

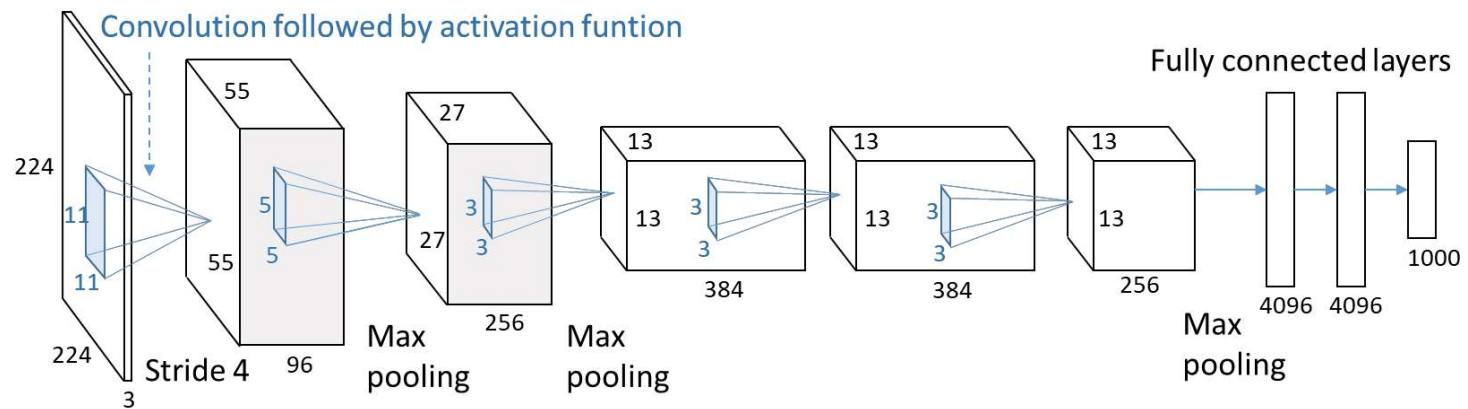
# Object detection and segmentation

# This time

- Classification (recap)
- What image belongs to what category?
- Segmentation
- What pixels belong to what category?
- Detection
- What pixels belong to what instance of what category?

