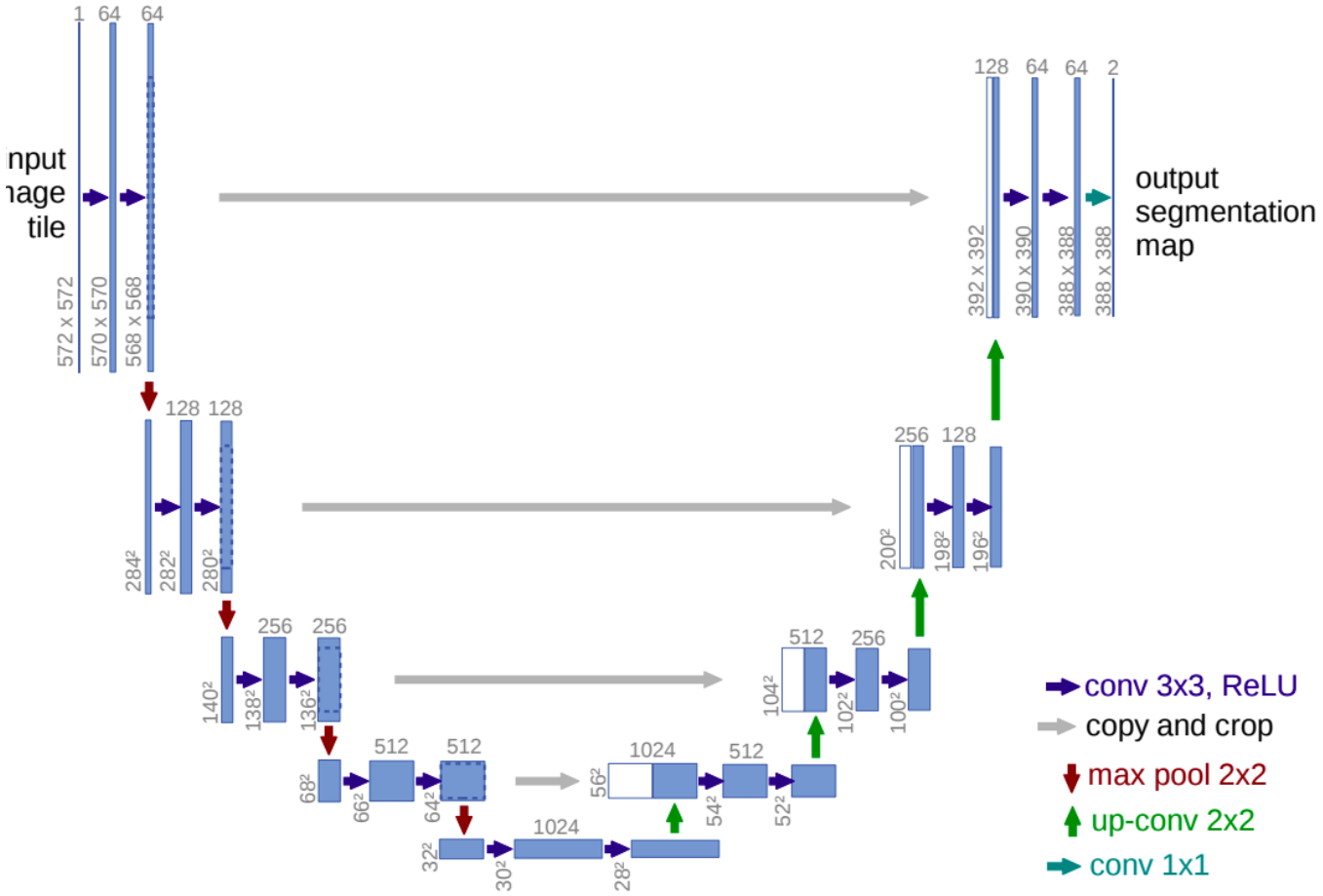




	pixel acc.	mean acc.	mean IU	f.w. IU
FCN-32s-fixed	83.0	59.7	45.4	72.0
FCN-32s	89.1	73.3	59.4	81.4
FCN-16s	90.0	75.7	62.4	83.0
FCN-8s	**90.3**	**75.9**	**62.7**	**83.2**

