Panoptic Segmentation: Task and Approaches

CVPR 2019 Tutorial
Visual Recognition and Beyond

Alexander Kirillov

facebook Artificial Intelligence

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

In this tutorial:

- panoptic segmentation task – unified semantic segmentation task

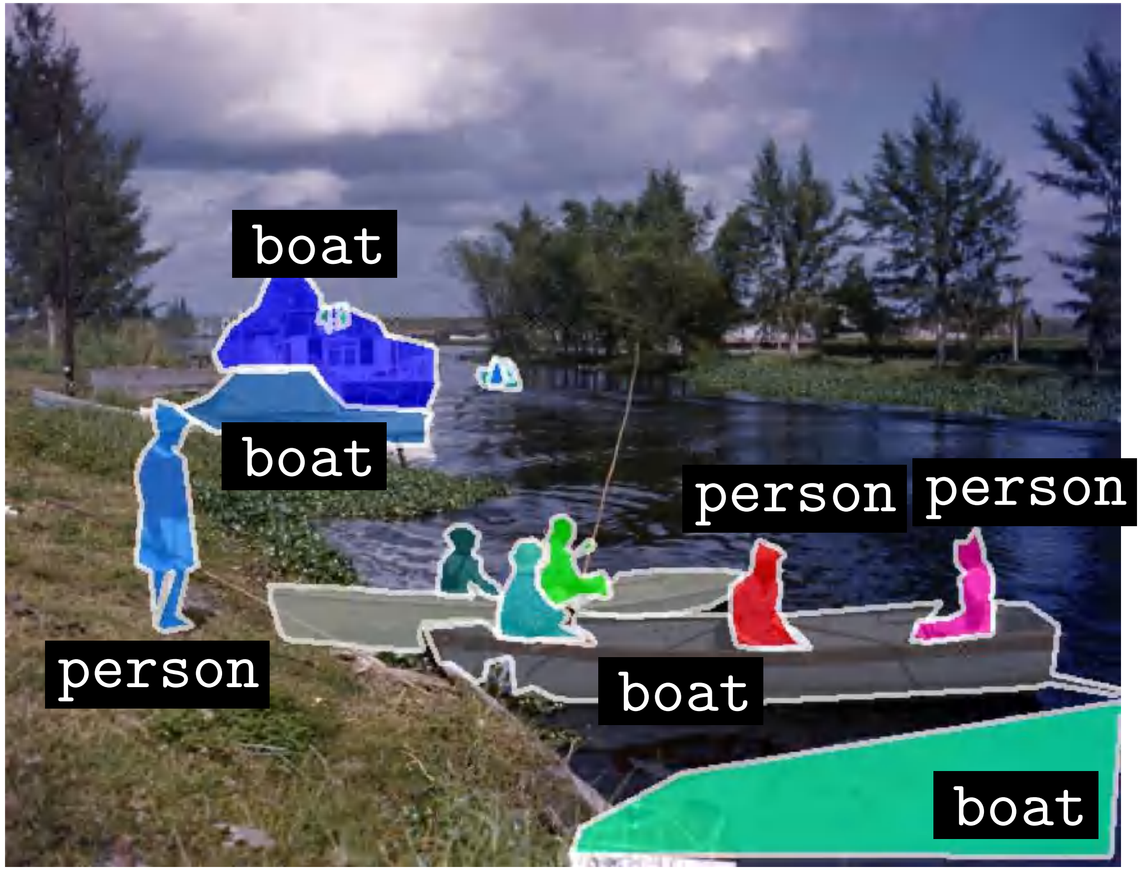
- approaches for the task

In this tutorial:

- panoptic segmentation task – unified semantic segmentation task

- approaches for the task

Image segmentation tasks last 10 years



instance segmentation

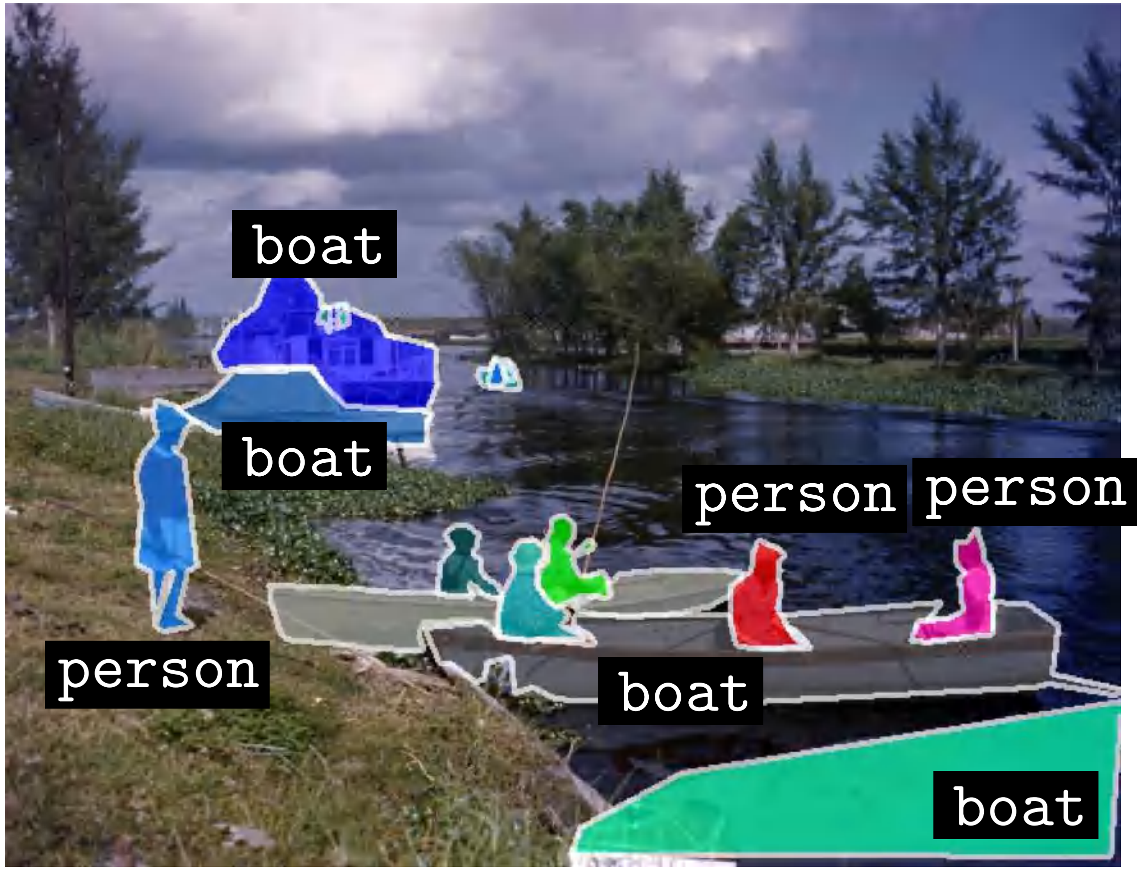
delineate each
object with a mask

Image segmentation tasks last 10 years

assign semantic label
to each pixel

semantic segmentation

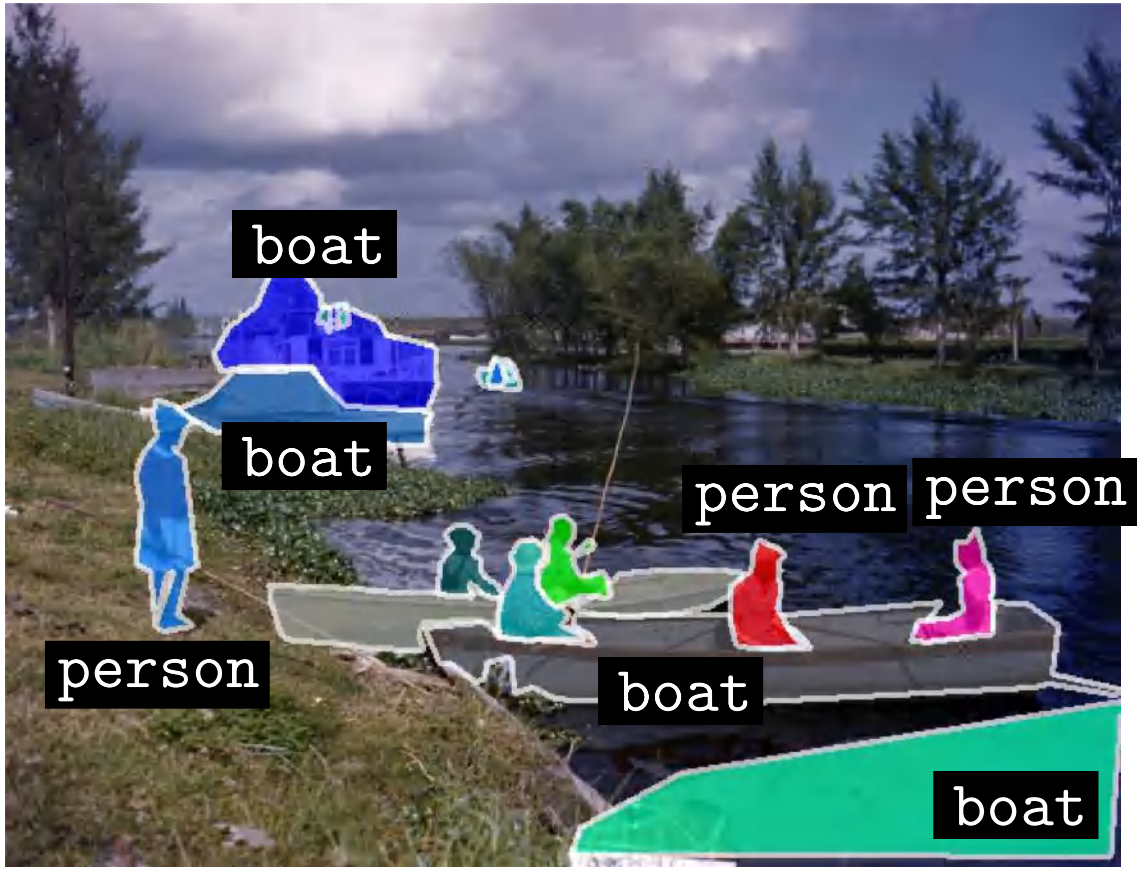
Image segmentation tasks last 10 years



instance segmentation

semantic segmentation

Image segmentation tasks last 10 years

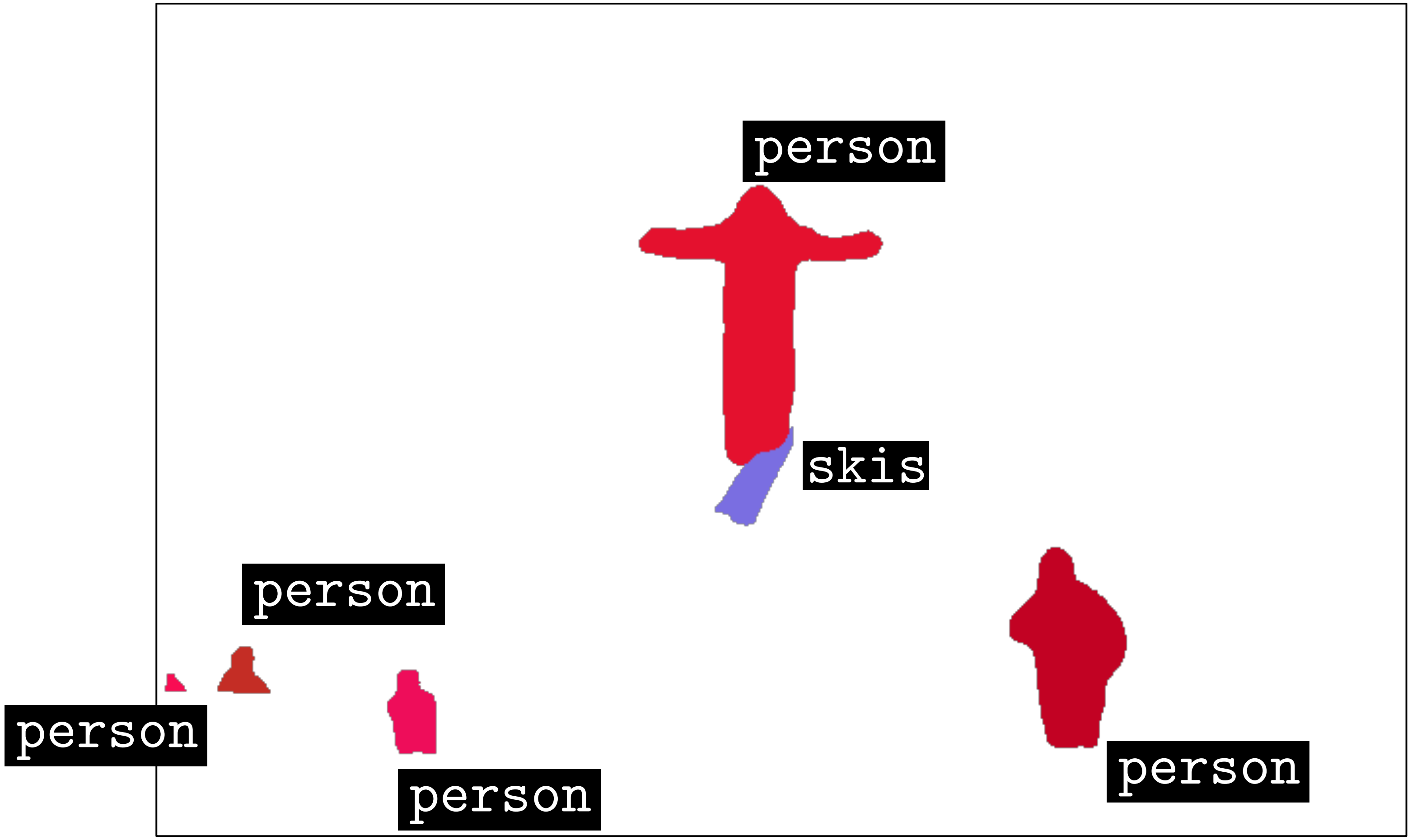


instance segmentation

semantic segmentation

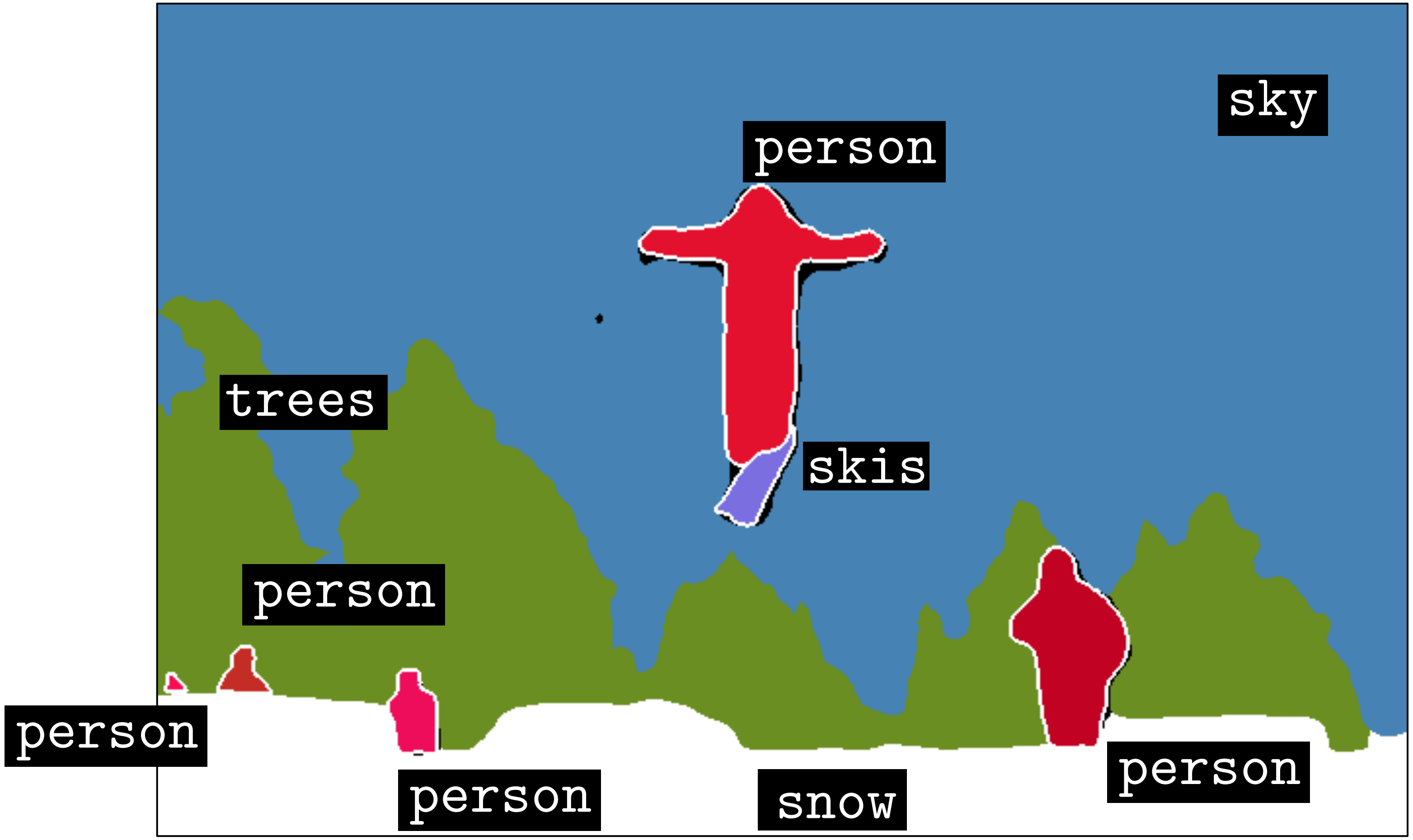
real-world application likely requires both modalities

What do instance segmentation methods "see"?



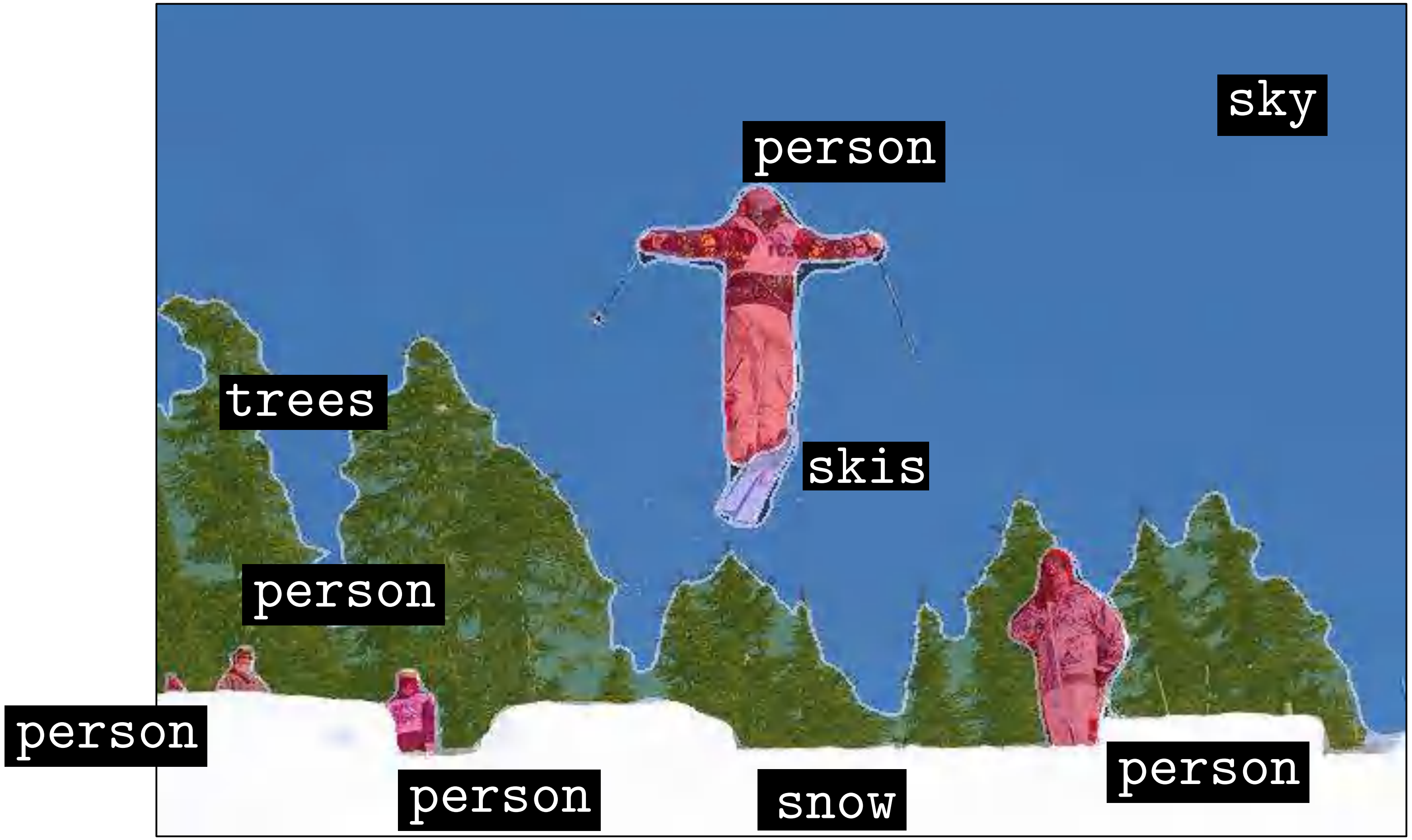
no understanding of the general scene layout

What do instance segmentation methods "see"?



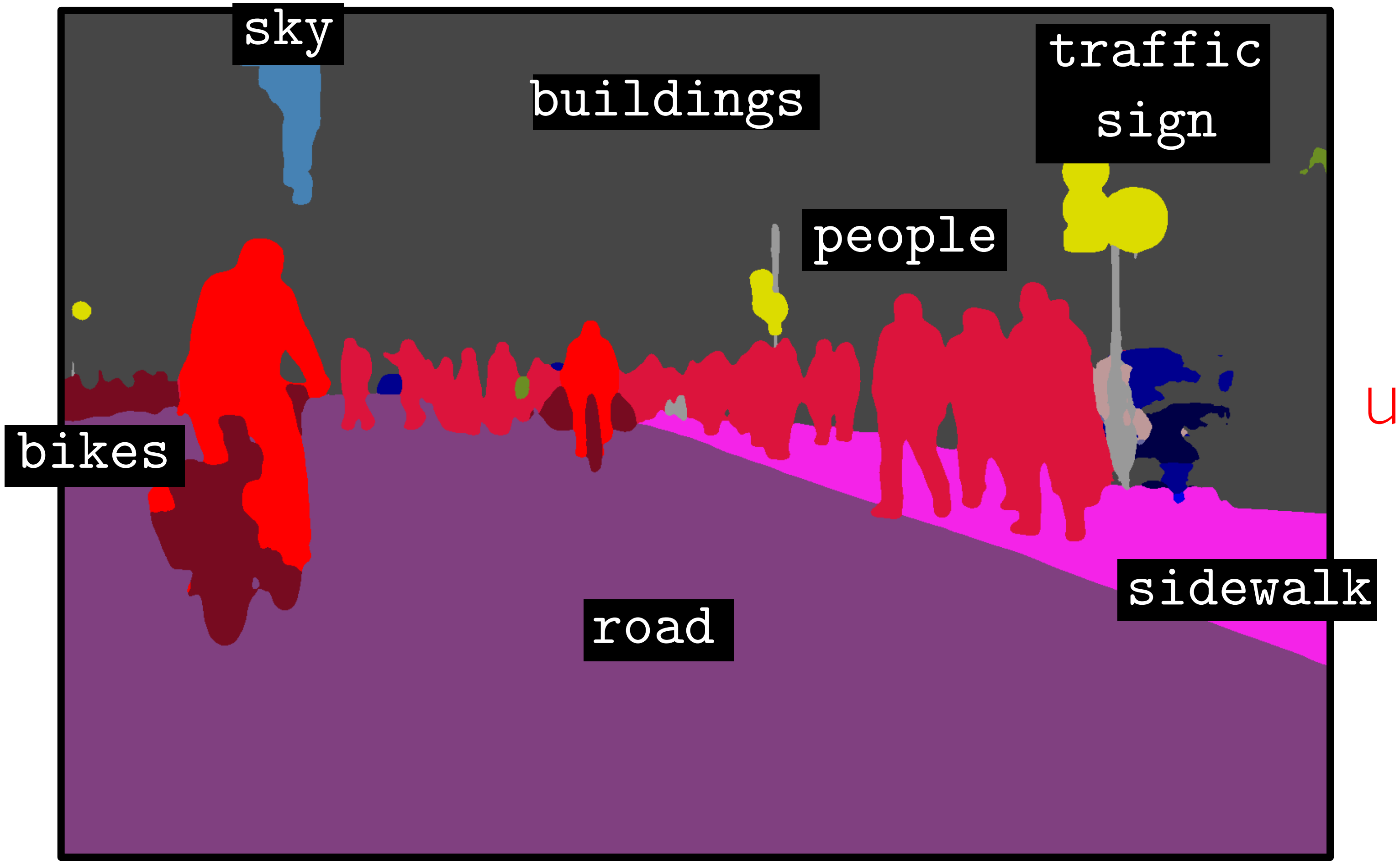
combined with
semantic segmentation
prediction

What do instance segmentation methods “see”?



combined with
semantic segmentation
prediction

What do semantic segmentation methods "see"?