# Supervised learning

- So far: just classification



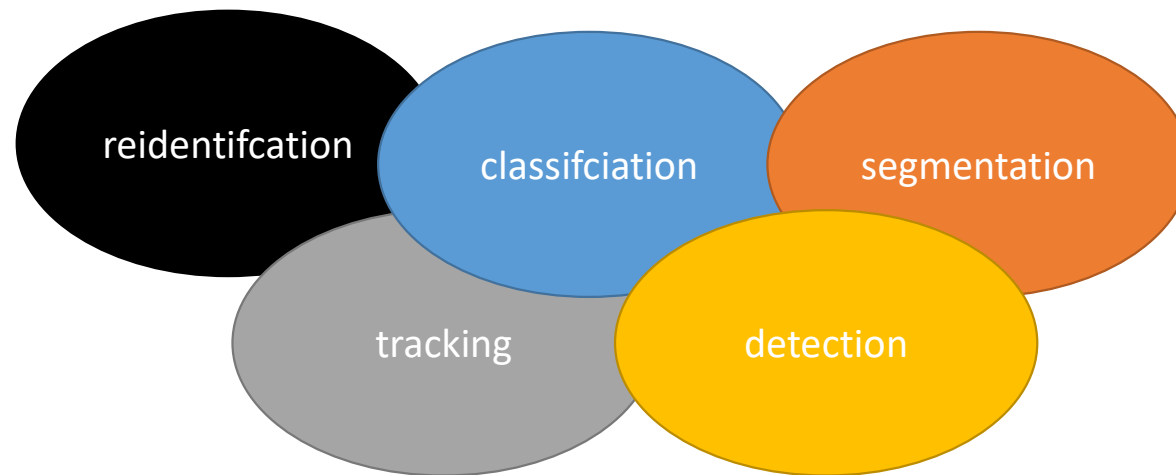
DOG

0.01  
0.04  
0.82  
...  
0.00  
0.02


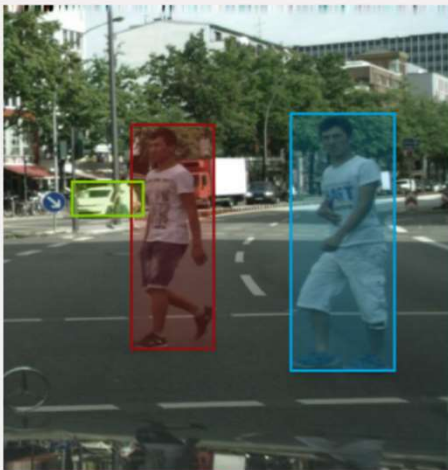
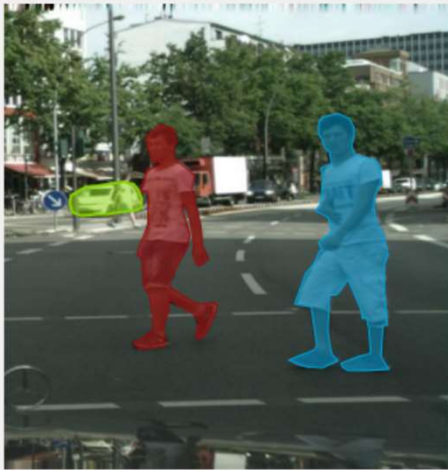
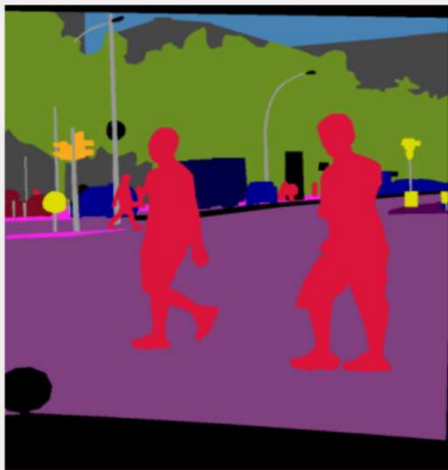
Image in → label out (score vector)

Typically an image shows more than a single class though...

# Supervised learning



# Supervised learning

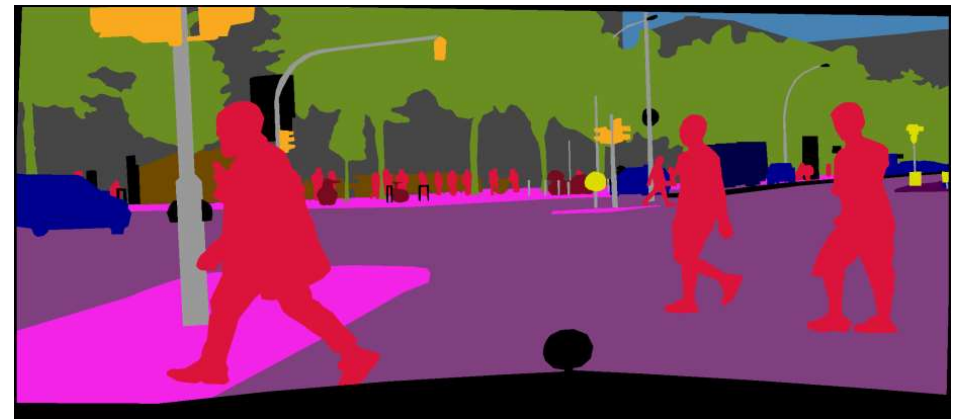
Classification+localization	Object detection	Instance segmentation	Semantic segmentation
			
<div>Car</div>	<div>Pedestrian</div> <div>Car</div> <div>Pedestrian</div>	<div>Pedestrian</div> <div>Car</div> <div>Pedestrian</div>	<div>Truck</div> <div>Road</div> <div>Vegetation</div> <div>Pedestrian</div> <div>Sidewalk</div> <div>Car</div>
Single object	Multiple objects	Multiple objects	No objects, just pixels

# Semantic segmentation

- Assign a class to each pixel in an image
- Do not distinguish between several instances of the same category