Fully Convolutional Networks for Semantic Segmentation. Long et al., CVPR, 2015

U-Net



Convolution

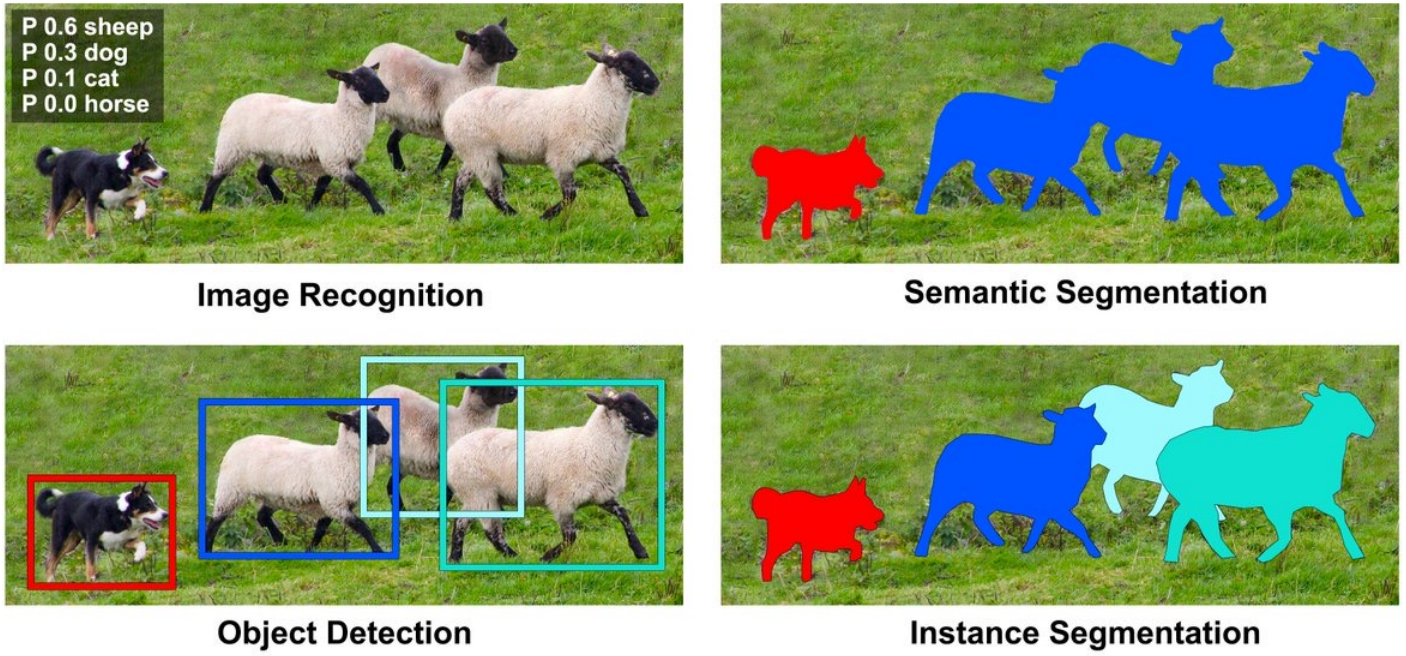
Deconvolution

U-Net: Convolutional Networks for Biomedical Image Segmentation, Ronneberger et al., MICCAI 2015