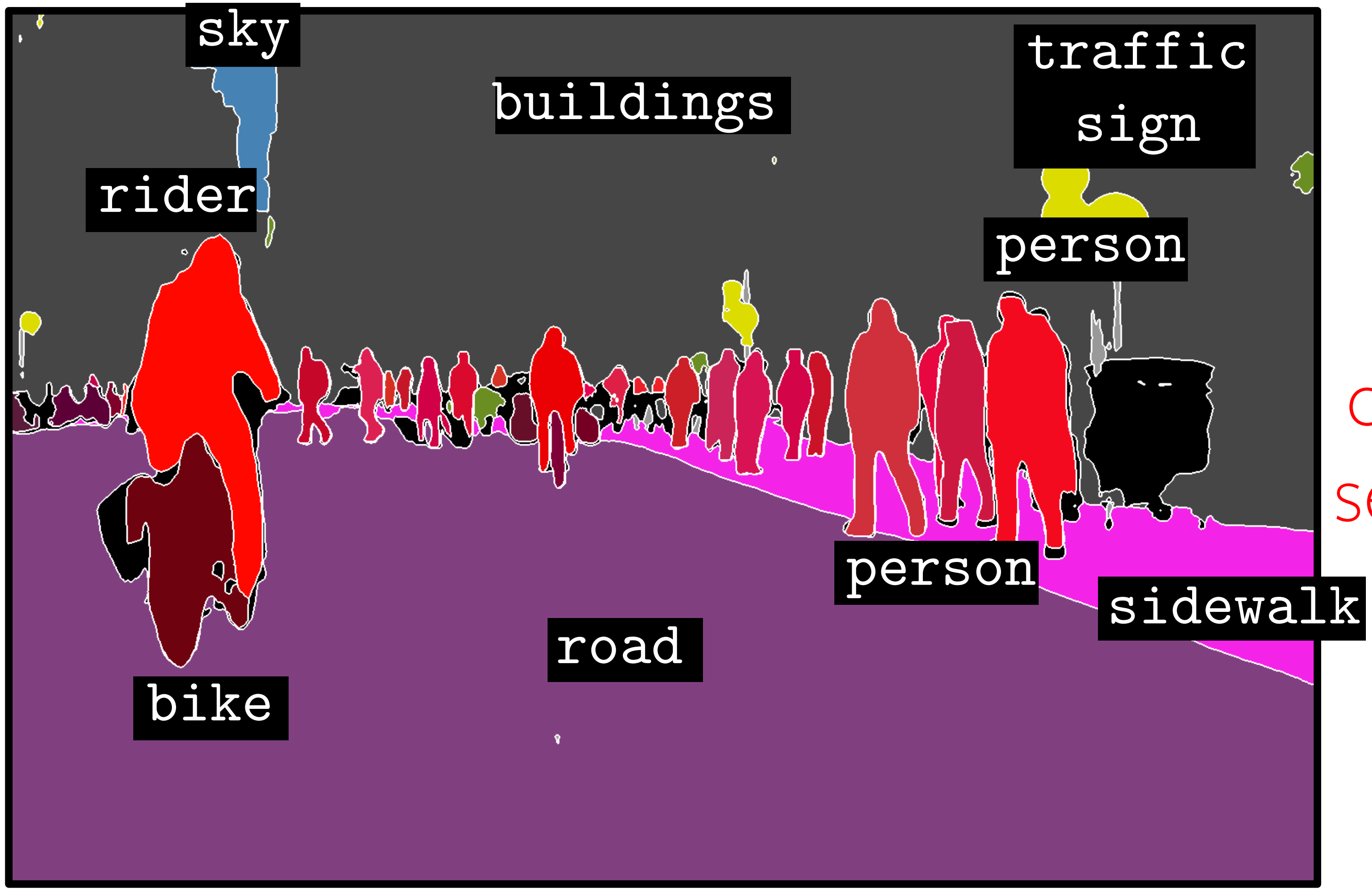


unable to reason about separate objects

What do semantic segmentation methods “see”?

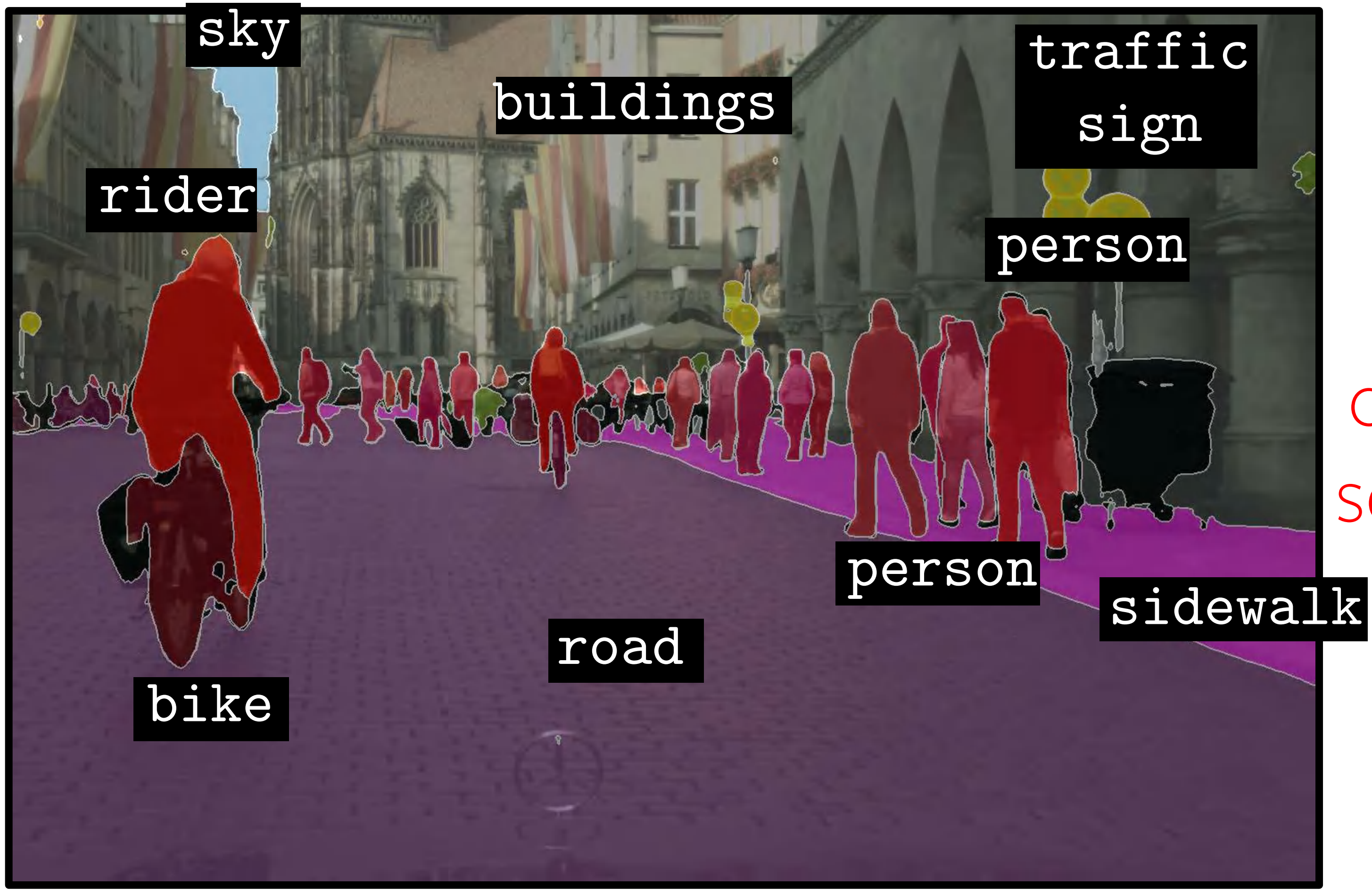
does this blob of people want to cross the road?

What do semantic segmentation methods "see"?



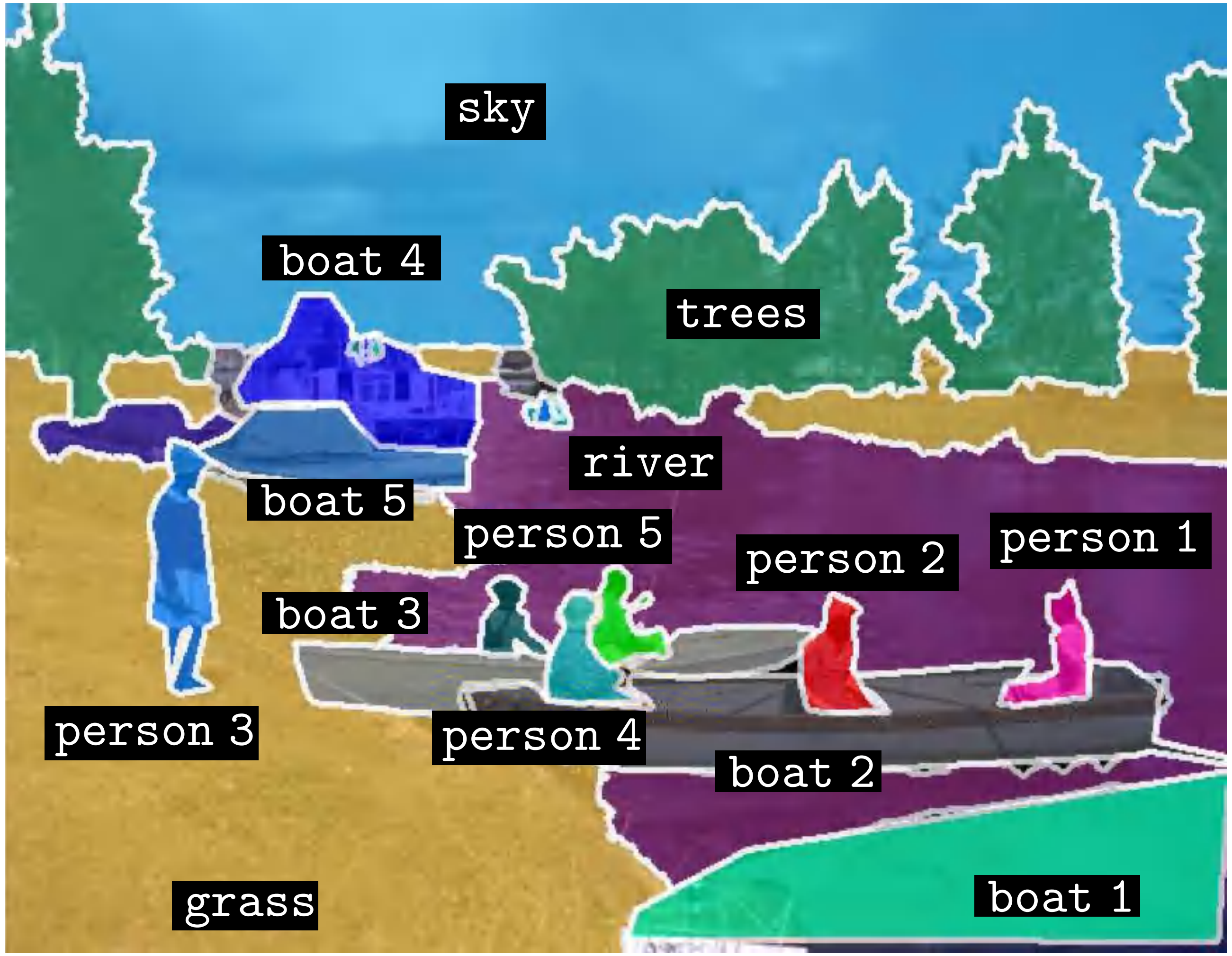
combined with instance segmentation prediction

What do semantic segmentation methods "see"?



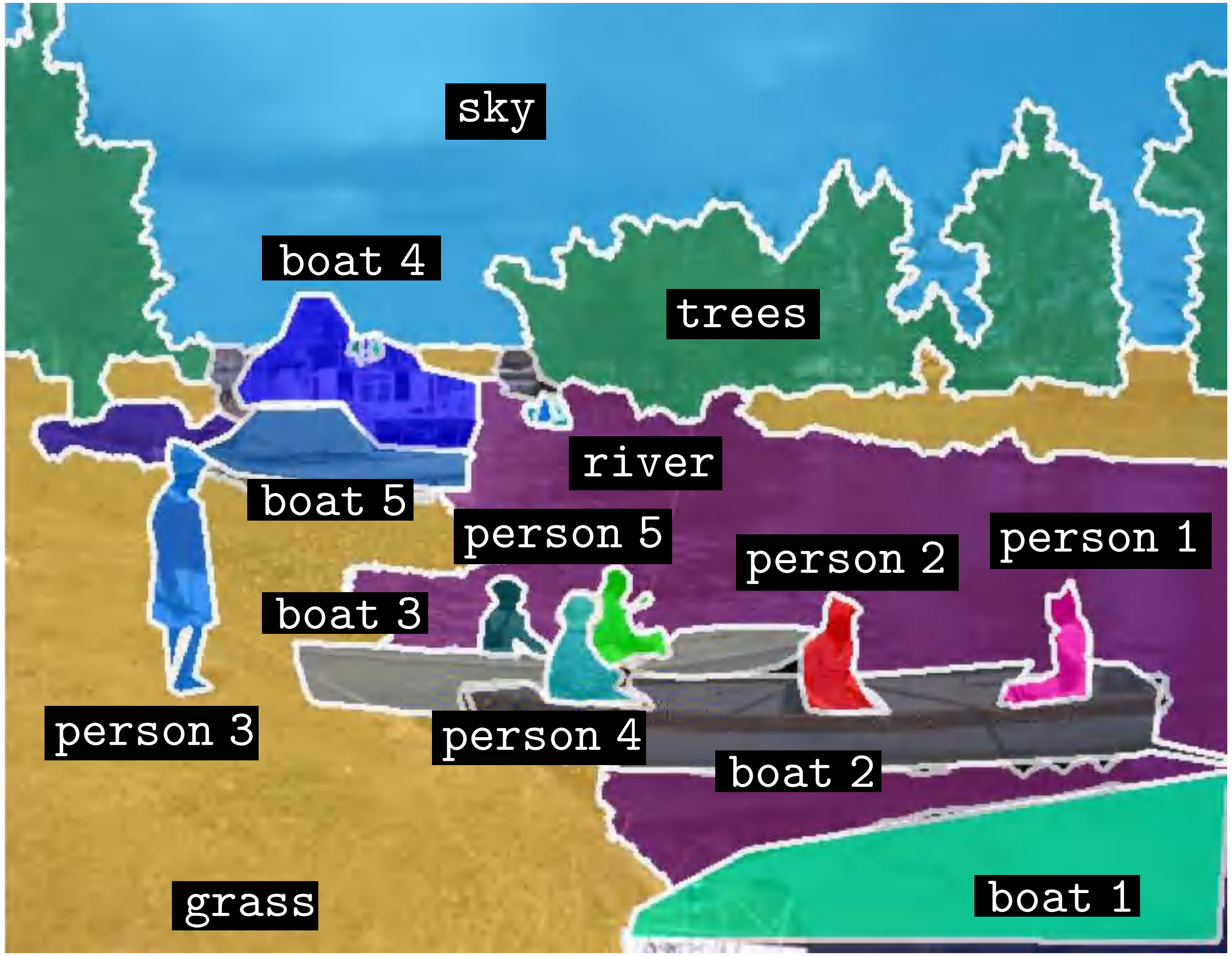
combined with instance segmentation prediction

Unified segmentation task



single task that combines semantic and instance segmentation

Unified segmentation task



single task that combines semantic and instance segmentation

things: categories with instance-level annotation (person, boat)
stuff: categories without the notion of instances (sky, road)

Unified segmentation task

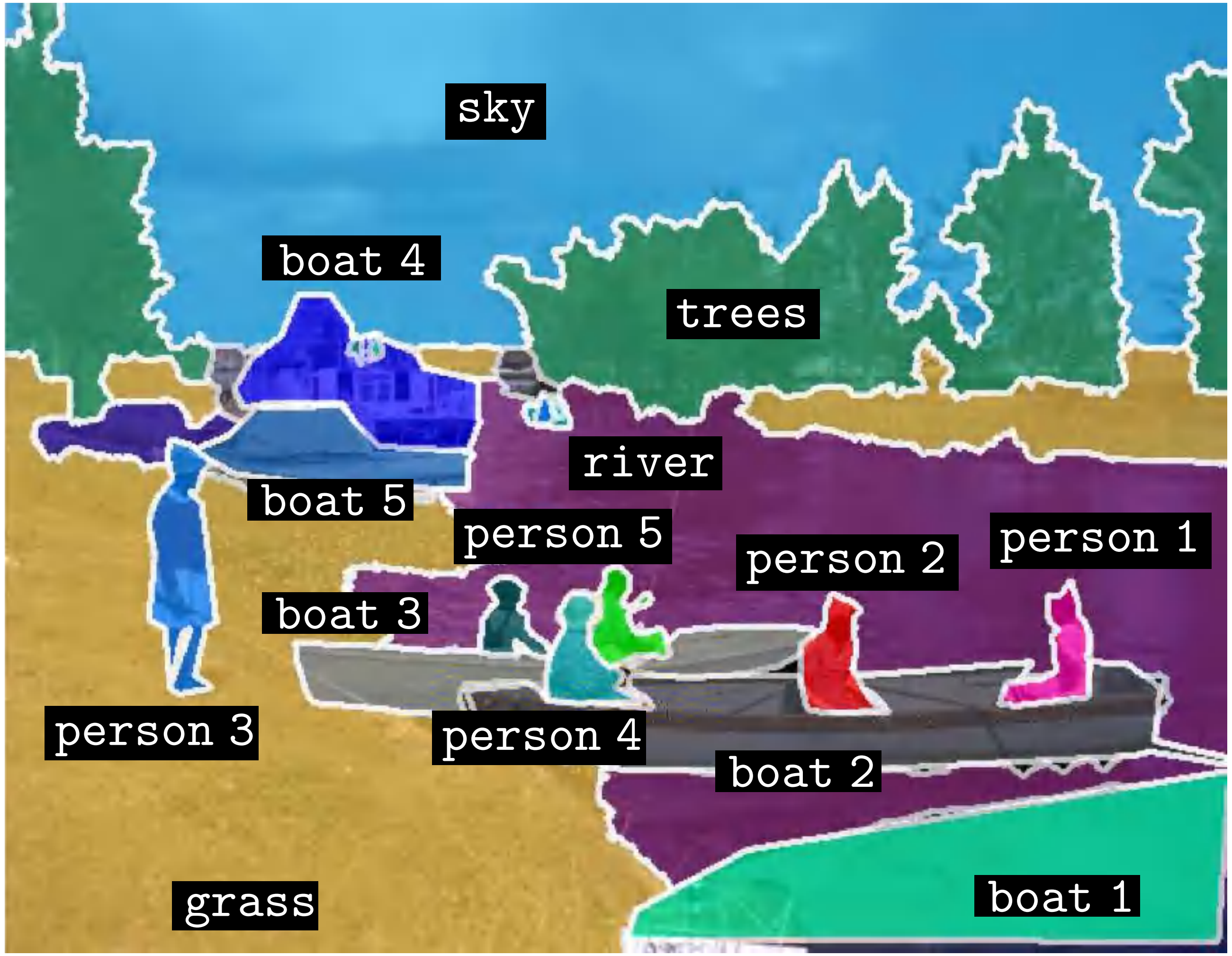
1. Tu et al. Image parsing: Unifying segmentation, detection, and recognition, IJCV 2005
2. Yao et al. Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation, CVPR 2012
3. Tighe et al. Finding things: **Image parsing** with regions and per-exemplar detectors, CVPR 2013
4. Tighe et al. **Scene parsing** with object instances and occlusion ordering, CVPR 2014
5. Sun et al. Relating things and stuff via object property interactions, PAMI 2014
6. Kirillov et al. **Panoptic segmentation**, CVPR 2019

Unified segmentation task

1. Tu et al. Image parsing: Unifying segmentation, detection, and recognition, IJCV 2005
2. Yao et al. Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation, CVPR 2012
3. Tighe et al. Finding things: **Image parsing** with regions and per-exemplar detectors, CVPR 2013
4. Tighe et al. **Scene parsing** with object instances and occlusion ordering, CVPR 2014
5. Sun et al. Relating things and stuff via object property interactions, PAMI 2014
6. Kirillov et al. Panoptic segmentation, CVPR 2019