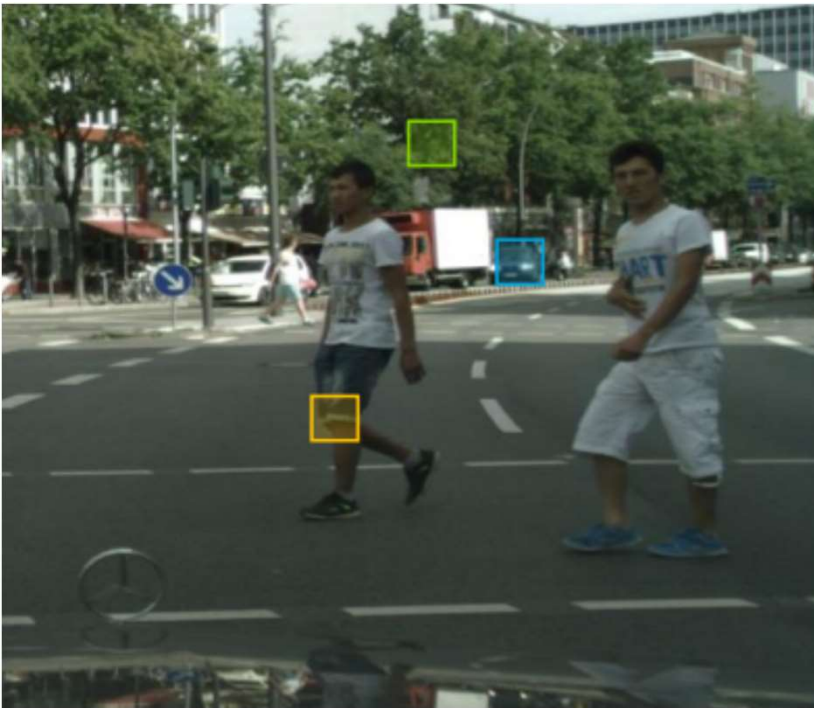
Main difference  
with instance  
segmentation?

Pedestrian Road  
Vegetation  
Sidewalk  
Car  
Truck

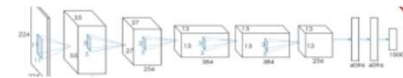


# Semantic segmentation

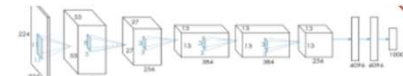
Idea: sliding window



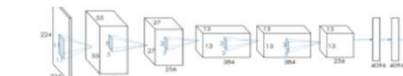
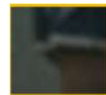
Why is this not a good idea?



Tree



Car



Pedestrian

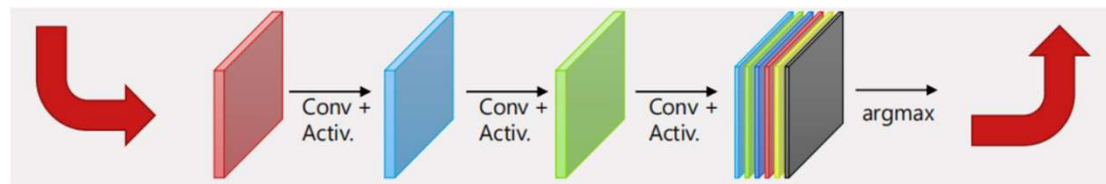


# Semantic segmentation

How could we characterize the loss in this case?

- Idea: fully convolutional

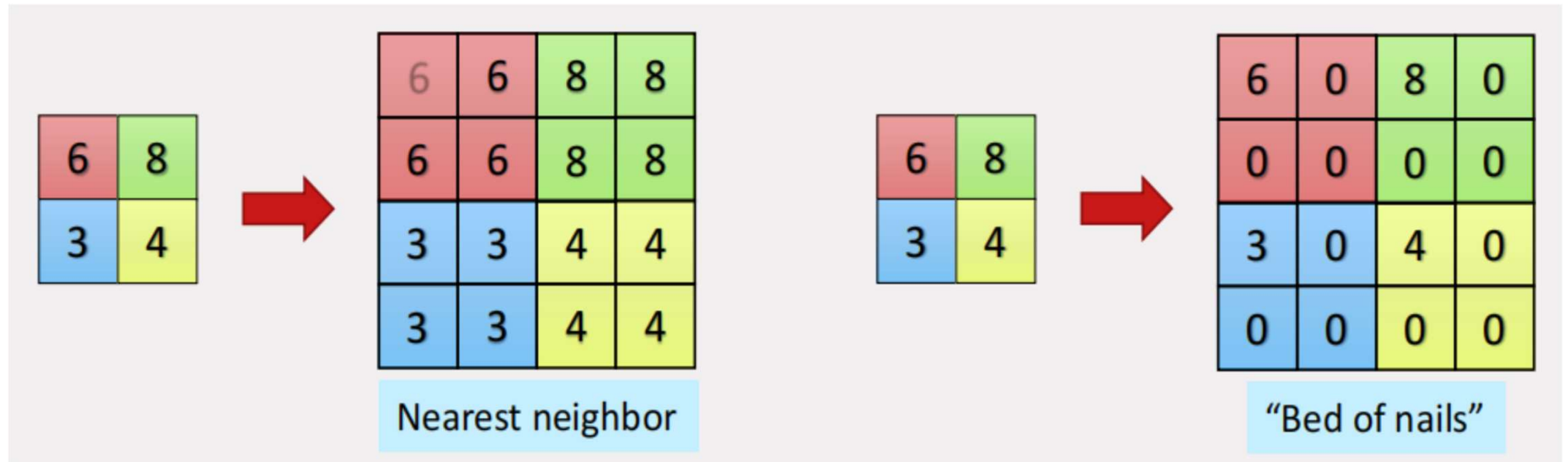
Problem in this case?