Instance Segmentation

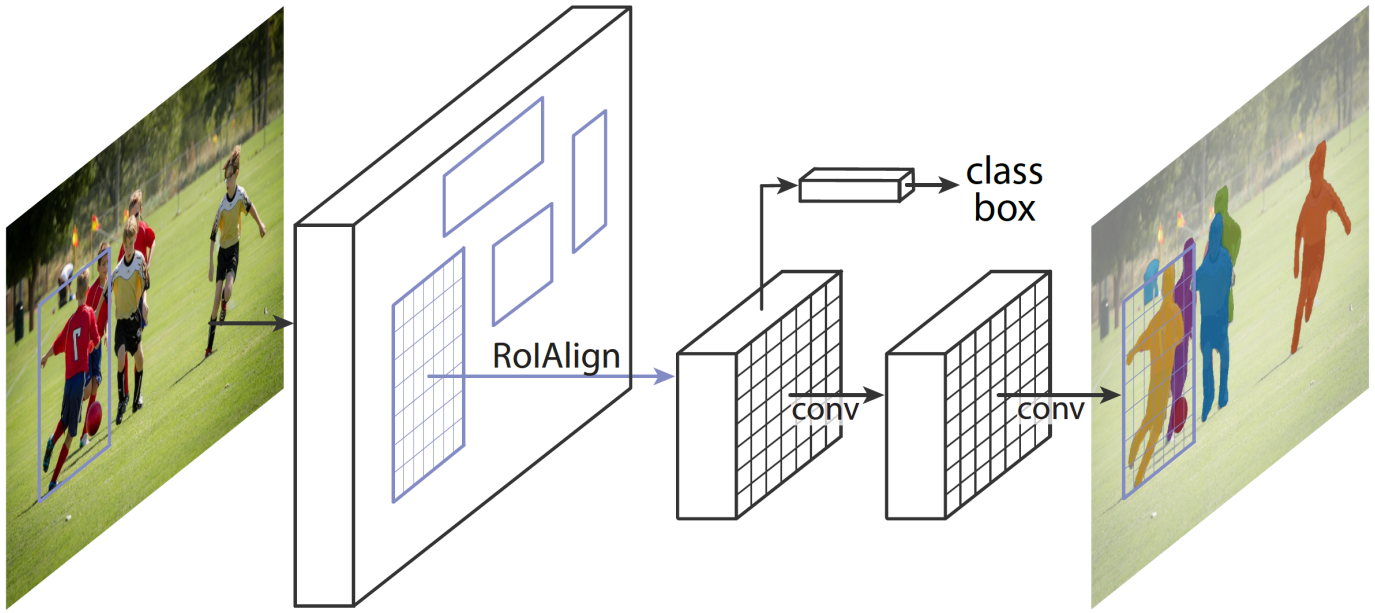
Separate object instances in the same class