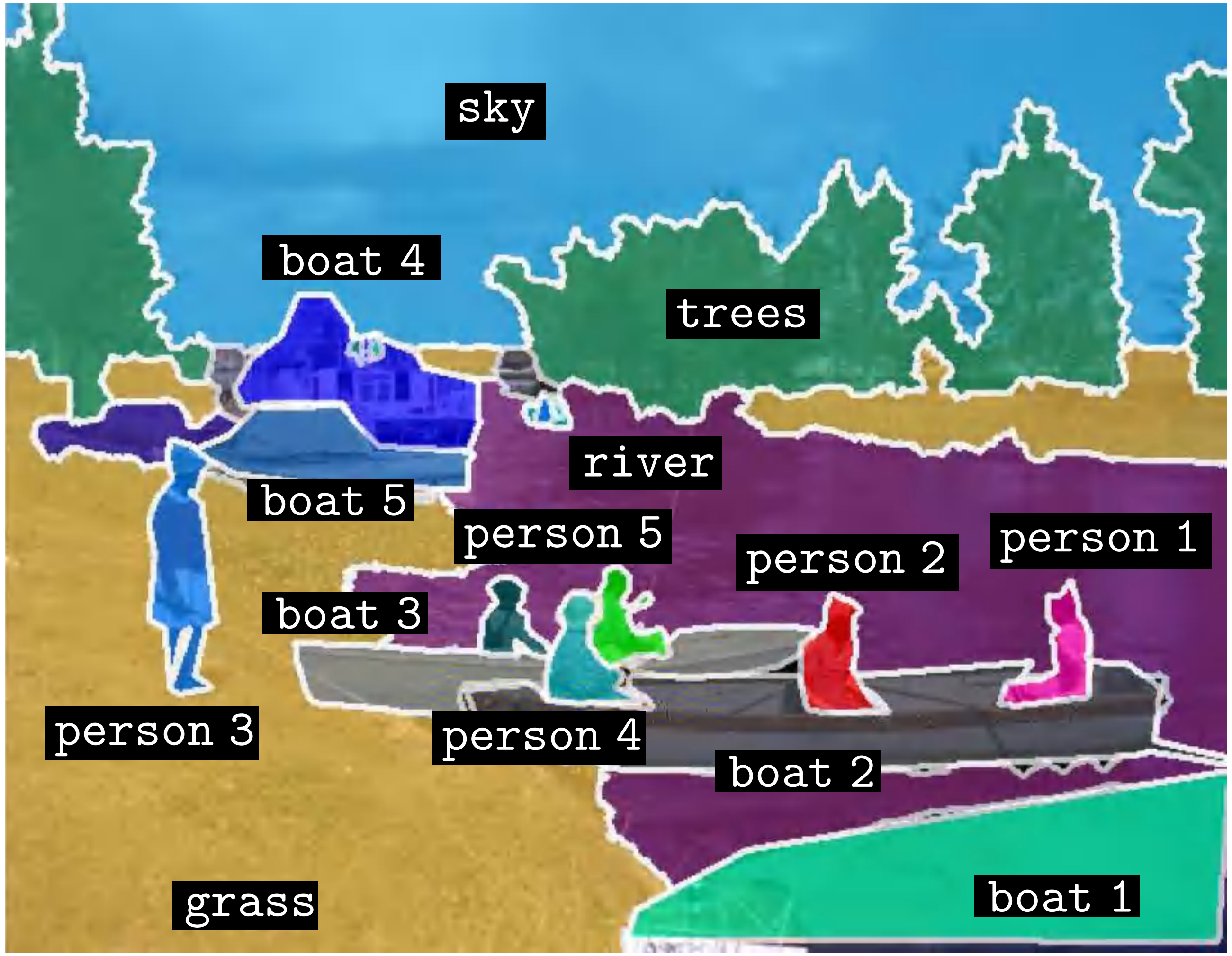
"panoptic" – seeing everything at once

Panoptic segmentation



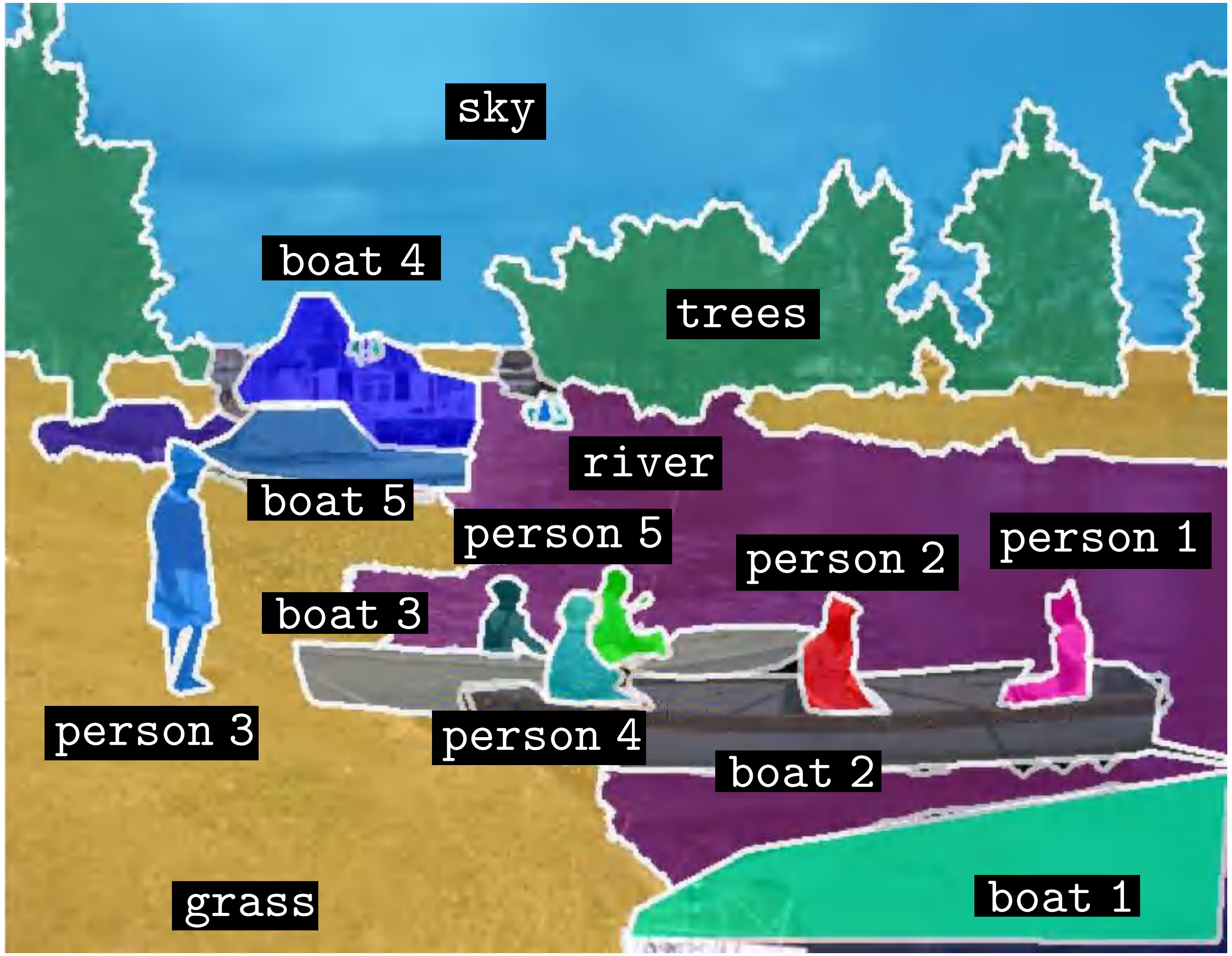
task: ?

Panoptic segmentation



assign semantic labels to pixels + segment each instance separately

Panoptic segmentation



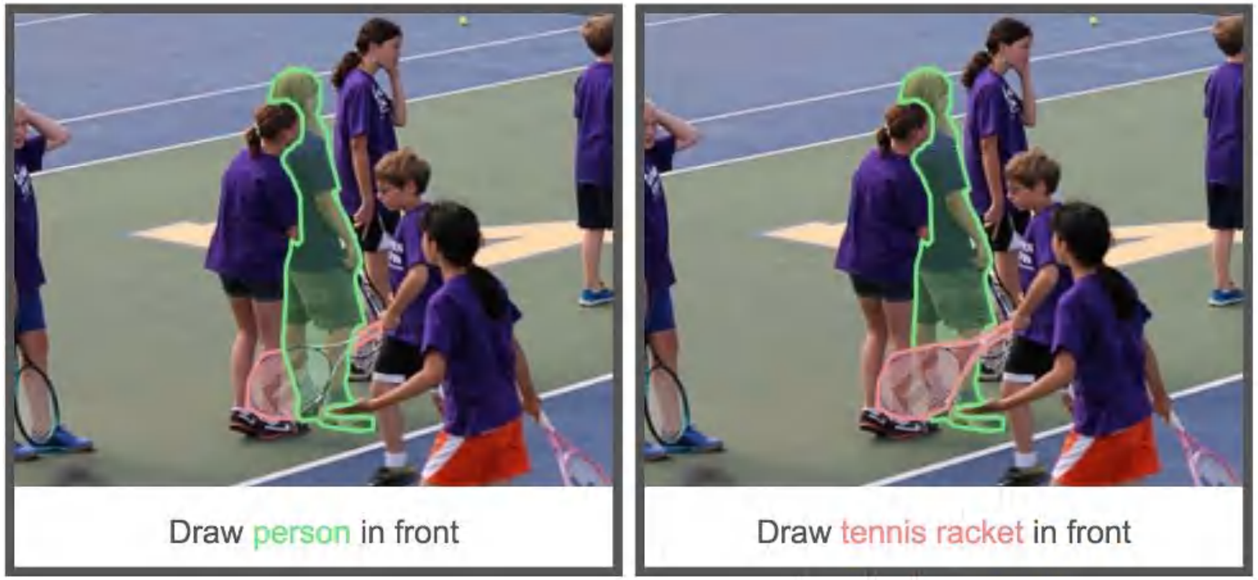
generalization of both semantic and instance segmentation tasks

Overlapping Segments

instance segmentation formulation
allows overlapping instances

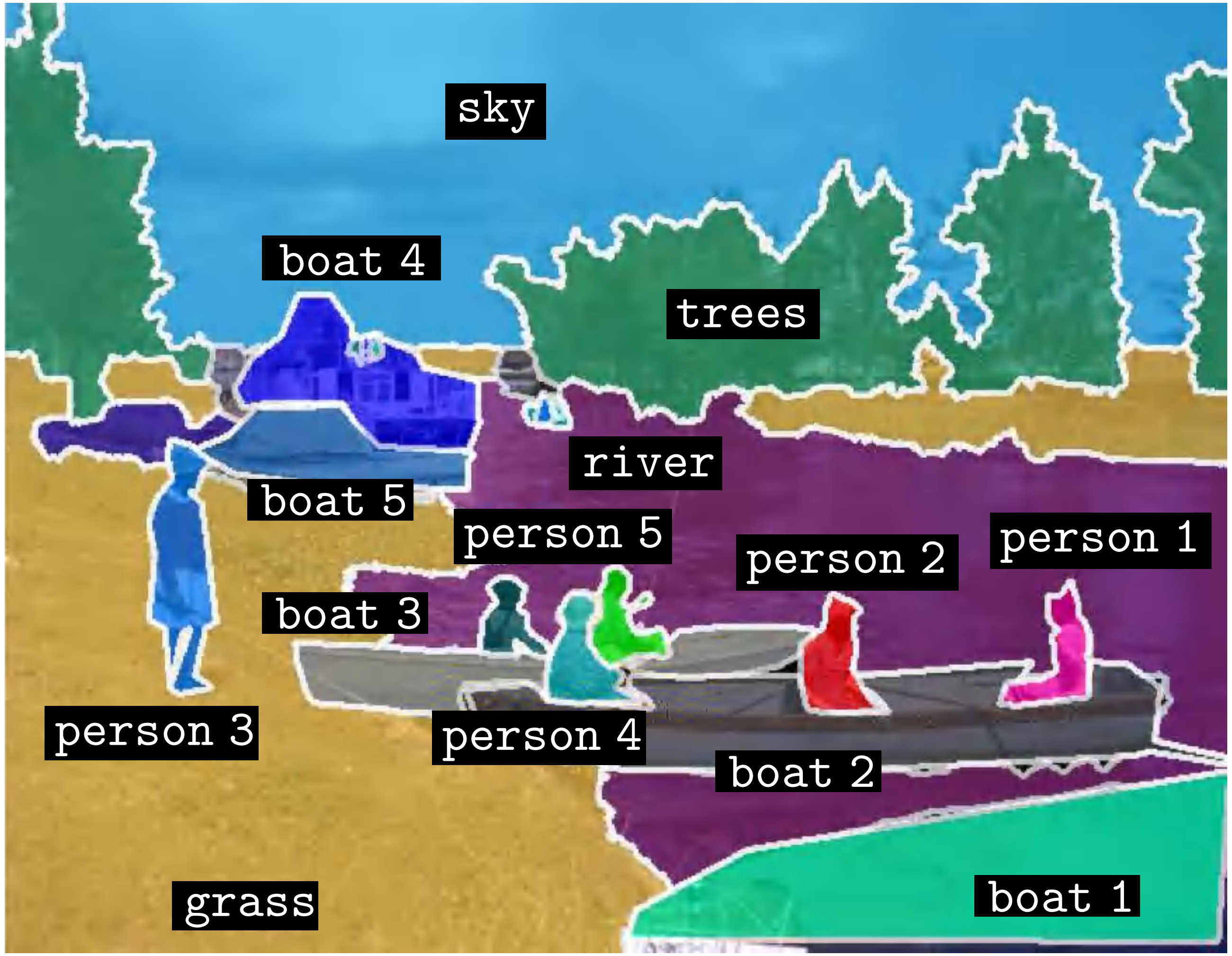
Overlapping Segments

instance segmentation formulation
allows overlapping instances



in panoptic segmentation
each pixel has only one label

Panoptic segmentation



task: ✓

datasets: ?

Available panoptic segmentation datasets

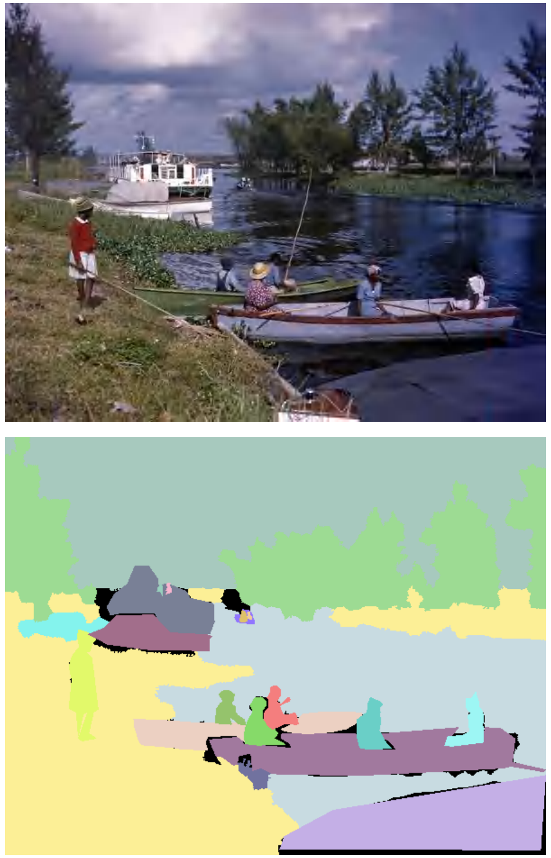
COCO (2014) + COCO-stuff (2017)
~200k images, 133 categories

Available panoptic segmentation datasets

COCO (2014) + COCO-stuff (2017)

COCO-panoptic challenges:
ECCV`18, **ICCV`19**

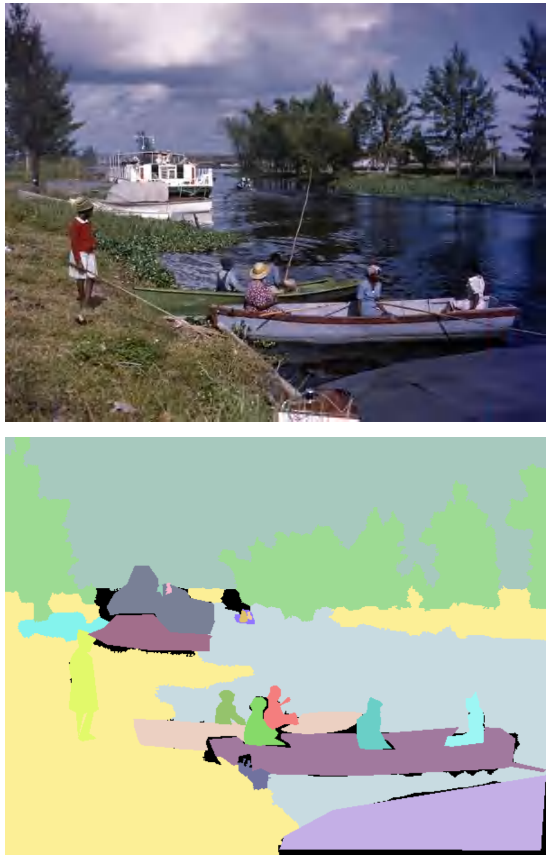
Available panoptic segmentation datasets



COCO (2014) + COCO-stuff (2017)
COCO-panoptic challenges:
ECCV`18, **ICCV`19**

Mapillary Vistas (2017)
~25k images, 66 categories

Available panoptic segmentation datasets



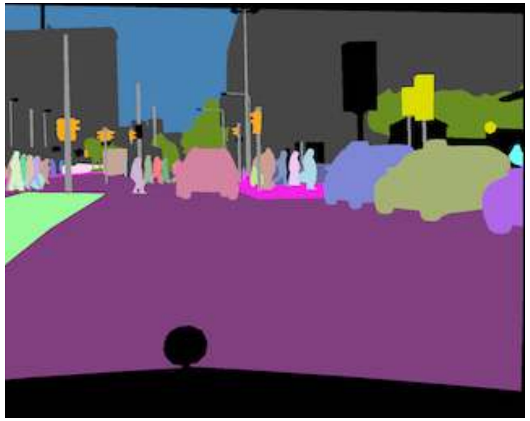
COCO (2014) + COCO-stuff (2017)
COCO-panoptic challenges:
ECCV`18, **ICCV`19**

Mapillary Vistas (2017)
Vistas-panoptic challenges:
ECCV`18, **ICCV`19**

Available panoptic segmentation datasets

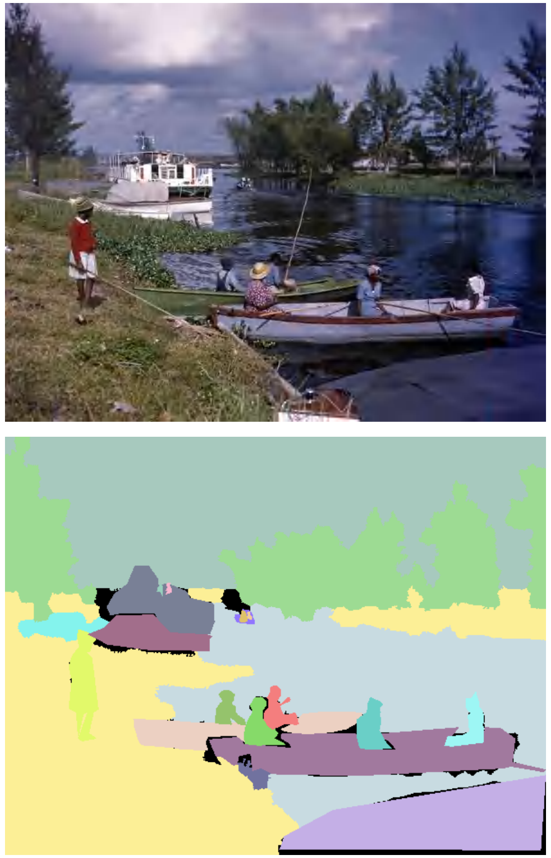
COCO (2014) + COCO-stuff (2017)
COCO-panoptic challenges:
ECCV`18, **ICCV`19**

Mapillary Vistas (2017)
Vistas-panoptic challenges:
ECCV`18, **ICCV`19**



Cityscapes (2015)
5k images, 19 categories

Available panoptic segmentation datasets

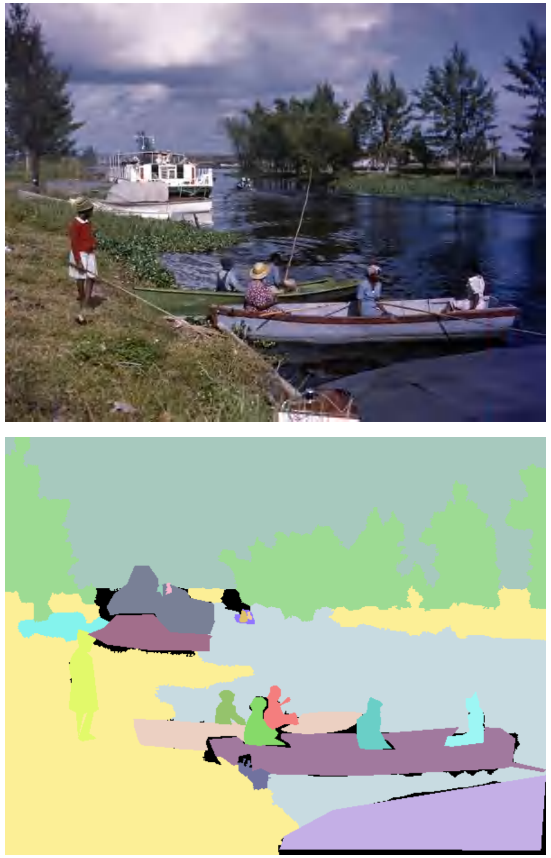


COCO (2014) + COCO-stuff (2017)
COCO-panoptic challenges:
ECCV`18, **ICCV`19**

Mapillary Vistas (2017)
Vistas-panoptic challenges:
ECCV`18, **ICCV`19**

Cityscapes (2015)
panoptic test set
leaderboard (2019)

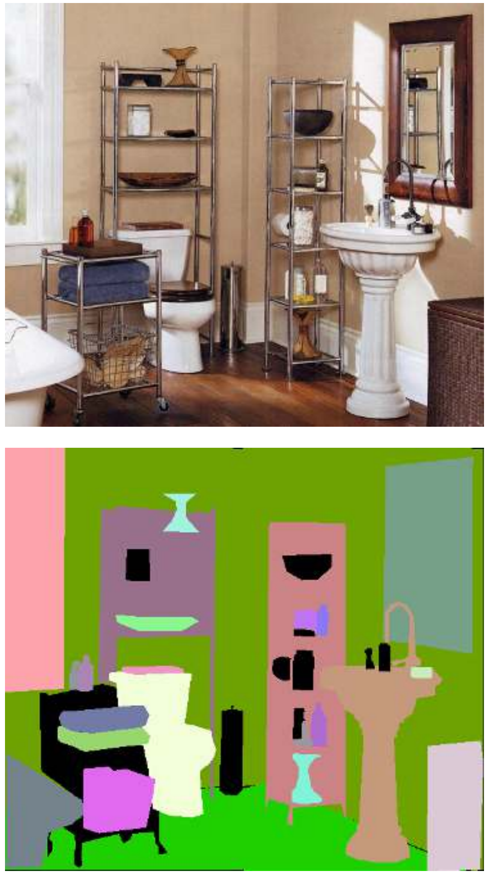
Available panoptic segmentation datasets



COCO (2014) + COCO-stuff (2017)
COCO-panoptic challenges:
ECCV`18, **ICCV`19**

Mapillary Vistas (2017)
Vistas-panoptic challenges:
ECCV`18, **ICCV`19**

Cityscapes (2015)
panoptic test set
leaderboard (2019)

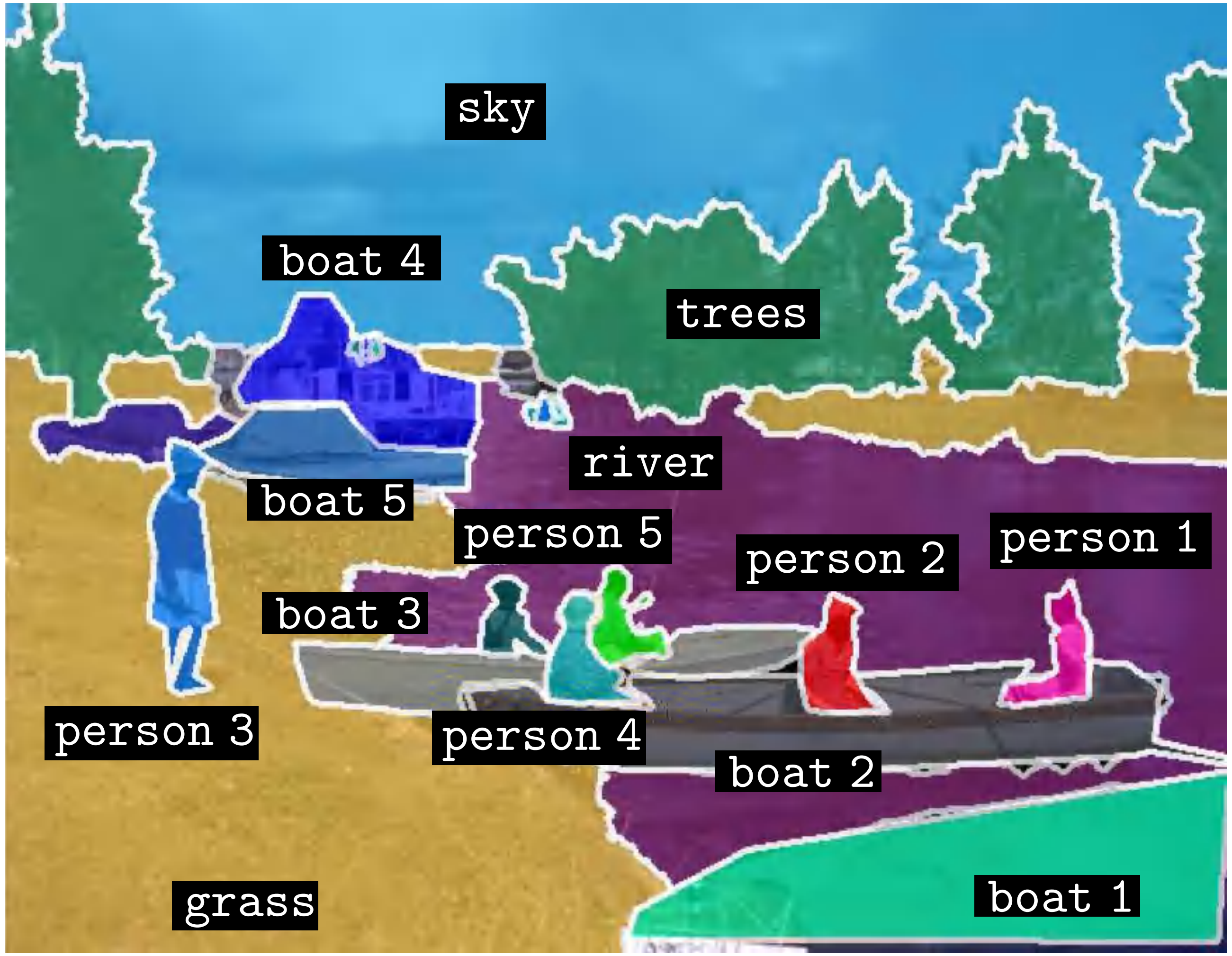


ADE20k (2016)
>22k images, 150 categories

Available panoptic segmentation datasets

1. Lin et al. Microsoft COCO: Common Objects in Context, ECCV 2014
2. Caesar et al. COCO-Stuff: Thing and Stuff Classes in Context, CVPR 2018
3. Neuhold et al. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes, ICCV 2017
4. Cordts et al. The Cityscapes Dataset for Semantic Urban Scene Understanding, CVPR 2016
5. Zhou et al. Semantic understanding of scenes through the ade20k dataset, IJCV 2016

Panoptic segmentation



task: ✓

datasets: ✓

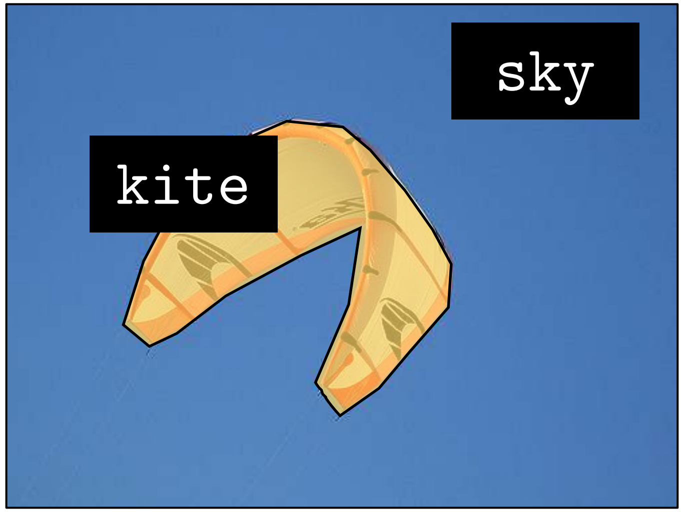
evaluation: ?