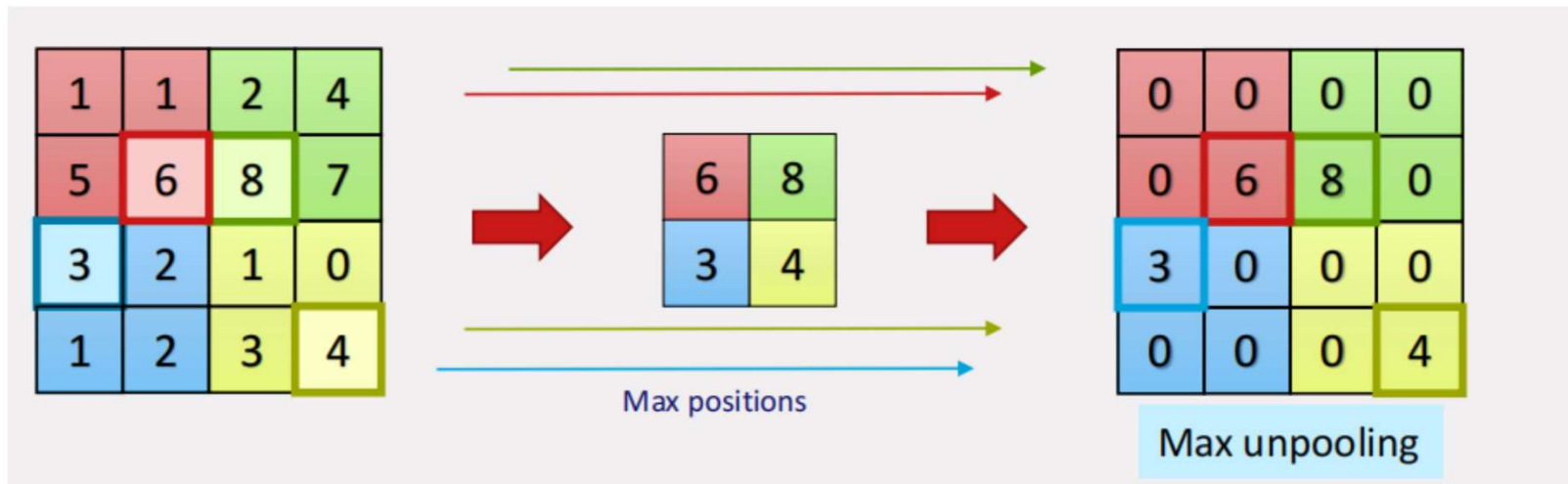
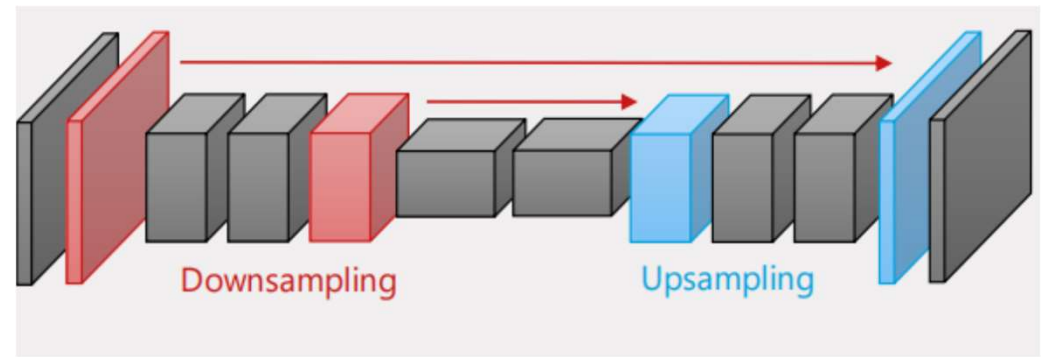
# Semantic segmentation