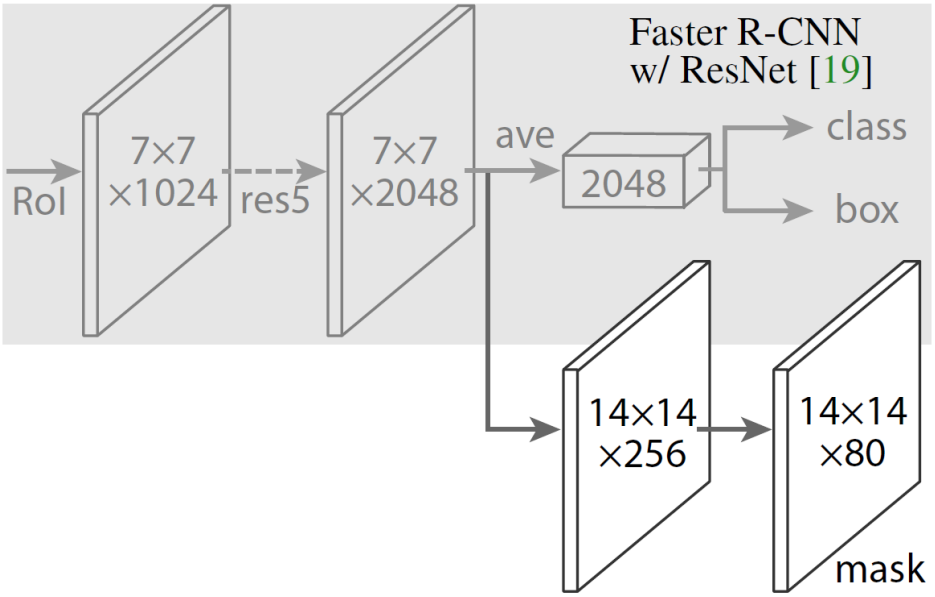
Detection + segmentation



https://ai-pool.com/d/could-you-explain-me-how-instance-segmentation-works

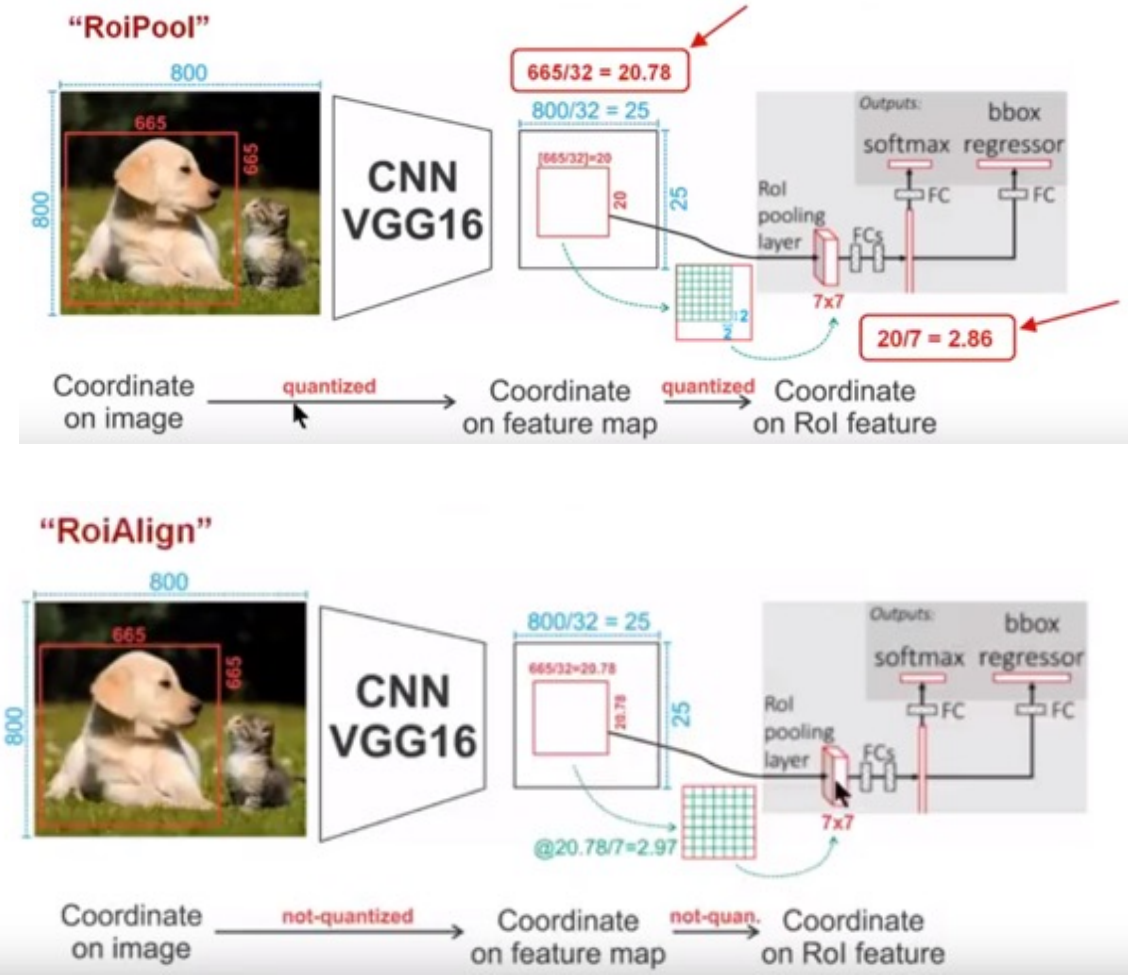
Mask R-CNN