Image segmentation evaluation

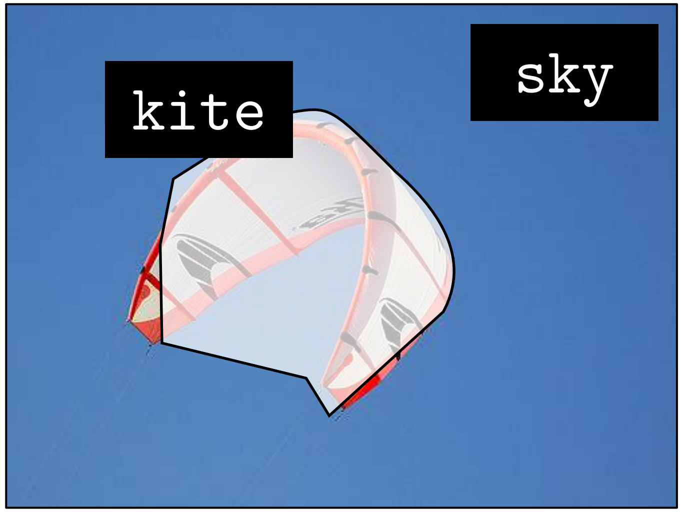
- semantic segmentation
  - Intersection-over-union (IoU), per-pixel metric

Image segmentation evaluation

- semantic segmentation
  - Intersection-over-union (IoU), per-pixel metric



ground truth



prediction

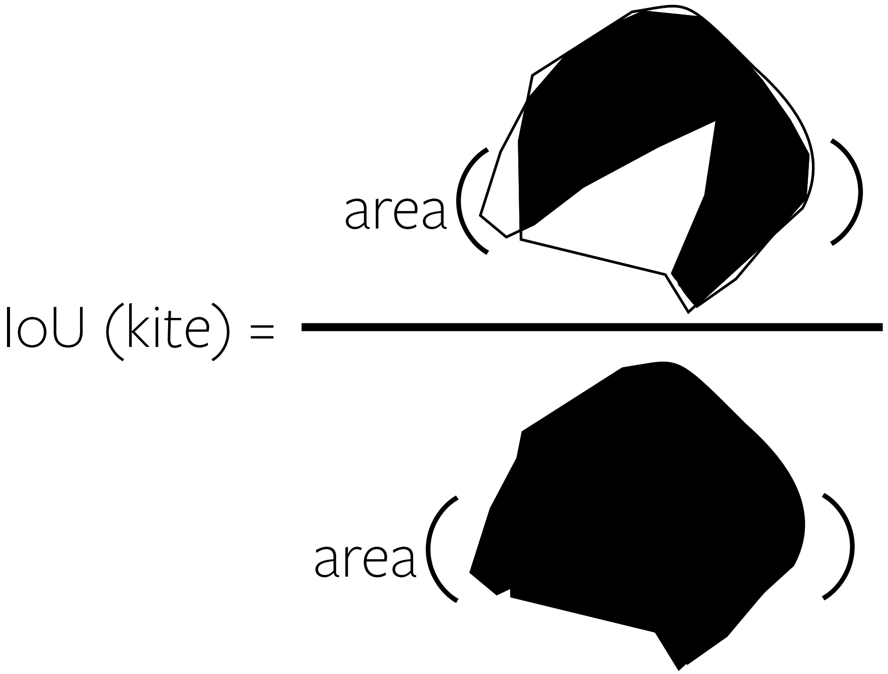
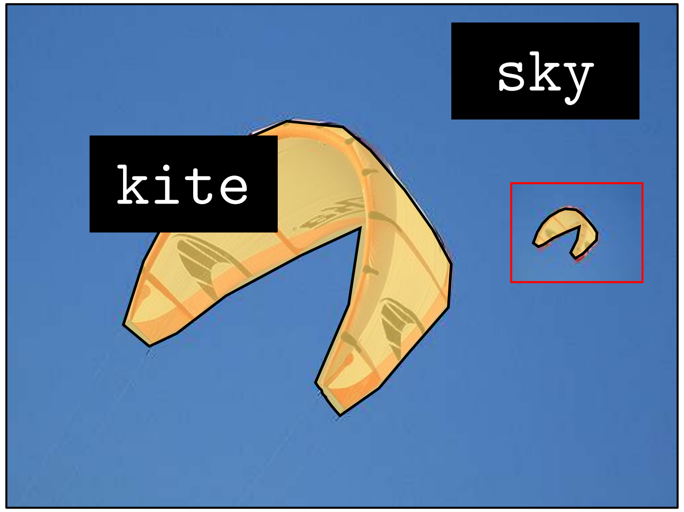
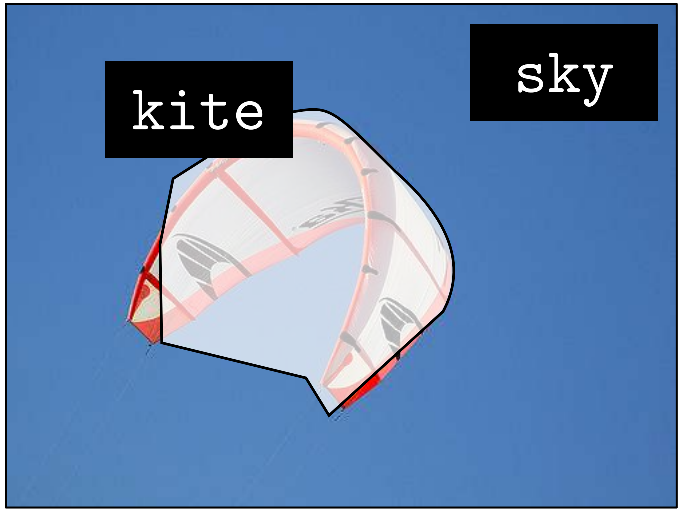


Image segmentation evaluation

- semantic segmentation
  - Intersection-over-union (IoU), per-pixel metric



ground truth



prediction

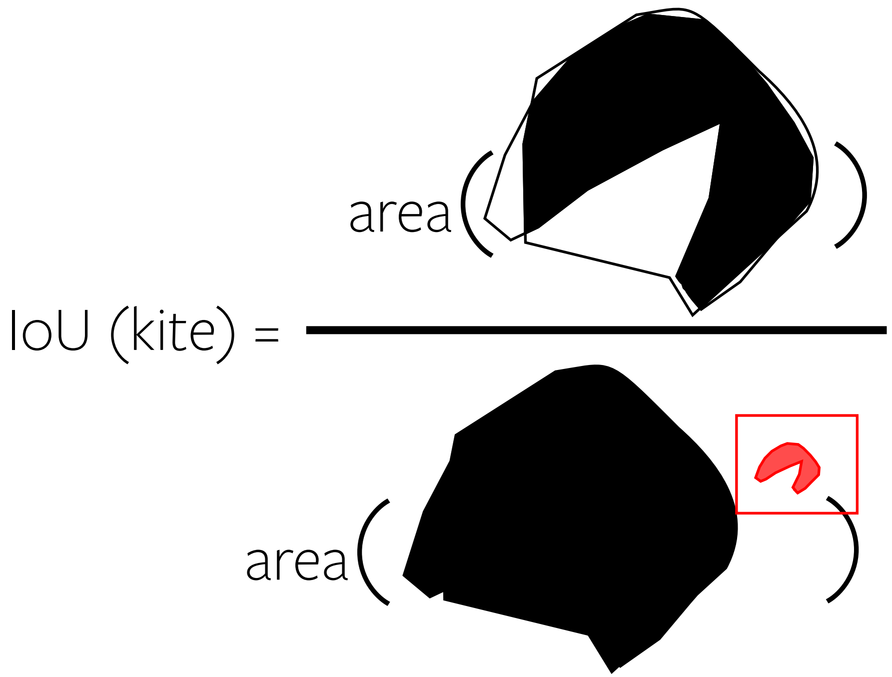
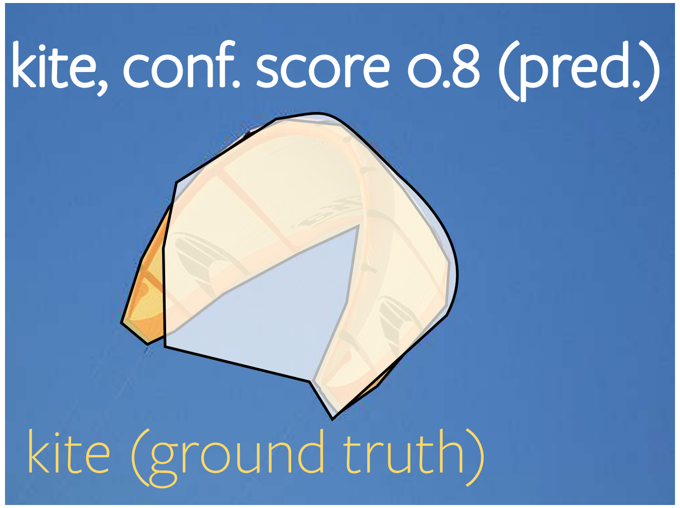


Image segmentation evaluation

- semantic segmentation
  - intersection-over-union (IoU), per-pixel metric
- instance segmentation
  - average precision (AP) over several IoU thresholds (0.5:0.05:0.95), object size-agnostic

Image segmentation evaluation

- semantic segmentation
  - intersection-over-union (IoU), per-pixel metric
- instance segmentation
  - average precision (AP) over several IoU thresholds (0.5:0.05:0.95), object size-agnostic



IoU = 0.75
> threshold?

yes → true positive

no → false positive + false negative

Image segmentation evaluation

- semantic segmentation
  - intersection-over-union (IoU), per-pixel metric
- instance segmentation
  - average precision (AP) over several IoU thresholds (0.5:0.05:0.95), object size-agnostic
- panoptic segmentation

neither IoU nor AP alone works
for panoptic segmentation

Image segmentation evaluation

- semantic segmentation
  - intersection-over-union (IoU), per-pixel metric
- instance segmentation
  - average precision (AP) over several IoU thresholds (0.5:0.05:0.95), object size-agnostic
- panoptic segmentation
  - IoU + AP

asymmetric for classes with and without instance-level annotation

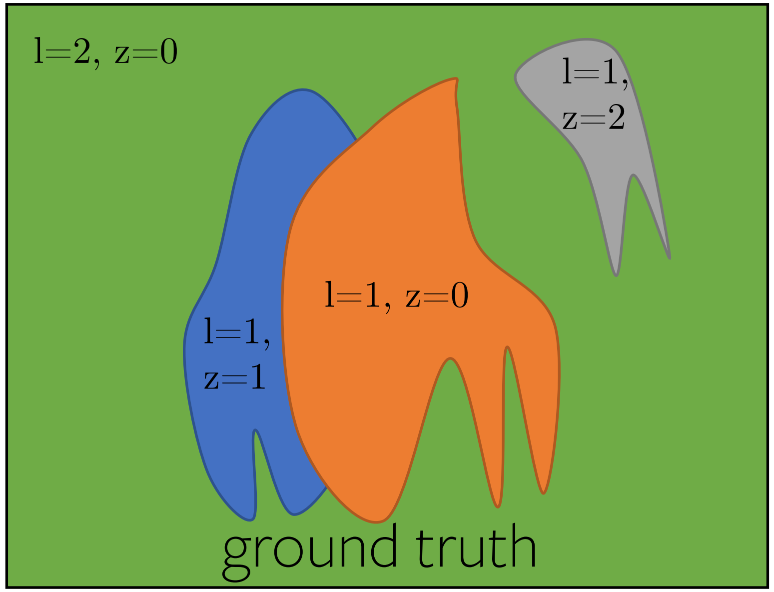
Image segmentation evaluation

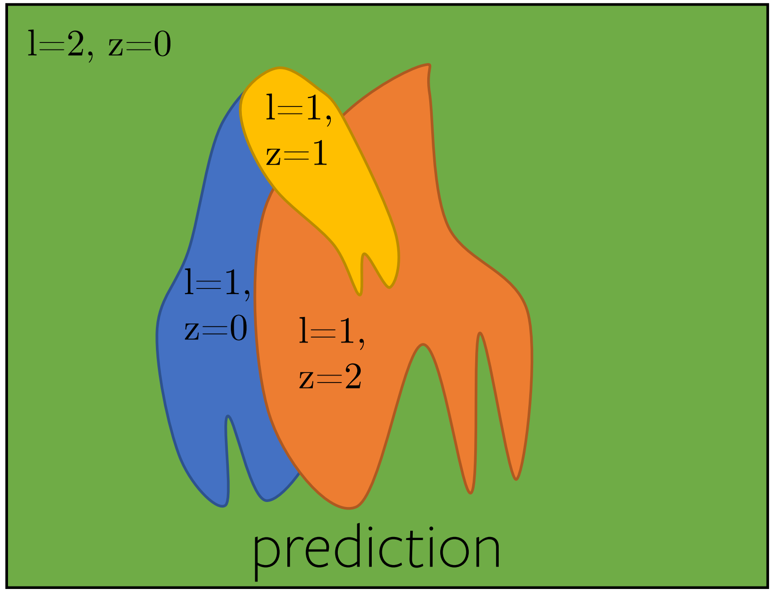
- semantic segmentation
  - intersection-over-union (IoU), per-pixel metric
- instance segmentation
  - average precision (AP) over several IoU thresholds (0.5:0.05:0.95), object size-agnostic
- panoptic segmentation
  - IoU + AP
  - **panoptic quality (PQ)**, segment size-agnostic

metric that treats all
categories in the same way

Kirillov et al. Panoptic segmentation, CVPR`19

Panoptic quality (PQ) measure

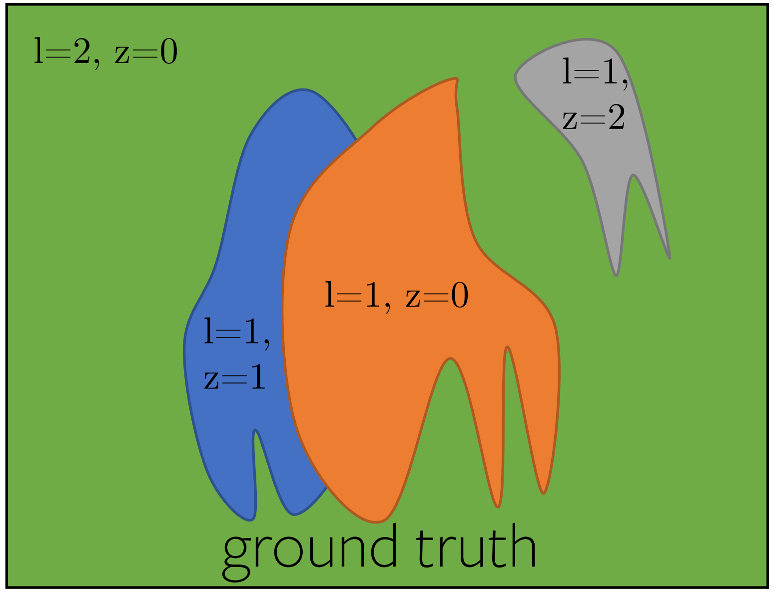


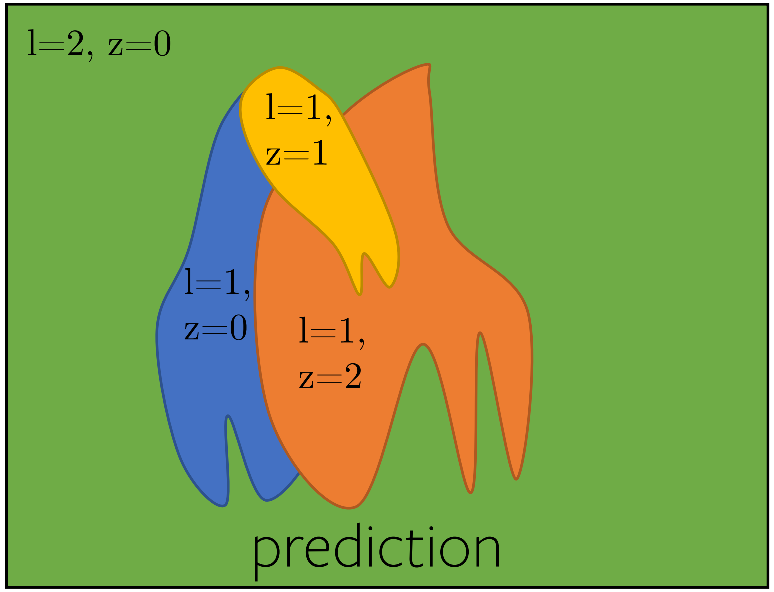


PQ computation:

- matching
- calculation

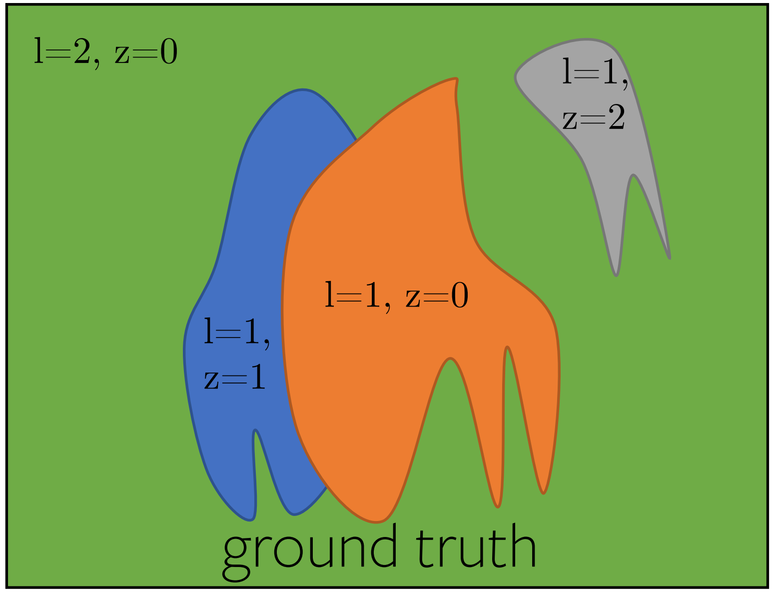
Panoptic quality (PQ) measure

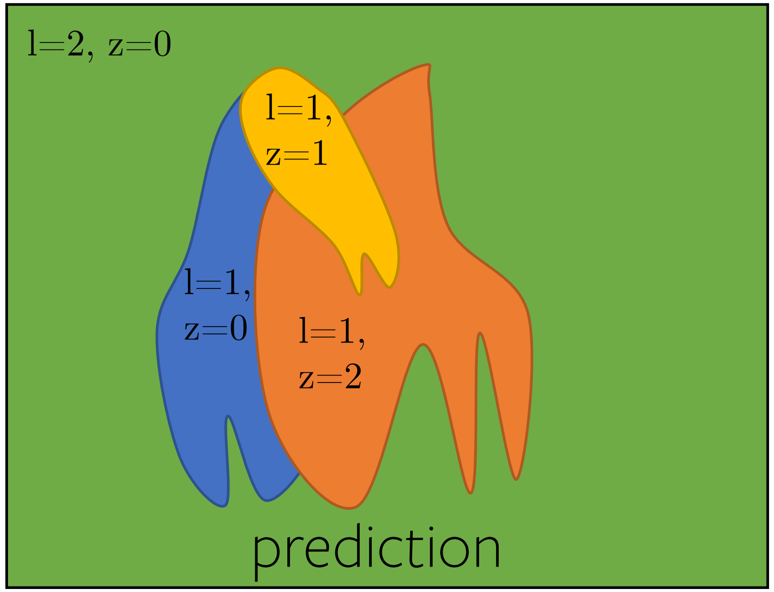




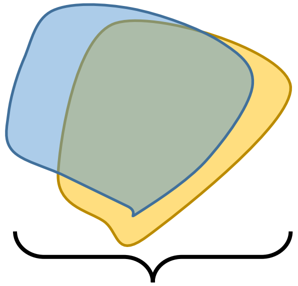
- matching rule: two segments match if their IoU > 0.5

Panoptic quality (PQ) measure



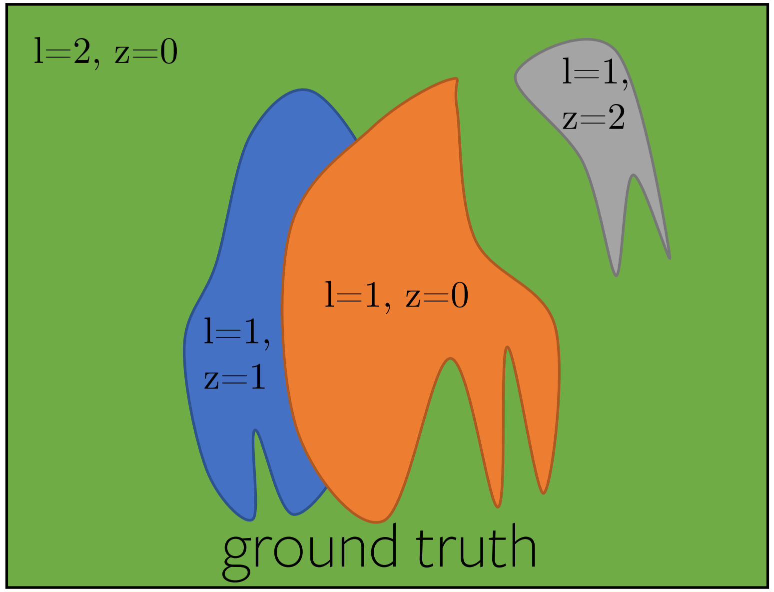


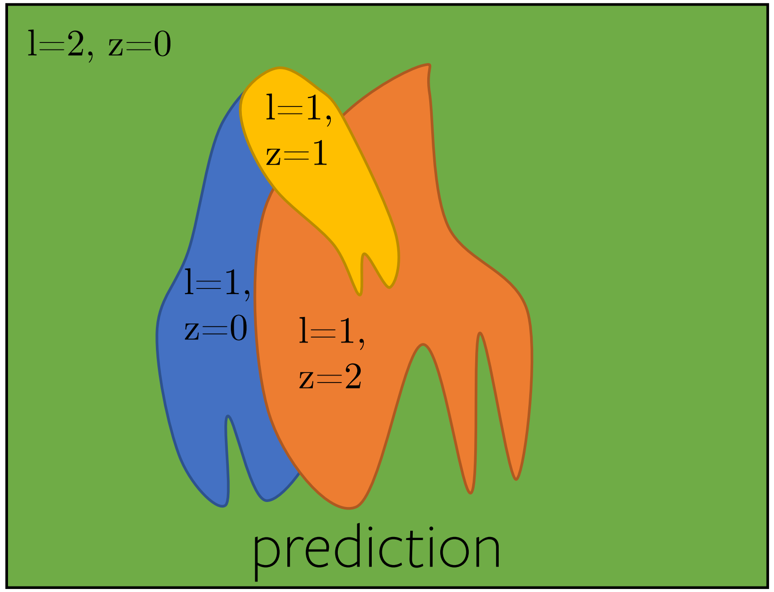
- matching rule: two segments match if their IoU > 0.5
- the matching is unique:

if  then there is no other non overlapping object that has IoU > 0.5

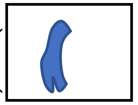
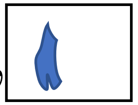
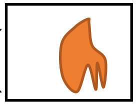
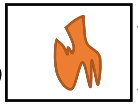
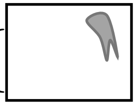
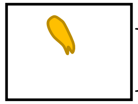
IoU > 0.5