- Upsampling... but how?
- Unpooling

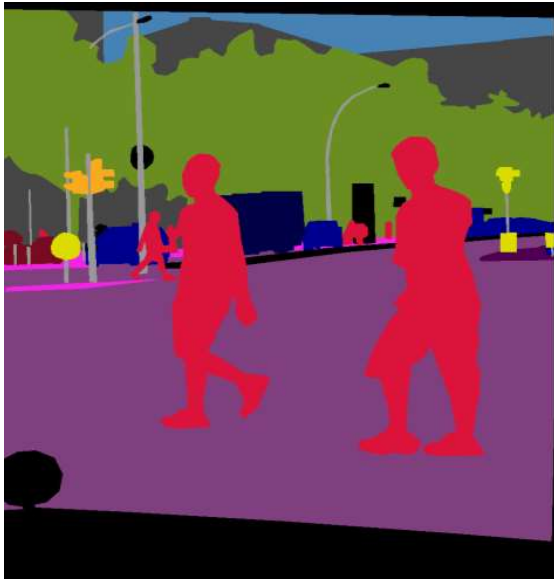


# Semantic segmentation

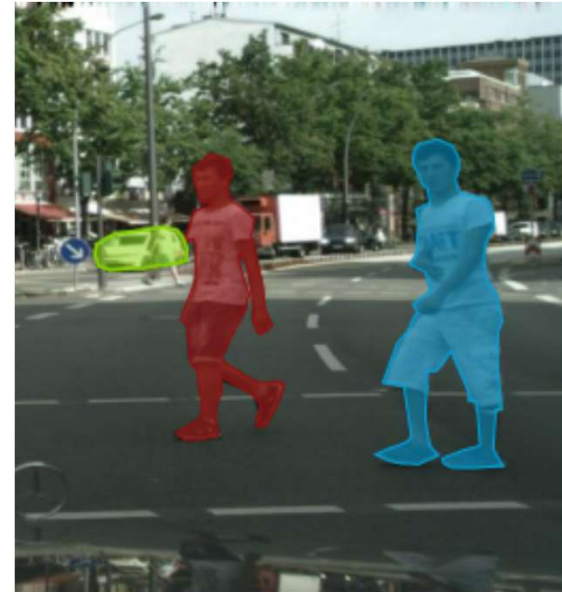
- Upsampling... but how?
- Unpooling



# Instance segmentation

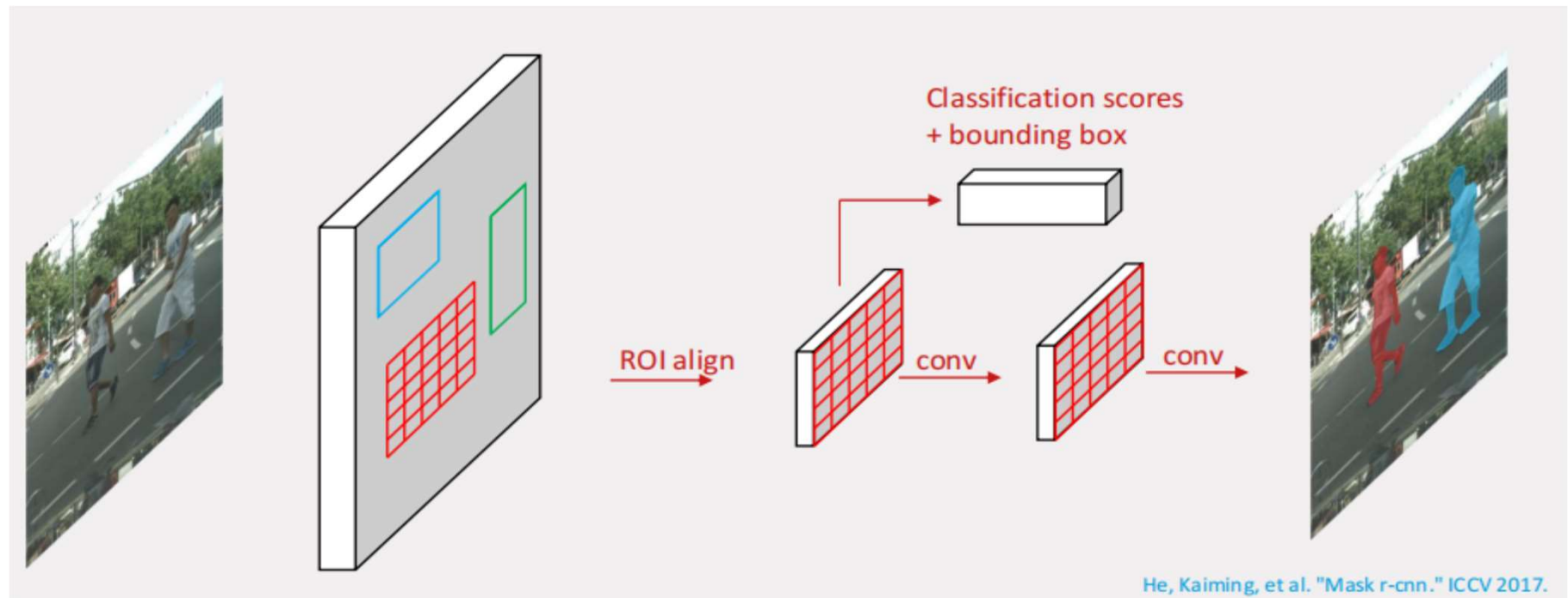


Pedestrian



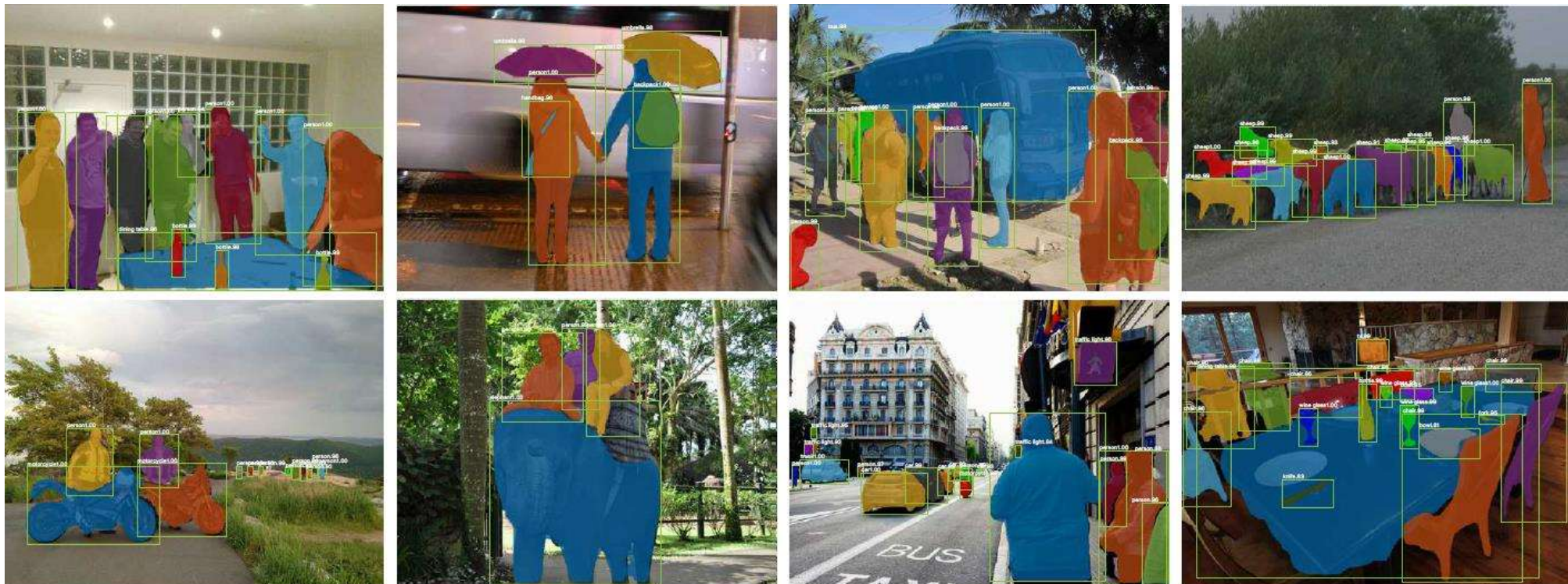
Pedestrian 1

# Instance segmentation



# Instance segmentation

Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick, “Mask R-CNN”, The IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2961-2969 ([link](#))





# Summary

- Different forms of supervised learning we've seen so far
  - Classification
  - Predict category of an image
- Semantic segmentation
  - Predict category of pixels within an image
  - Classification + localization
  - Predict category of an image + bounding box of the location of the object within the image
- Object detection
  - Detect multiple instances of different categories within a single image
  - Instance segmentation
  - Detect and segment multiple instances of different categories within a single image