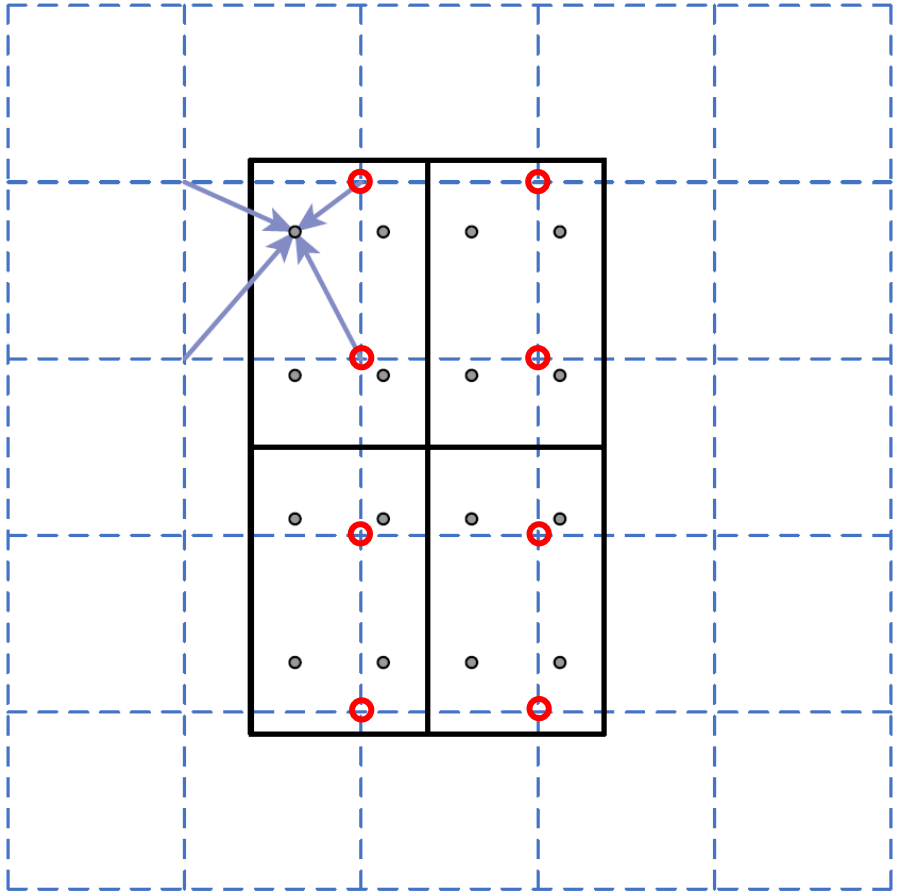
Mask R-CNN. He et al., ICCV, 2017

RoI Pooling vs. RoI Align



Credit: Ardian Umam

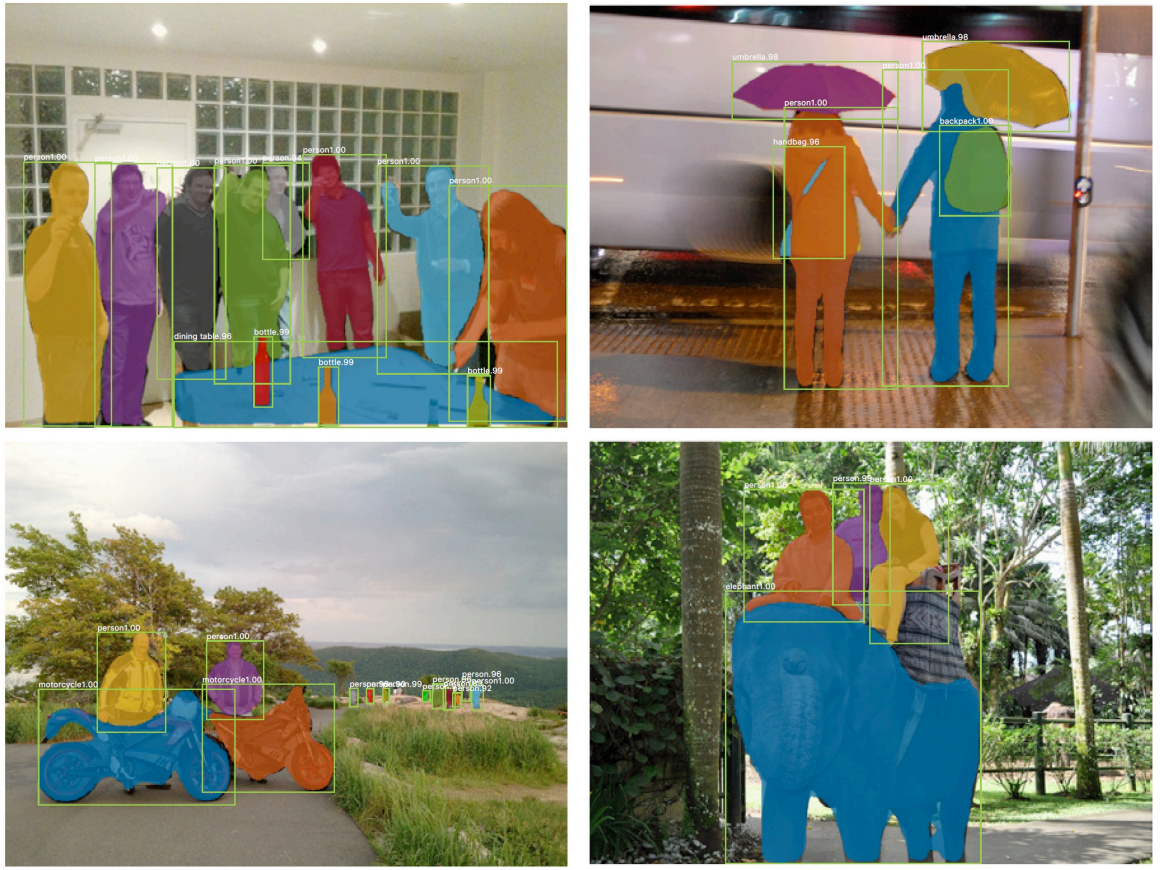
Bilinear interpolation for non-integer positions in RoI align