Panoptic quality (PQ) measure

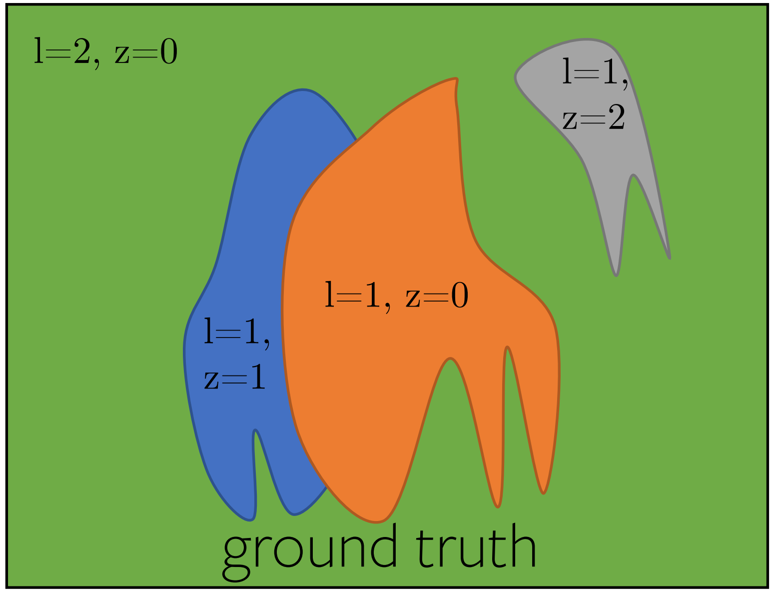


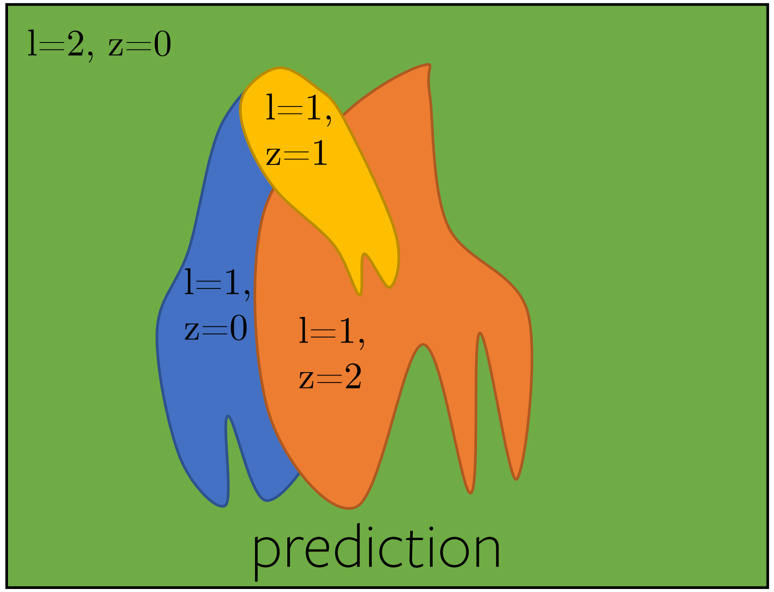


- matching rule: two segments match if their IoU > 0.5

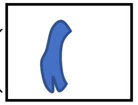
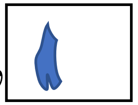
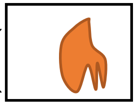
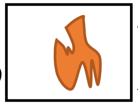
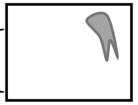
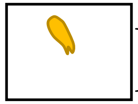
$TP = \{($  ,  $), ($  ,  $)\}, \; FN = \{$  $\}, \; FP = \{$  $\}$

Panoptic quality (PQ) measure





- matching rule: two segments match if their IoU > 0.5

$TP = \{($  ,  $), ($  ,  $)\},\ FN = \{$  $\},\ FP = \{$  $\}$

- calculation:

$$\mathbf{PQ} = \frac{\sum_{(p,g)\in TP} \mathbf{IoU}(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$$

Panoptic quality (PQ) measure

$$\text{PQ} = \frac{\sum_{(p,g) \in TP} \text{IoU}(p, g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$$

Panoptic quality (PQ) measure

$$\mathbf{PQ} = \frac{\sum_{(p,g)\in TP} \mathbf{IoU}(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} = \underbrace{\frac{\sum_{(p,g)\in TP} \mathbf{IoU}(p,g)}{|TP|}}_{\text{Segmentation Quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{Recognition Quality (RQ)}}$$

Panoptic quality (PQ) measure

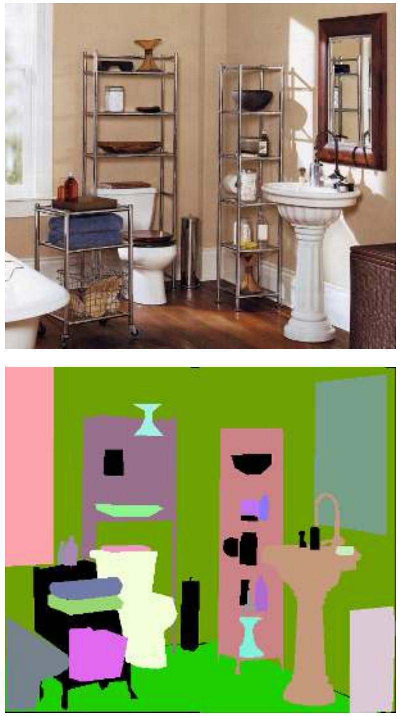
$$\mathrm{PQ} = \frac{\sum_{(p,g)\in TP} \mathbf{IoU}(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} = \underbrace{\frac{\sum_{(p,g)\in TP} \mathbf{IoU}(p,g)}{|TP|}}_{\text{Segmentation Quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{Recognition Quality (RQ)}}$$

- symmetric
- unified for categories with and without instance-level annotation (analysis)

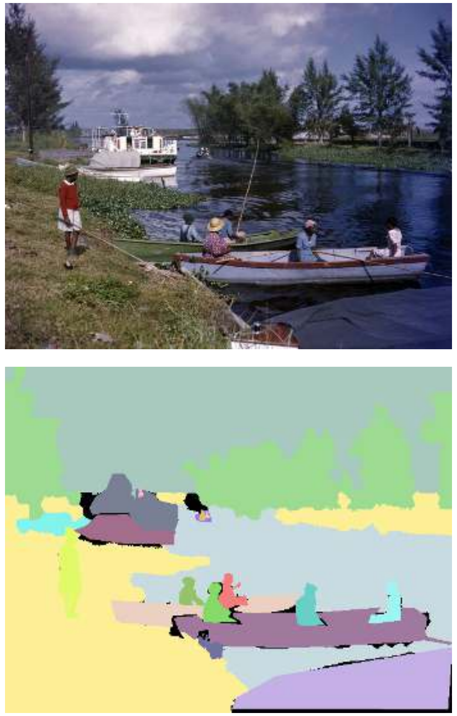
PQ analysis (human experimentation)

Cityscapes
30 images

Mapillary Vistas
46 images



ADE20k
64 images



COCO
5000 images

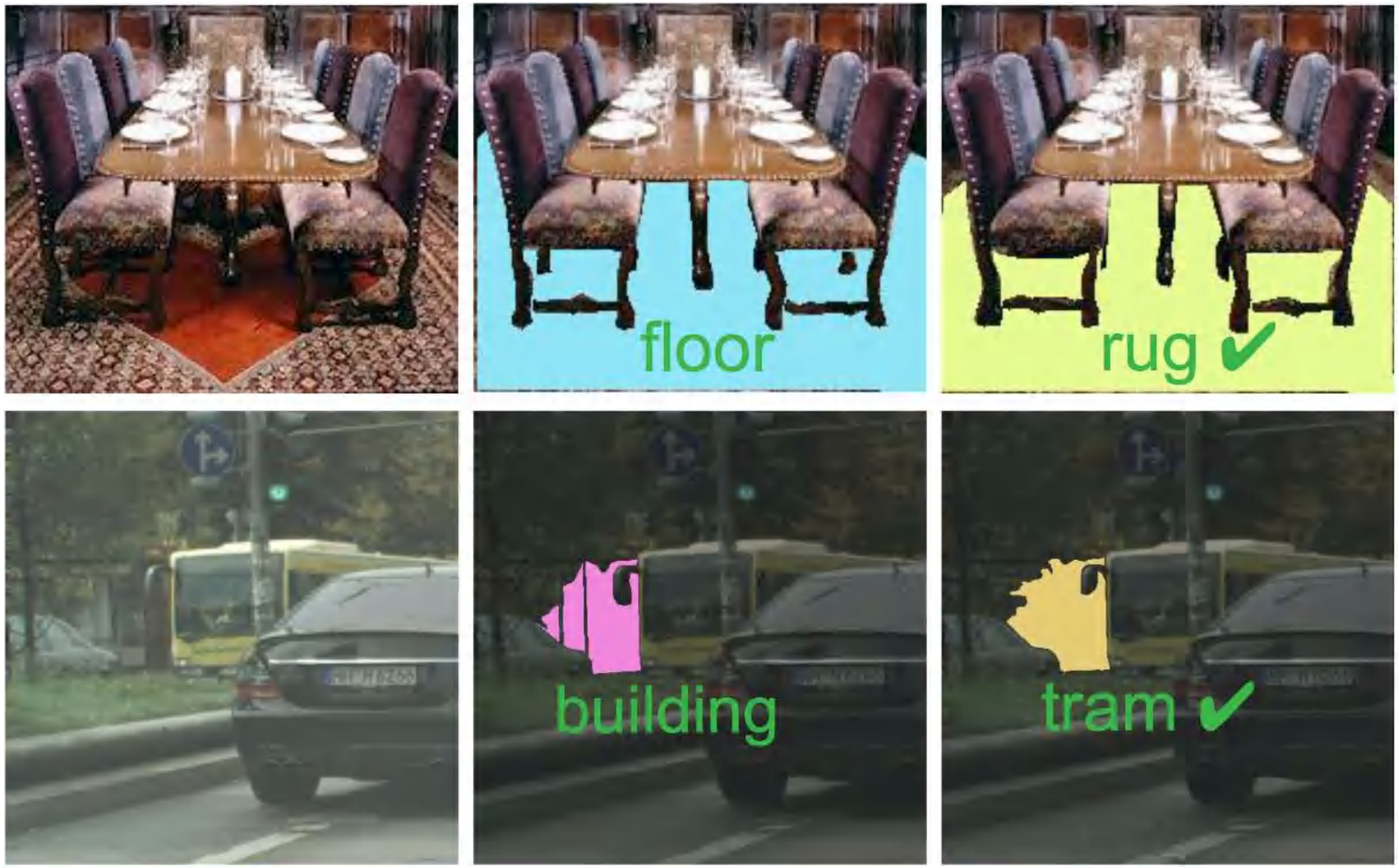
sets of images annotated twice independently

PQ analysis

inconsistency examples

annotator 1 annotator 2

PQ analysis



annotator 1 annotator 2

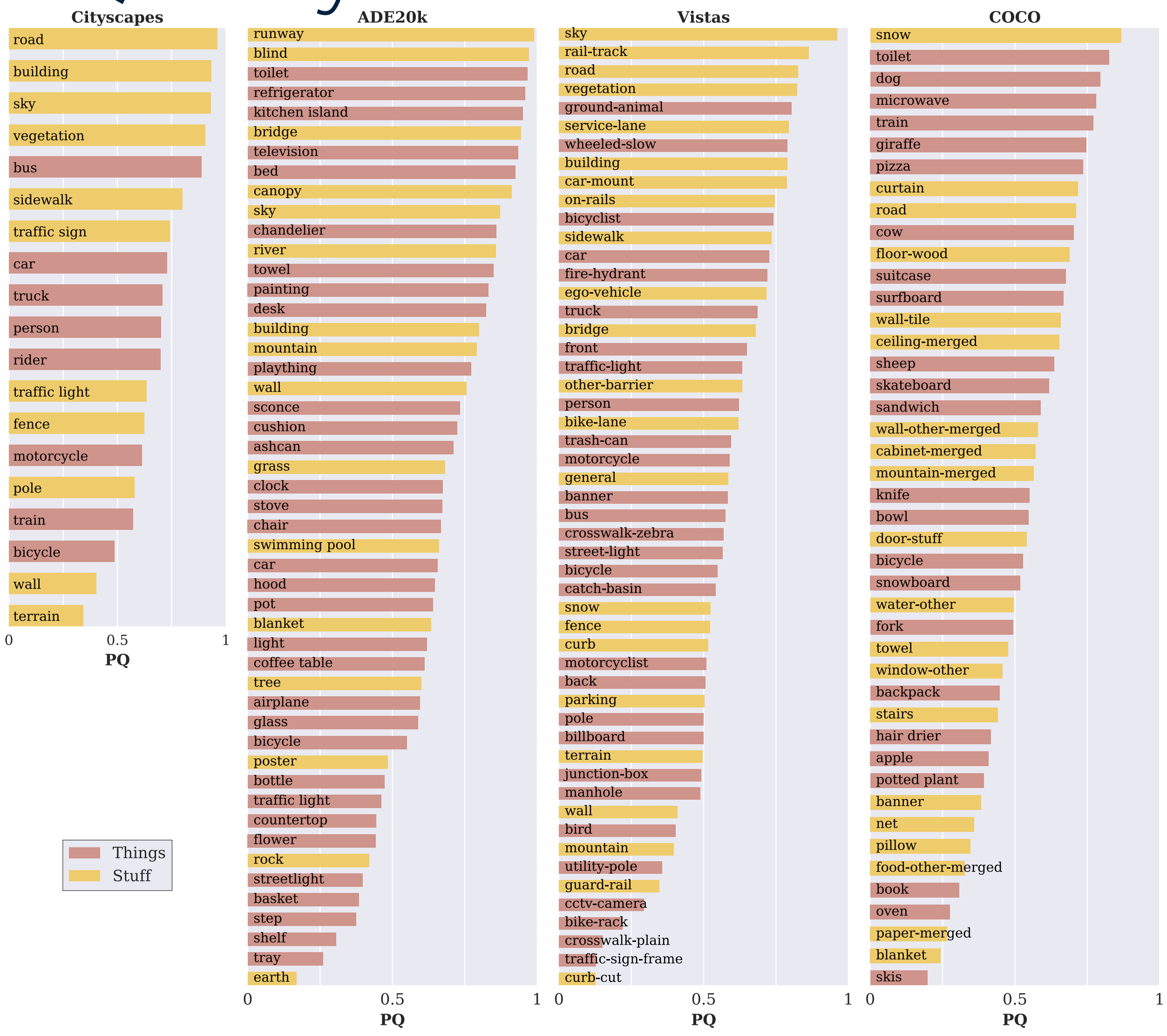
inconsistency examples

PQ analysis

	PQ	PQ^{St}	PQ^{Th}
Cityscapes	69.7	71.3	67.4
ADE20k	67.1	70.3	65.9
Vistas	57.5	62.6	53.4
COCO	53.5	47.1	57.8

PQ for stuff classes is close to PQ for things classes

PQ analysis



things and stuff are distributed evenly

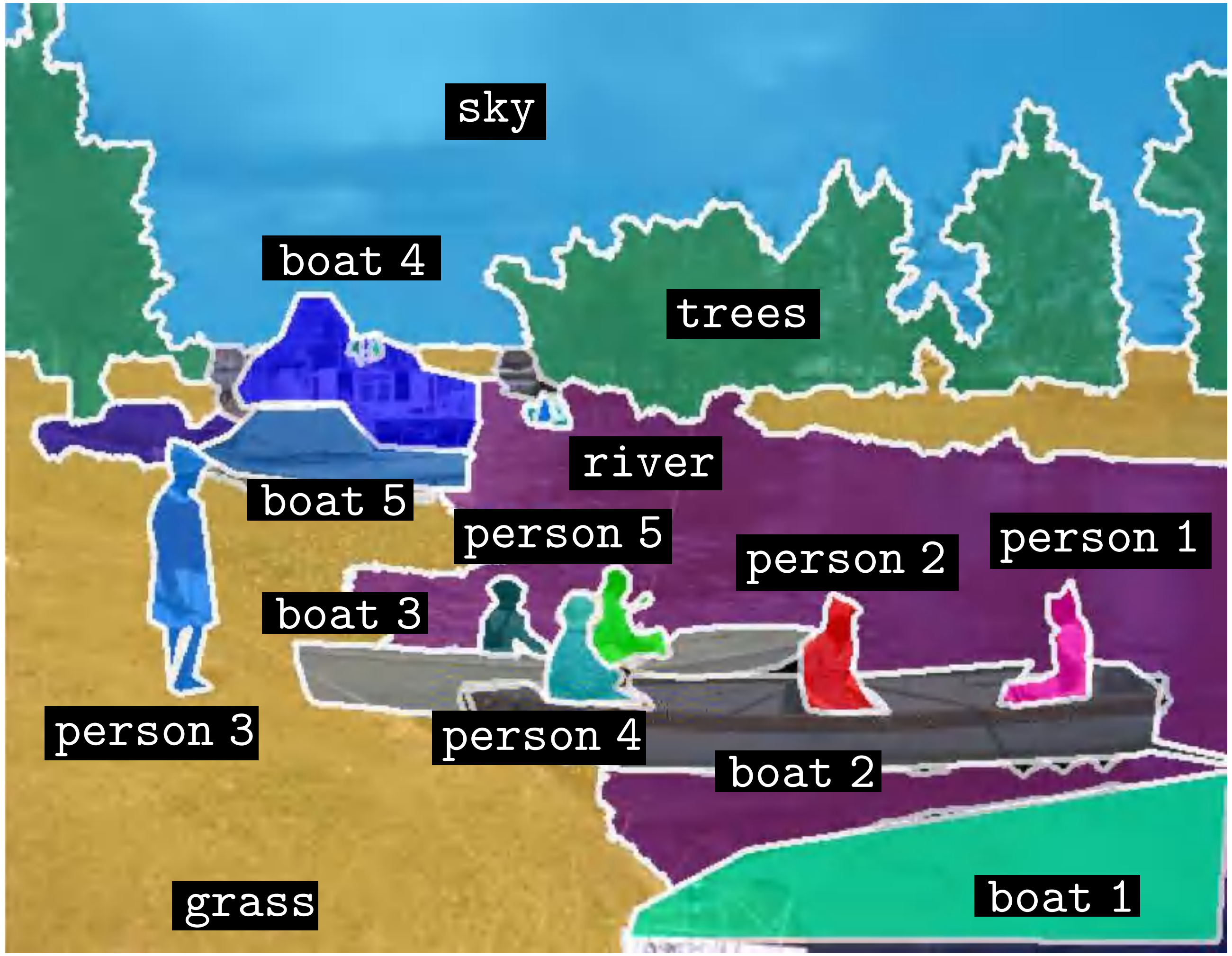
Panoptic quality (PQ) measure

$$\text{PQ} = \frac{\sum_{(p,g)\in TP} \text{IoU}(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} = \underbrace{\frac{\sum_{(p,g)\in TP} \text{IoU}(p,g)}{|TP|}}_{\text{Segmentation Quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{Recognition Quality (RQ)}}$$

- symmetric
- unified for categories with and without instance-level annotation (analysis)

evaluation code: https://github.com/cocodataset/panopticapi

Panoptic segmentation

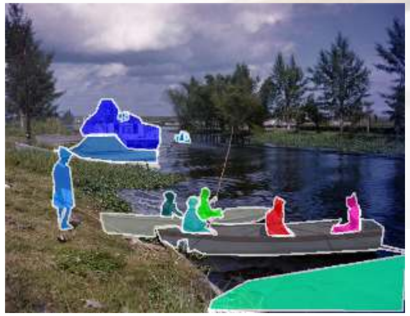


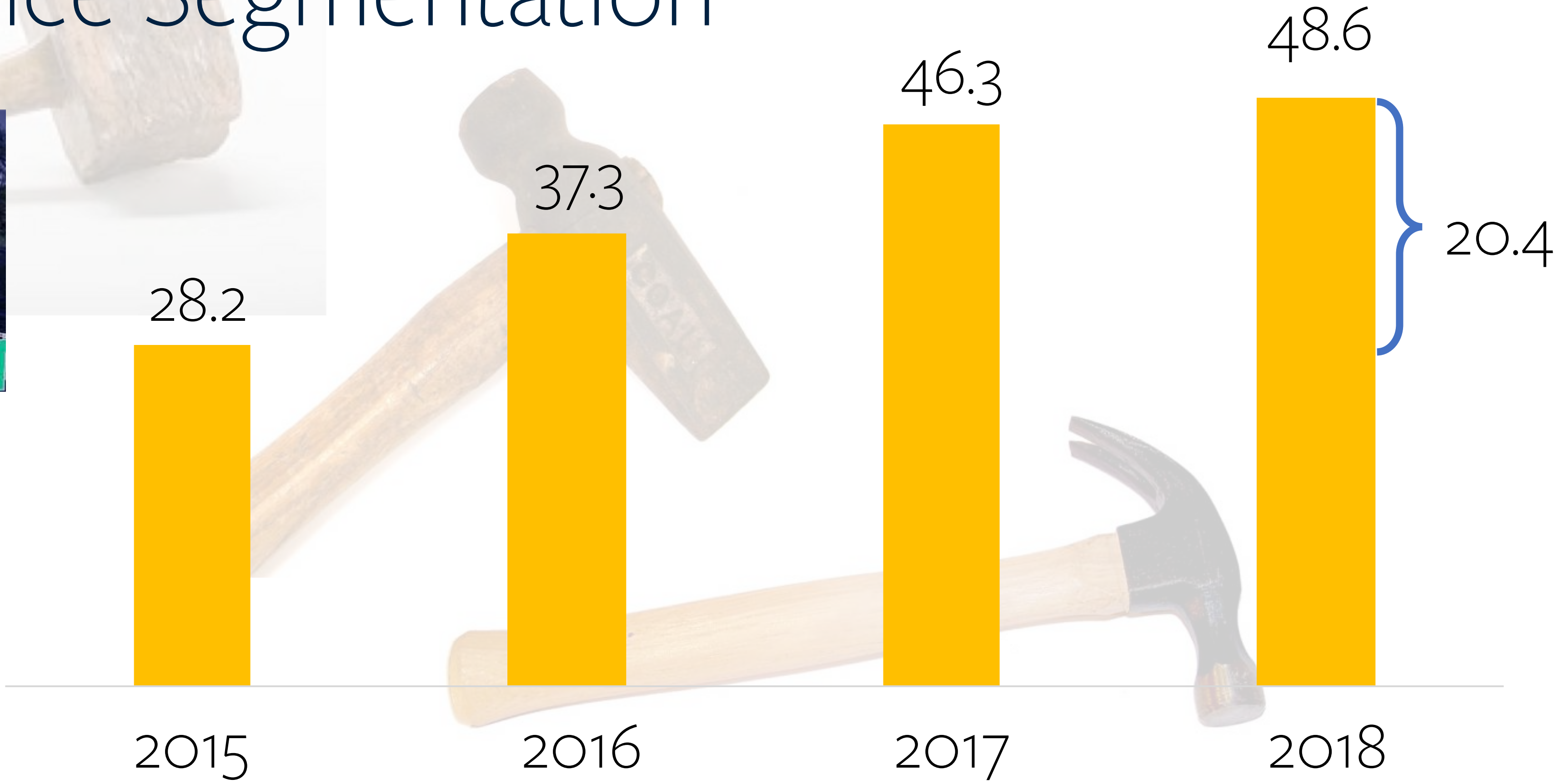
task: ✓
datasets: ✓
evaluation: ✓

In this tutorial:

- panoptic segmentation task – unified semantic segmentation task
- approaches for the task
  - instance segmentation (recap)
  - semantic segmentation (recap)
  - panoptic segmentation

Instance Segmentation

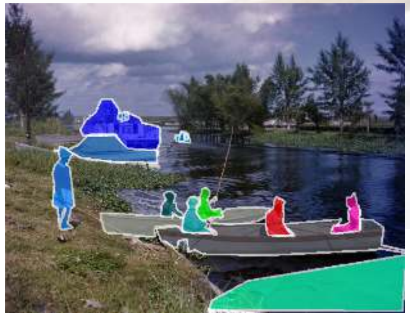




COCO-challenge winner instance segmentation AP (%)

Hammers credits:
Ross Girshick

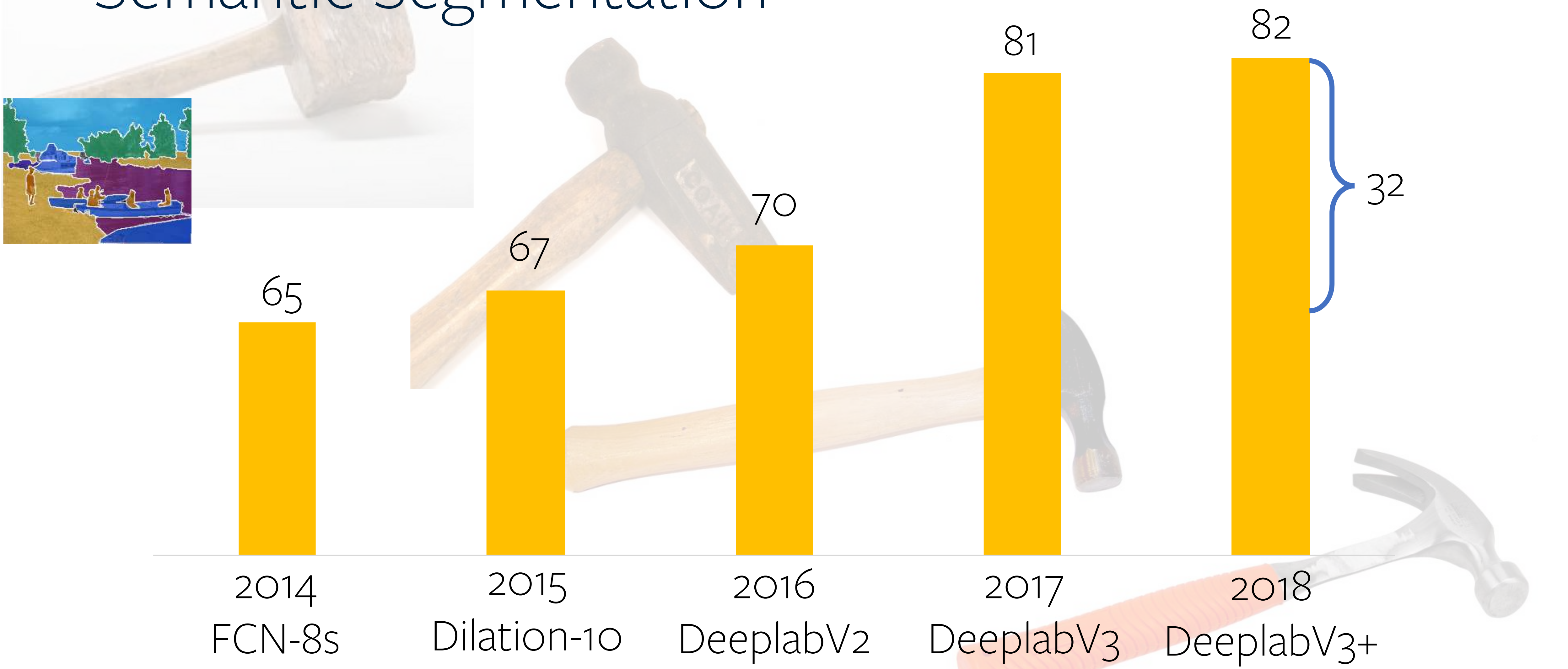
Instance Segmentation



COCO-challenge winner instance segmentation AP (%)

Hammers credits:
Ross Girshick

Semantic Segmentation



Cityscapes semantic segmentation IoU (%)

Hammers credits:
Ross Girshick

Cityscapes leaderboard
performance

Semantic Segmentation

1. Long et al. Fully Convolutional Networks for Semantic Segmentation, CVPR 2015
2. Yu et al. Multi-Scale Context Aggregation by Dilated Convolutions, ICLR 2016
3. Chen et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, TPAMI 2017
4. Chen et al. Rethinking Atrous Convolution for Semantic Image Segmentation, arXiv 2017
5. Chen et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation , ECCV 2018

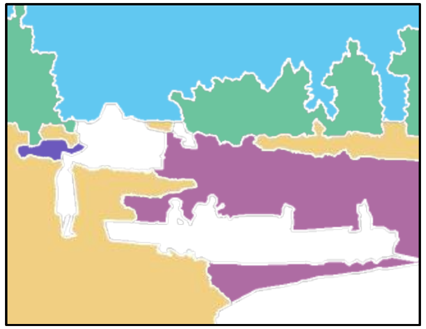
Hammers credits:
Ross Girshick

In this tutorial:

- panoptic segmentation task – unified semantic segmentation task
- approaches for the task
  - instance segmentation (recap)
  - semantic segmentation (recap)
  - panoptic segmentation