BI formula see wiki

Mask R-CNN

	align?	bilinear?	agg.	AP	AP_{50}	AP_{75}
RoIPool [12]			max	26.9	48.8	26.4
RoIWarp [10]		✓	max	27.2	49.2	27.1
		✓	ave	27.1	48.9	27.1
RoIAlign	✓	✓	max	**30.2**	**51.0**	**31.8**
	✓	✓	ave	**30.3**	**51.2**	**31.5**



Mask R-CNN. He et al., ICCV, 2017

Computer Vision

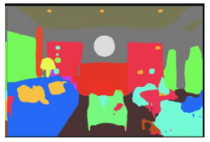
Semantic Segmentation

Edit

3673 papers with code • 97 benchmarks • 255 datasets

Semantic segmentation, or image segmentation, is the task of clustering parts of an image together which belong to the same object class. It is a form of pixel-level prediction because each pixel in an image is classified according to a category. Some example benchmarks for this task are Cityscapes, PASCAL VOC and ADE20K. Models are usually evaluated with the Mean Intersection-Over-Union (Mean IoU) and Pixel Accuracy metrics.

(Image credit: CSAILVision)



Benchmarks

Add a Result

These leaderboards are used to track progress in Semantic Segmentation

Trend	Dataset	Best Model	Paper	Code	Compare
	ADE20K	InternImage-H (M3I Pre-training)			See all
	Cityscapes test	InternImage-H			See all
	ADE20K val	BEiT-3			See all
	Cityscapes val	InternImage-H			See all
	NYU Depth v2	CMX (B5)			See all
	PASCAL Context	InternImage-H			See all
	PASCAL VOC 2012 test	DeepLabv3+ (Xception-65-JFT)			See all
	S3DIS	WindowNorm+StratifiedTransformer			See all
	DensePASS	Trans4PASS+ (multi-scale)			See all
	S3DIS Area5	PTv2			See all

Show all 97 benchmarks

Libraries

Use these libraries to find Semantic Segmentation models and implementations

PaddlePaddle/PaddleSeg	52 papers	6,625 ★
osmr/imgclsmob	30 papers	2,776 ★
rwightman/pytorch-image-models	27 papers	24,242 ★
open-mmlab/mmsegmentation	19 papers	5,431 ★

See all 31 libraries.

Content

Summary

Semantic segmentation

- Label pixels into object classes

Instance segmentation

- Separate object instances in the same class
- Detection + segmentation inside each box

Further Reading

FCN, 2015 https://arxiv.org/abs/1411.4038

Unet, 2015 https://arxiv.org/abs/1505.04597

Mask R-CNN, 2017 https://arxiv.org/abs/1703.06870

DeepLab, 2015 https://arxiv.org/abs/1606.00915

A semantic segmentation overview
https://www.jeremyjordan.me/semantic-segmentation/