Panoptic segmentation: naïve approach

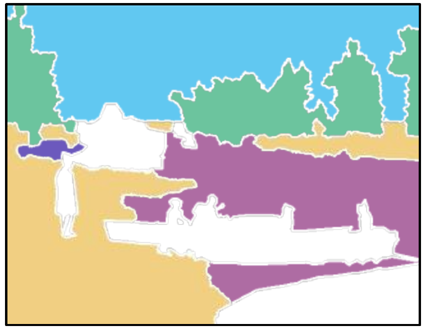
semantic segmentation



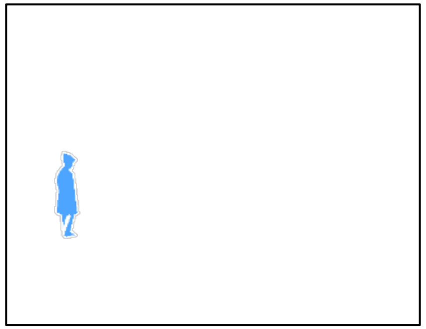
input

Panoptic segmentation: naïve approach

semantic segmentation

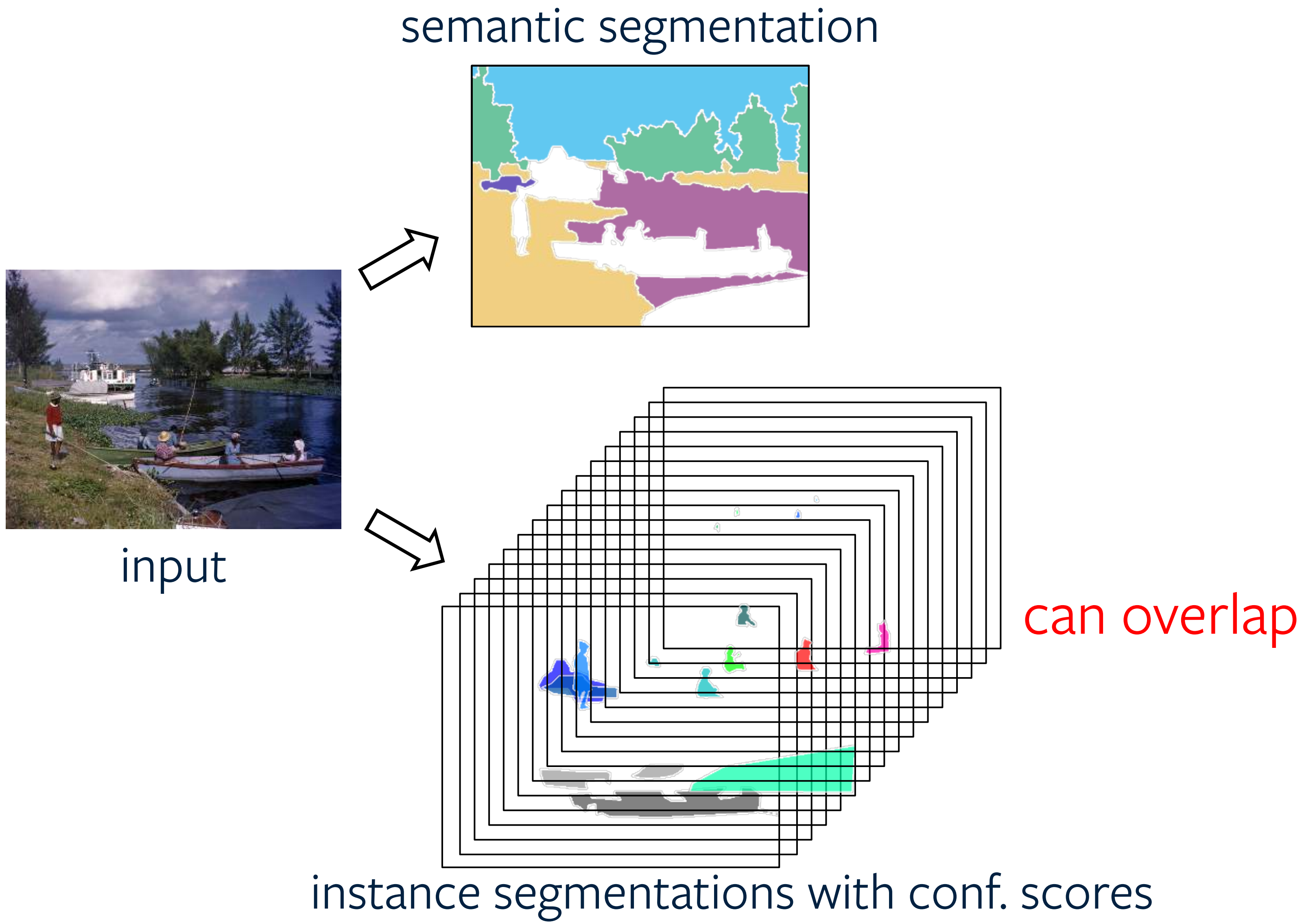


input

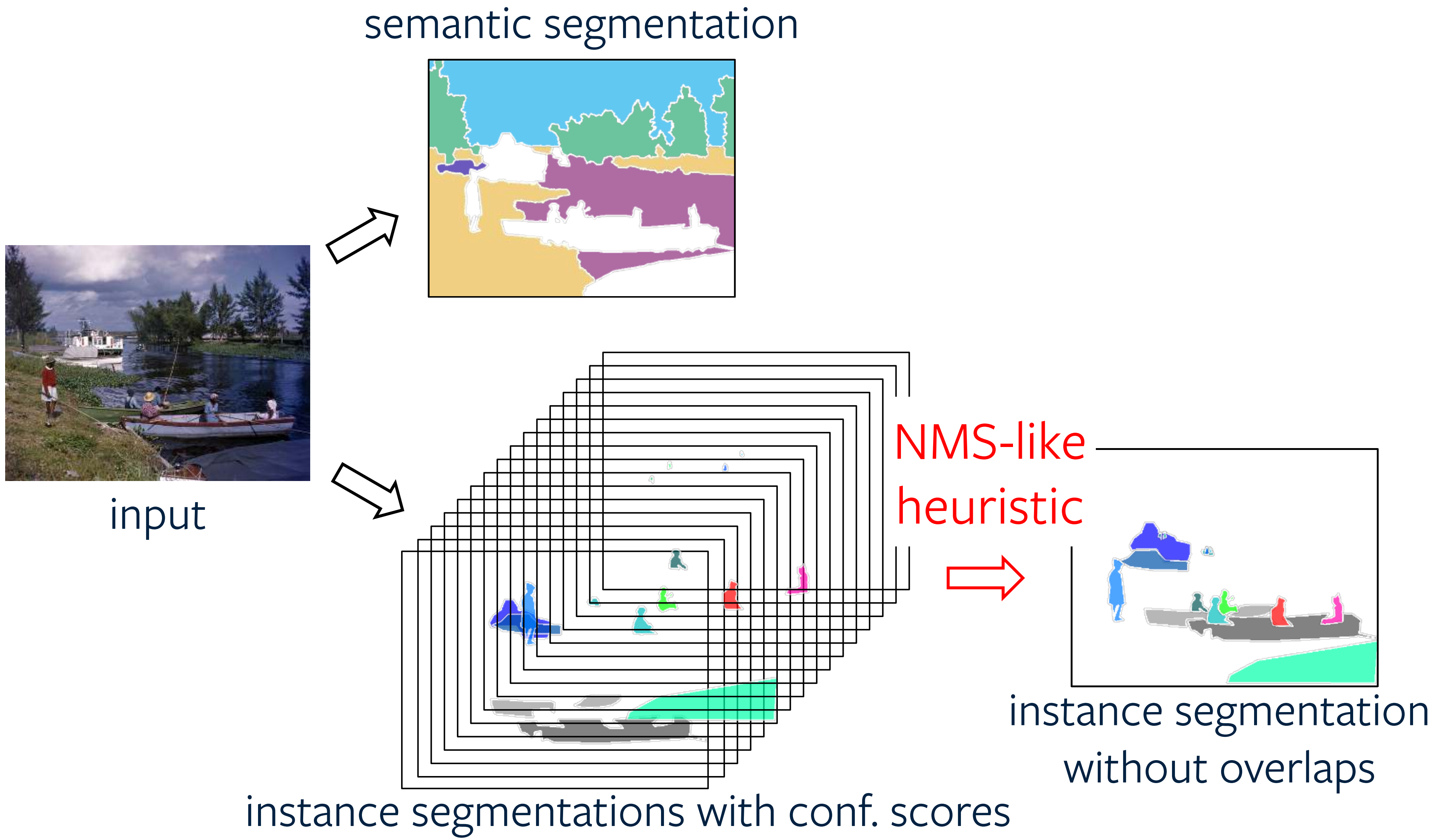


person, conf. score 0.93

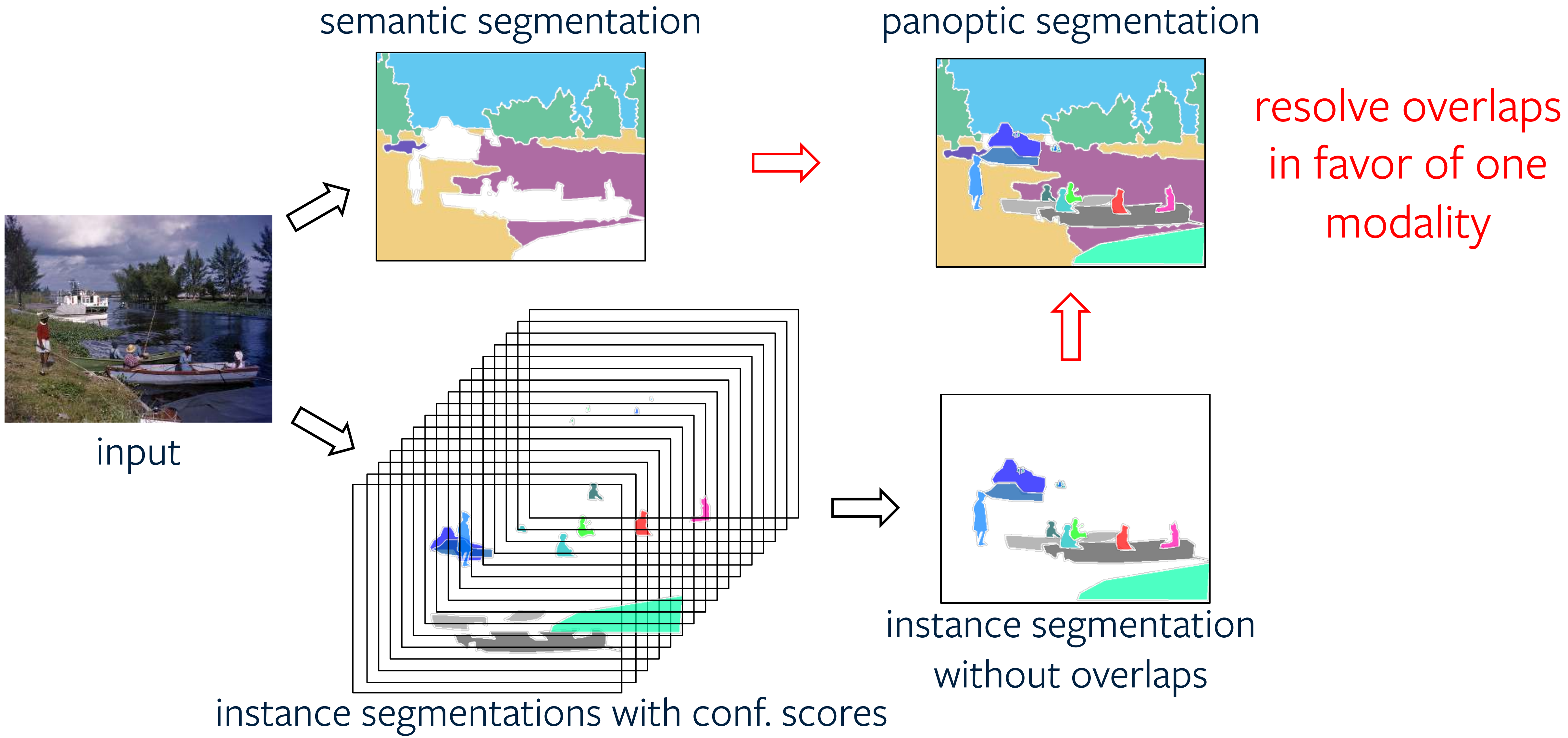
Panoptic segmentation: naïve approach



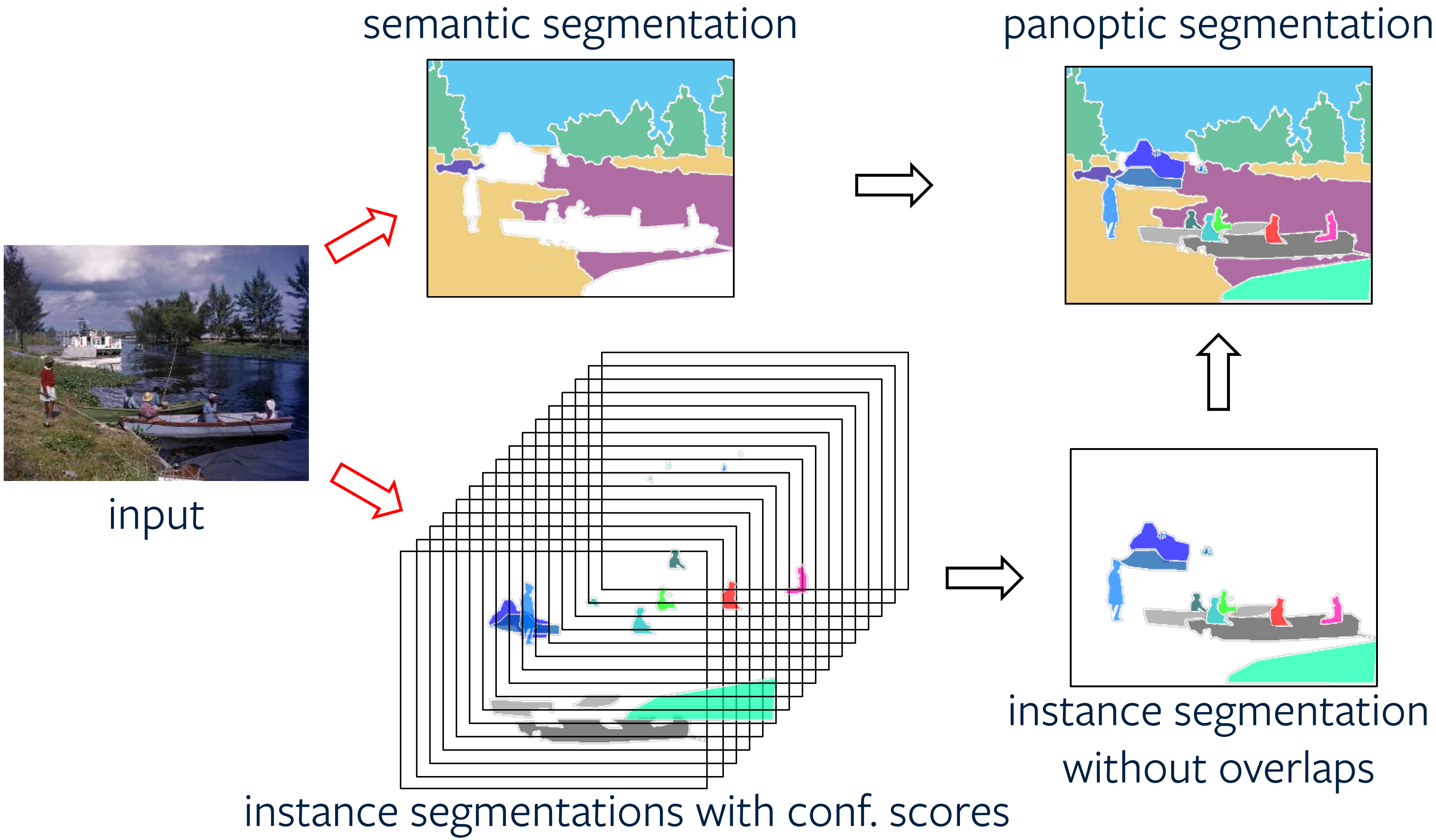
Panoptic segmentation: naïve approach



Panoptic segmentation: naïve approach

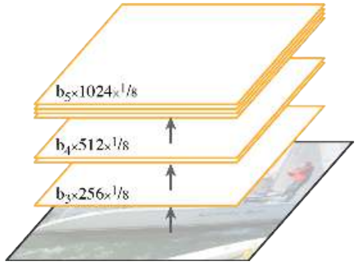


Panoptic segmentation: naïve approach



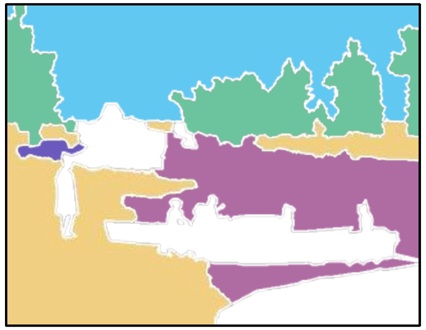
Panoptic segmentation: naïve approach

input



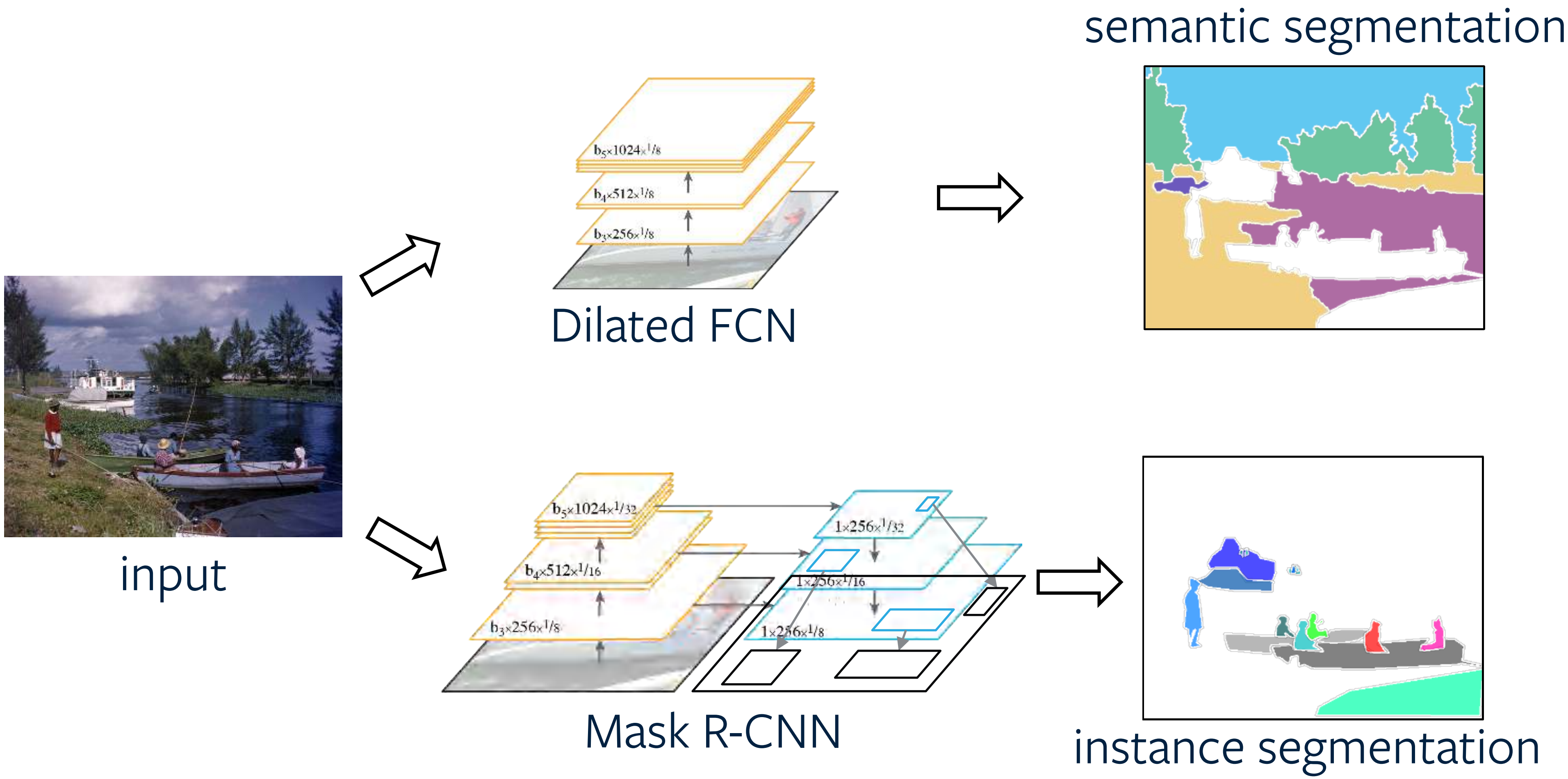
Dilated FCN

semantic segmentation



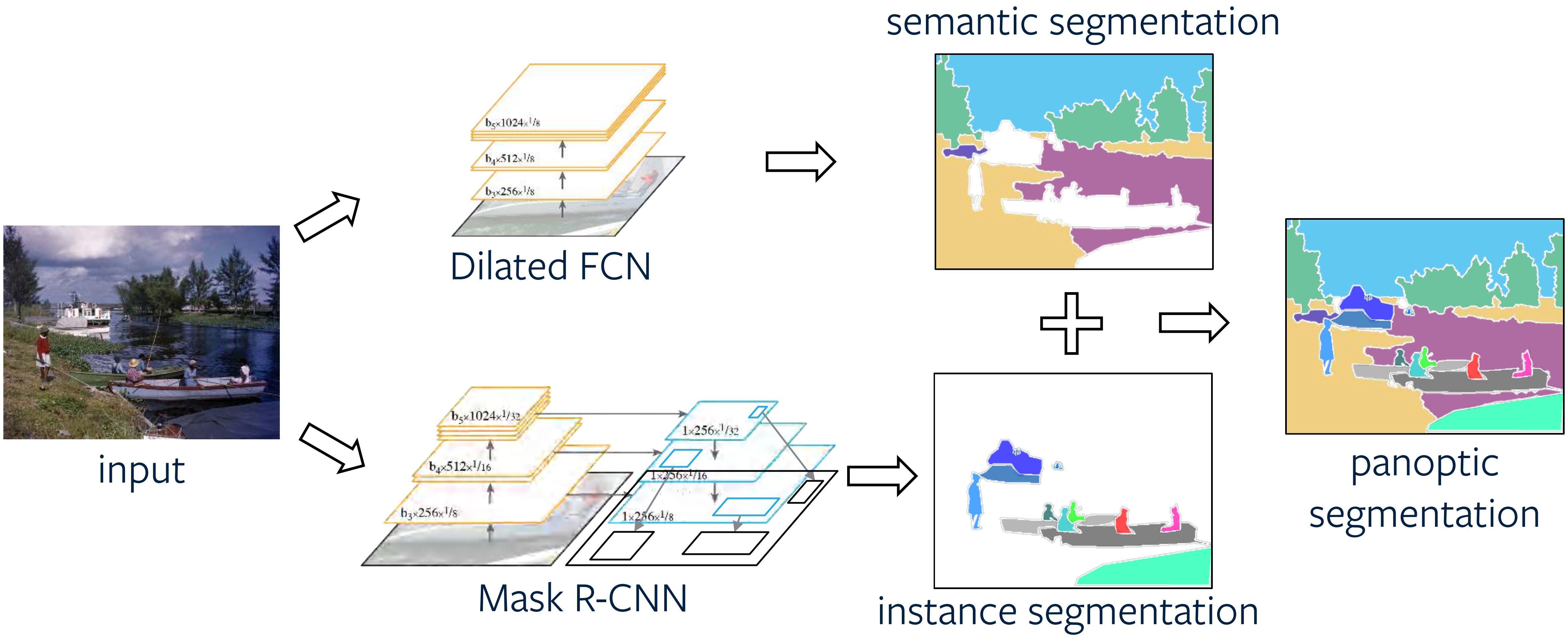
best known semantic segmentation method

Panoptic segmentation: naïve approach



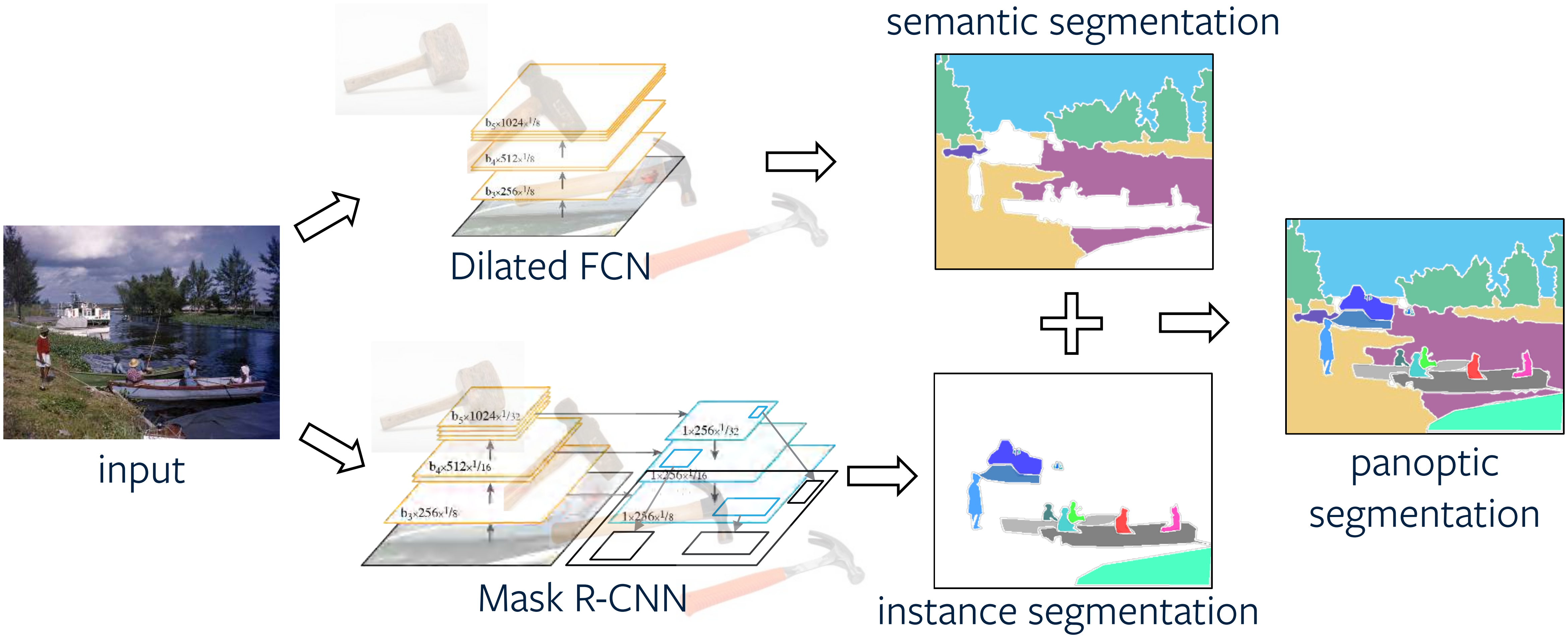
best known instance segmentation method

Panoptic segmentation: naïve approach



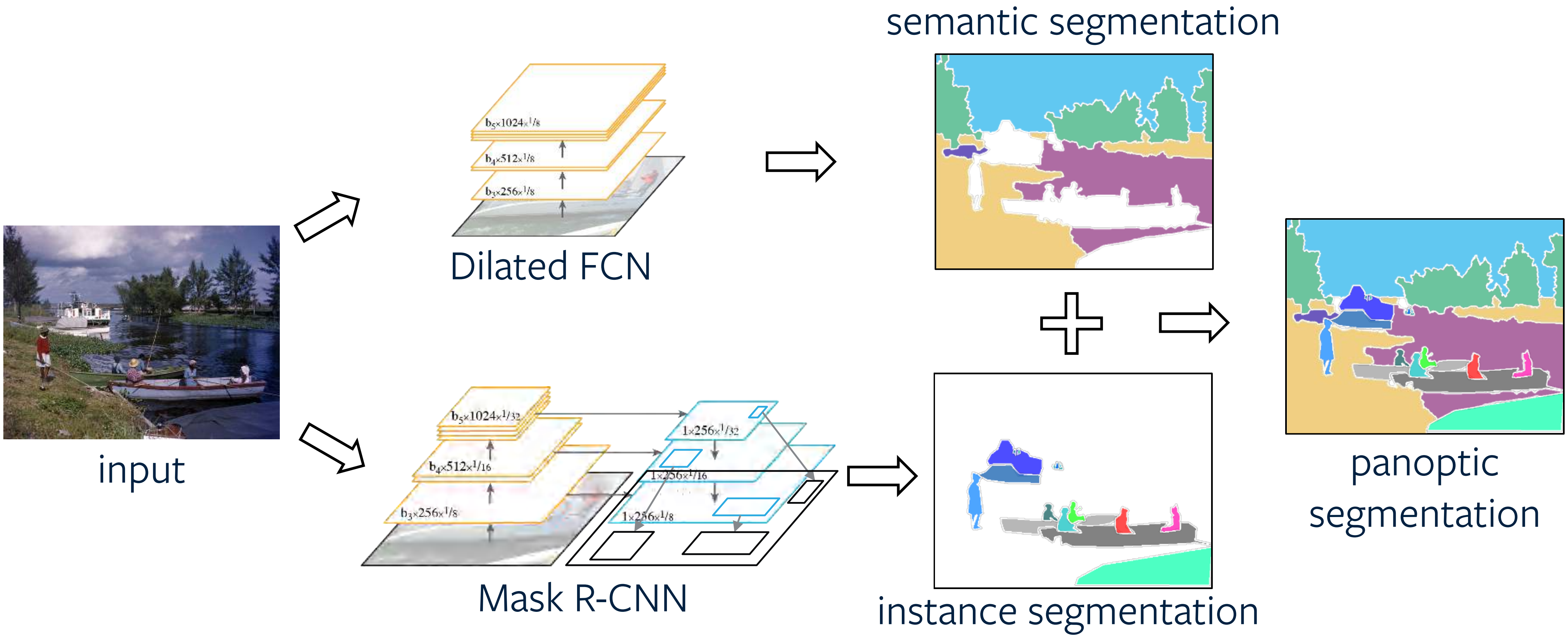
resolve overlaps between different
instances and stuff classes

Panoptic segmentation: naïve approach



better semantic or instance segmentation ->
better panoptic segmentation

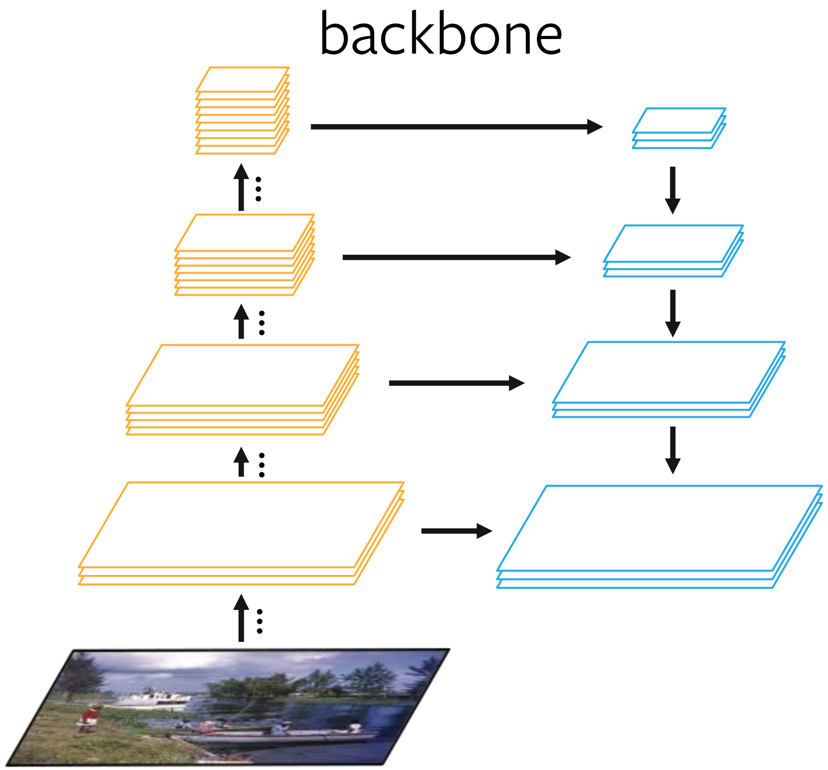
Panoptic segmentation: naïve approach



inefficient

hard to improve

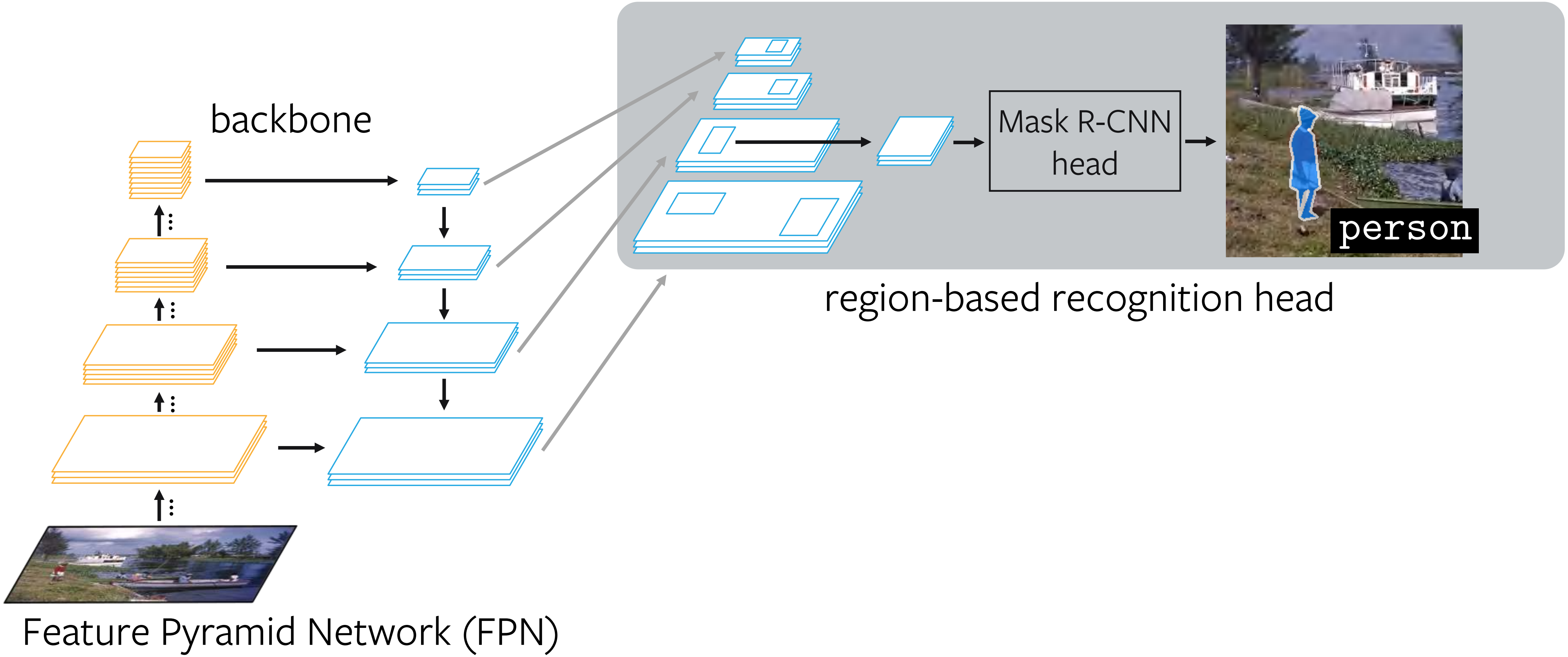
Panoptic FPN: unified framework



Feature Pyramid Network (FPN)

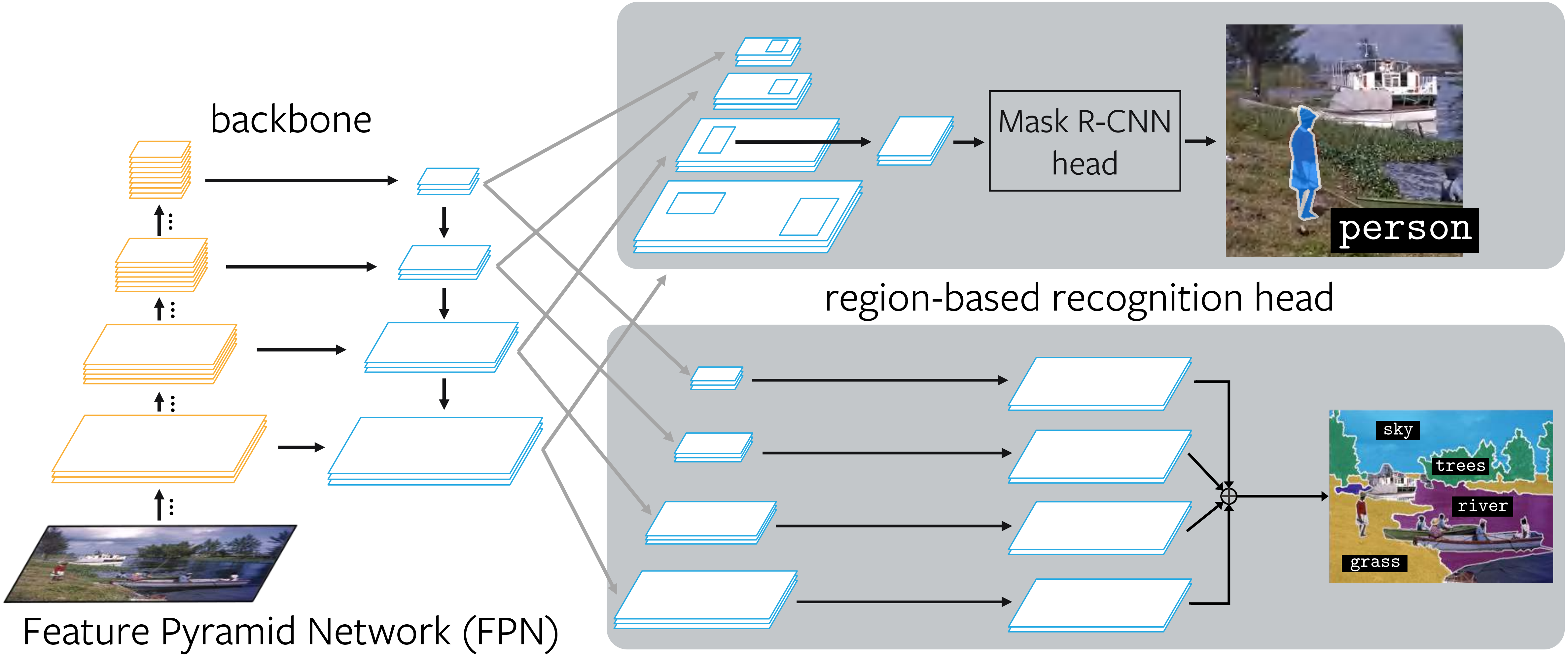
Lin et al. Feature Pyramid Networks for Object Detection, CVPR`17

Panoptic FPN: unified framework



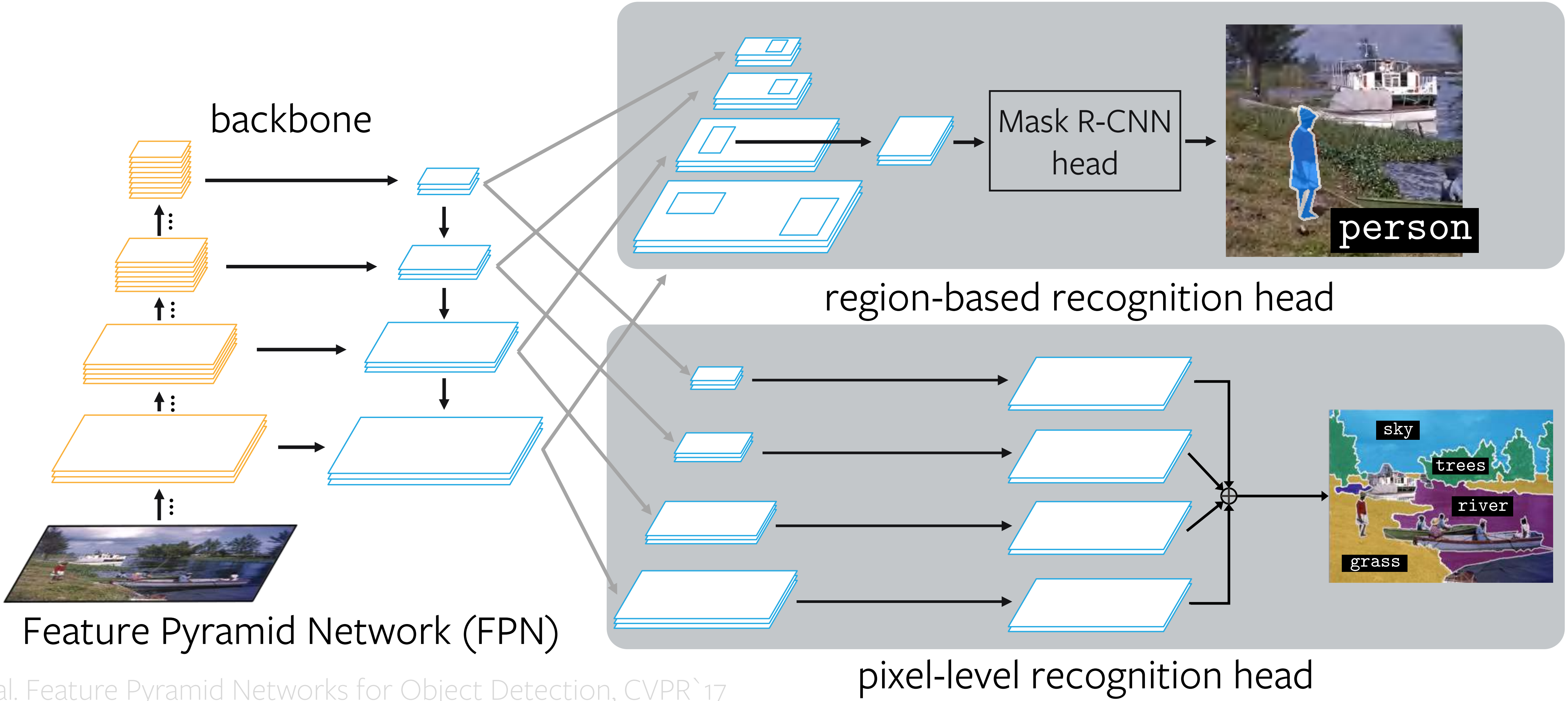
Lin et al. Feature Pyramid Networks for Object Detection, CVPR`17
He et al. Mask R-CNN, ICCV`17

Panoptic FPN: unified framework



Lin et al. Feature Pyramid Networks for Object Detection, CVPR`17
He et al. Mask R-CNN, ICCV`17
Kirillov et al. Panoptic Feature Pyramid Networks, CVPR`19

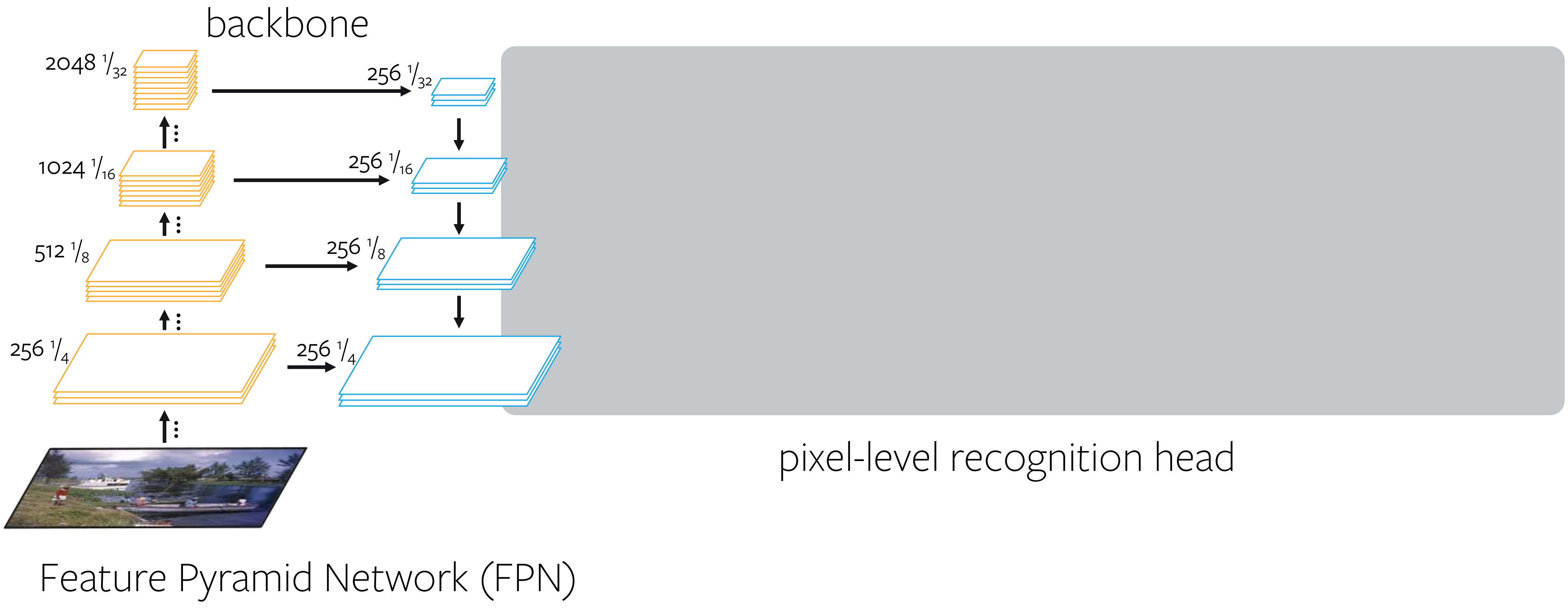
Panoptic FPN: unified framework



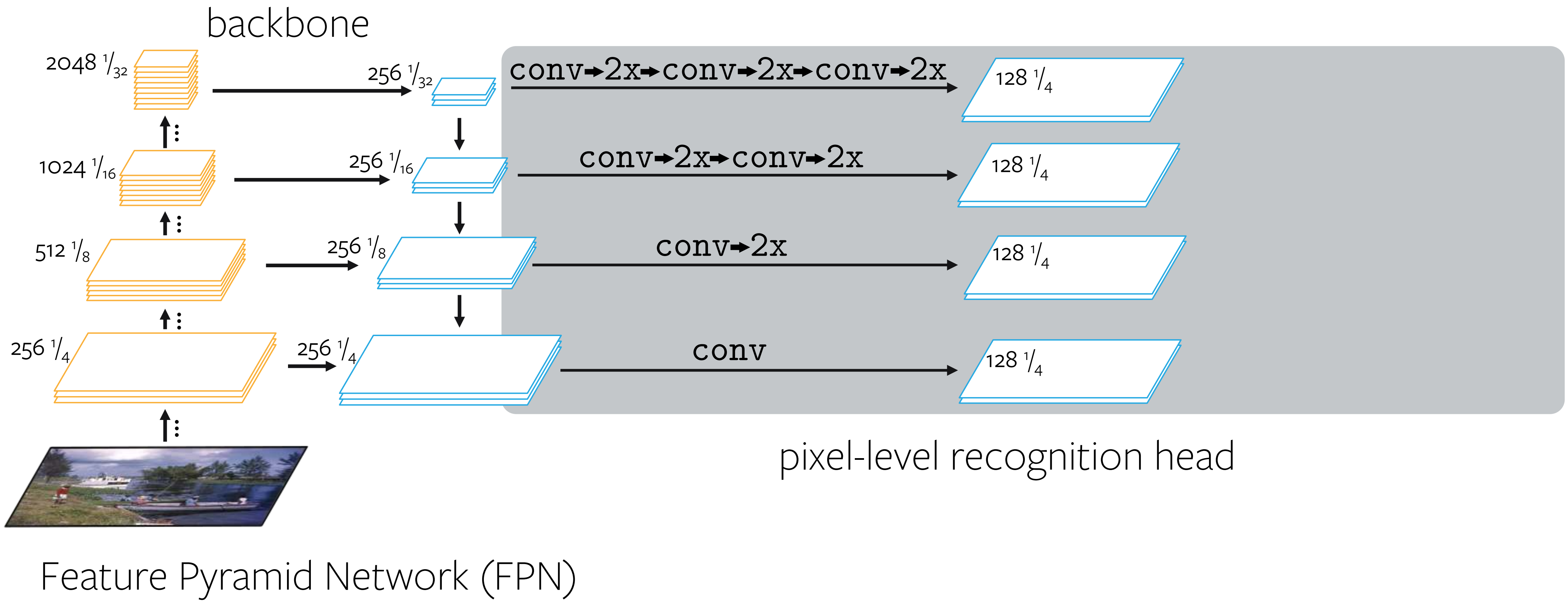
Lin et al. Feature Pyramid Networks for Object Detection, CVPR`17
He et al. Mask R-CNN, ICCV`17
Kirillov et al. Panoptic Feature Pyramid Networks, CVPR`19

not yet another R-CNN framework head

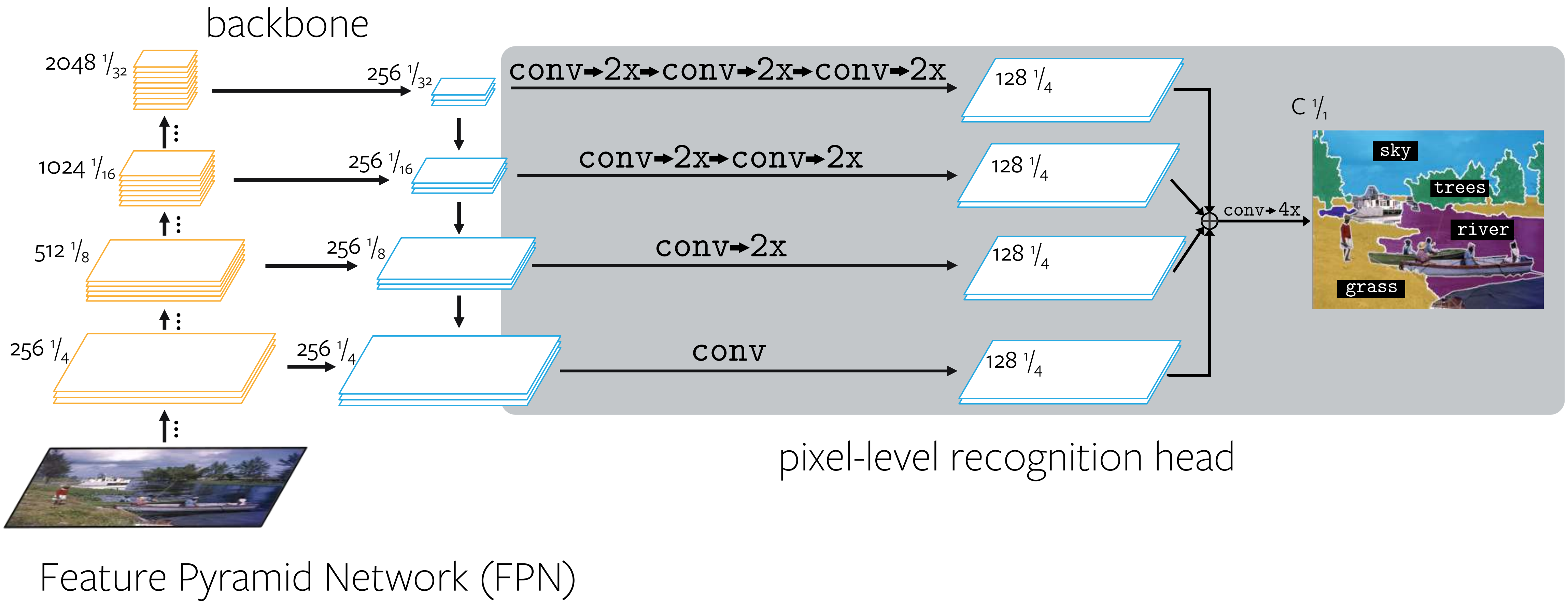
Semantic FPN



Semantic FPN



Semantic FPN



Semantic FPN

	backbone	mIoU	FLOPs	memory
DeeplabV3	ResNet-101-D8	77.8	1.9	1.9
PSANet101	ResNet-101-D8	77.9	2.0	2.0
Mapillary	WideResNet-38-D8	79.4	4.3	1.7
DeeplabV3+	X-71-D16	79.6	0.5	1.9
Semantic FPN	ResNet-101-FPN	77.7	0.5	0.8
Semantic FPN	ResNeXt-101-FPN	79.1	0.8	1.4

Cityscapes

on par performance with the best
semantic segmentation approaches

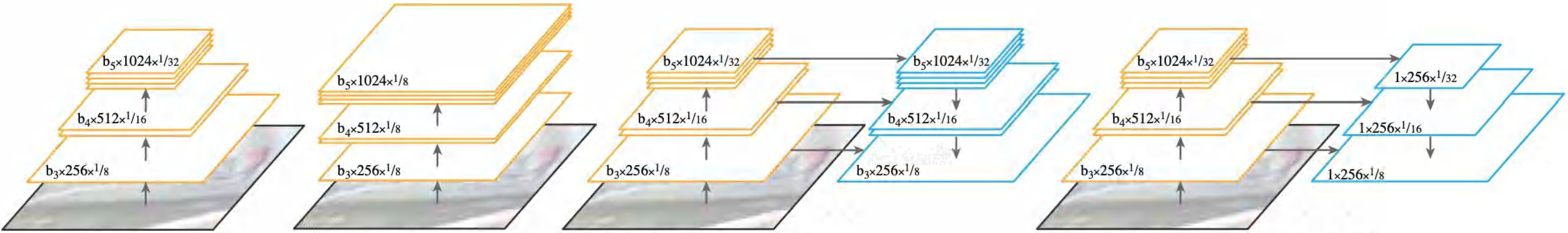
Semantic FPN

	backbone	mIoU	FLOPs	memory
DeeplabV3	ResNet-101-D8	77.8	1.9	1.9
PSANet101	ResNet-101-D8	77.9	2.0	2.0
Mapillary	WideResNet-38-D8	79.4	4.3	1.7
DeeplabV3+	X-71-D16	79.6	0.5	1.9
Semantic FPN	ResNet-101-FPN	77.7	0.5	0.8
Semantic FPN	ResNeXt-101-FPN	79.1	0.8	1.4

Cityscapes

computational and memory efficient

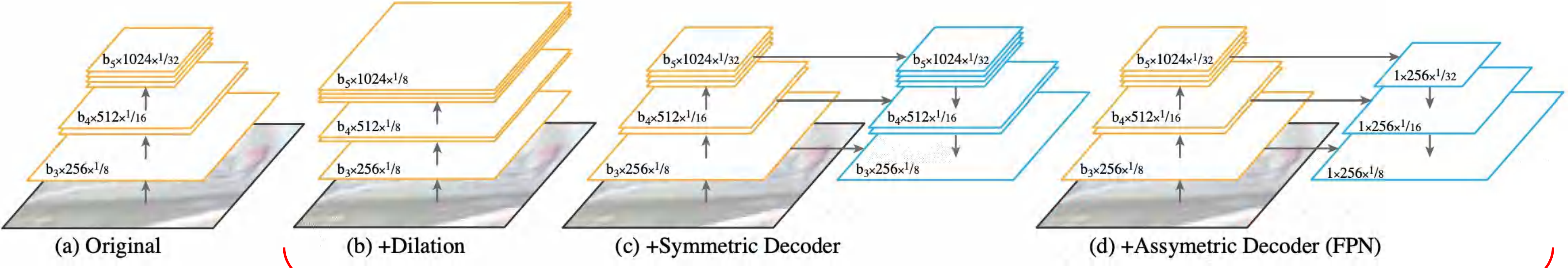
Semantic FPN



(a) Original (b) +Dilation (c) +Symmetric Decoder (d) +Assymetric Decoder (FPN)

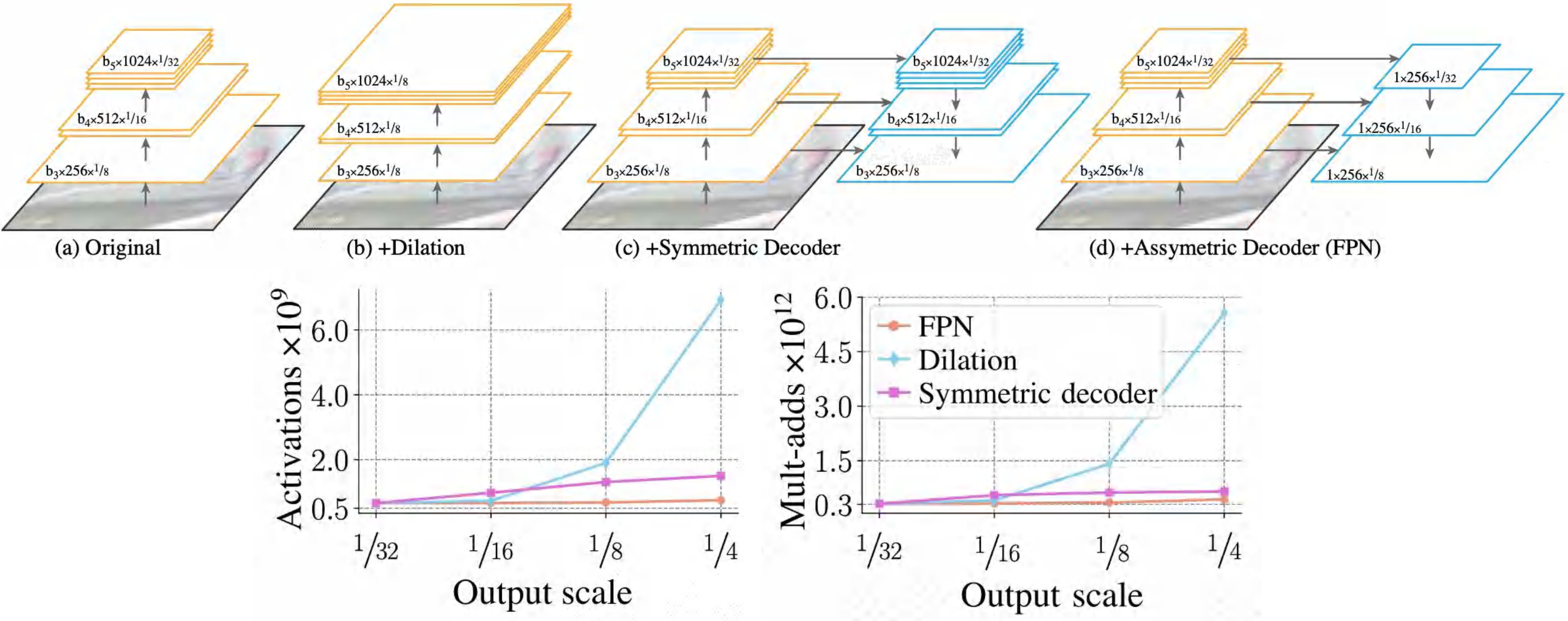
classification network

Semantic FPN



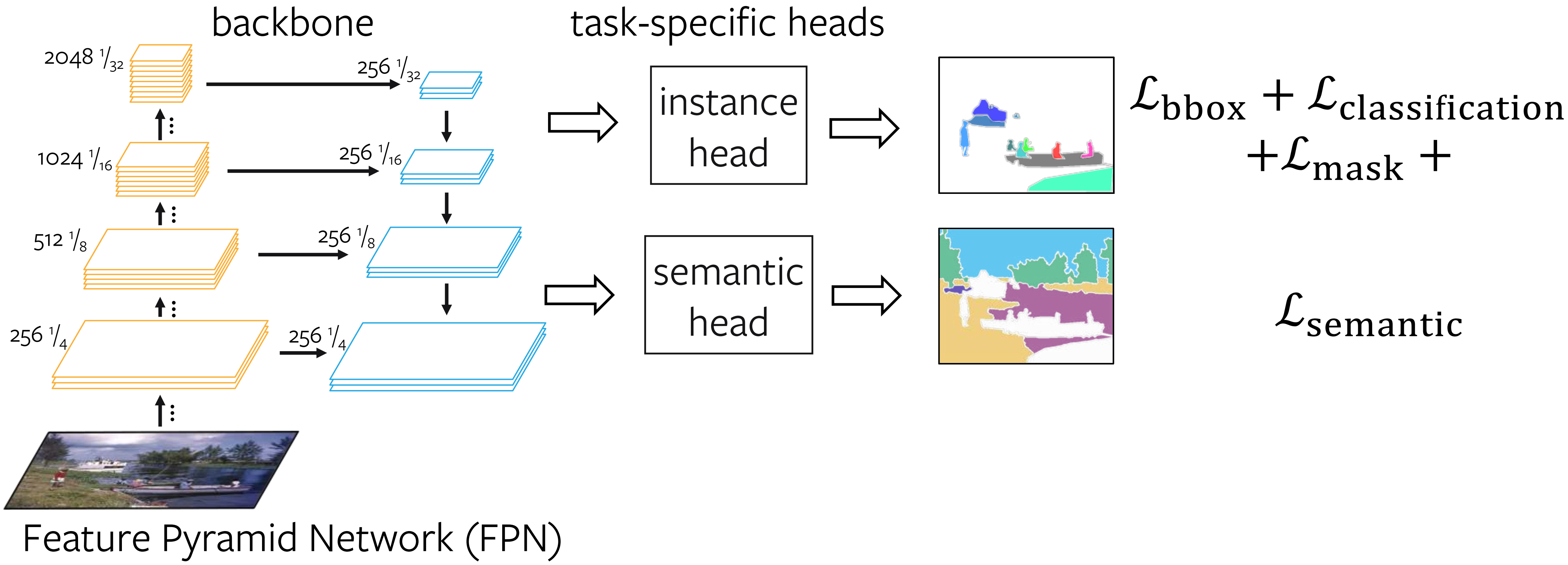
different approaches to preserve spatial resolution

Semantic FPN



FPN-based backbone is the most efficient

Panoptic FPN



Panoptic FPN vs. Mask R-CNN

$$\mathcal{L} = \lambda_i(\mathcal{L}_{\mathbf{bbox}} + \mathcal{L}_{\mathbf{classification}} + \mathcal{L}_{\mathbf{mask}}) + \lambda_s \mathcal{L}_{\mathbf{semantic}}$$

dataset	λ_i	λ_s	AP
COCO	1.0	**0.1**	+0.1
Cityscapes	1.0	**1.0**	+1.0

improves instance segmentation
compare with Mask R-CNN alone

Panoptic FPN vs. Semantic FPN

$$\mathcal{L} = \lambda_i(\mathcal{L}_{\text{bbox}} + \mathcal{L}_{\text{classification}} + \mathcal{L}_{\text{mask}}) + \lambda_s \mathcal{L}_{\text{semantic}}$$

dataset	$\boldsymbol{\lambda_i}$	λ_s	IoU
COCO	**1.0**	1.0	+1.2
Cityscapes	**0.25**	1.0	+1.0

improves semantic segmentation
compare to Semantic FPN alone

Panoptic FPN vs. Mask R-CNN + Semantic FPN

dataset	inst. segm.	sem. segm.	panoptic segm.
COCO	+1.3 AP	+1.9 IoU	+0.9 PQ
Cityscapes	+0.8 AP	+1.2 IoU	+0.3 PQ

given the same computational budget

Panoptic FPN takeaway

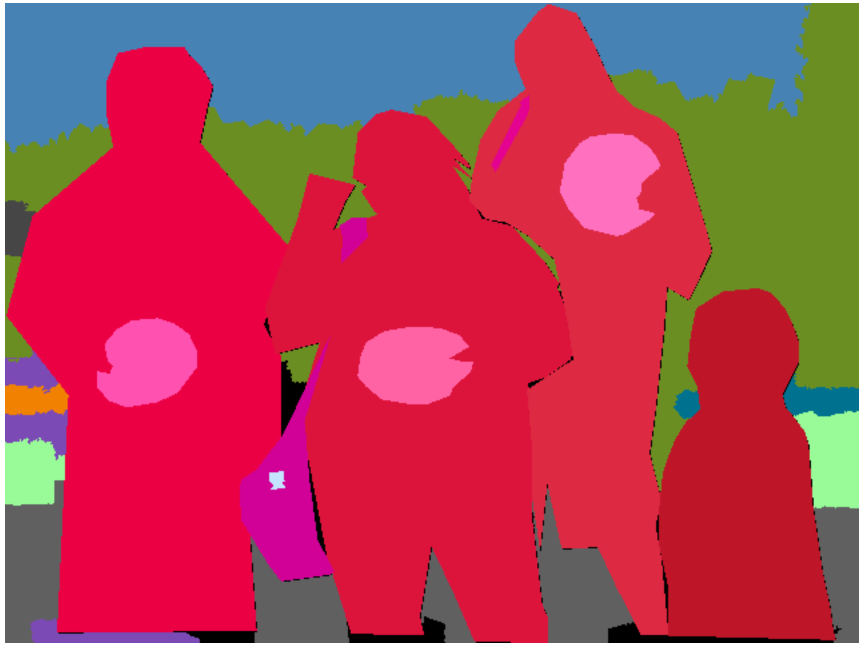
- straightforward and efficient baseline for panoptic segmentation
- OSS version later this year as a part of `Detectron2 (PyTorch)`
- lower bound for future panoptic methods

Panoptic FPN takeaway

- straightforward and efficient baseline for panoptic segmentation
- OSS version later this year as a part of **`Detectron2 (PyTorch)`**
- lower bound for future panoptic methods

Is panoptic segmentation solved?

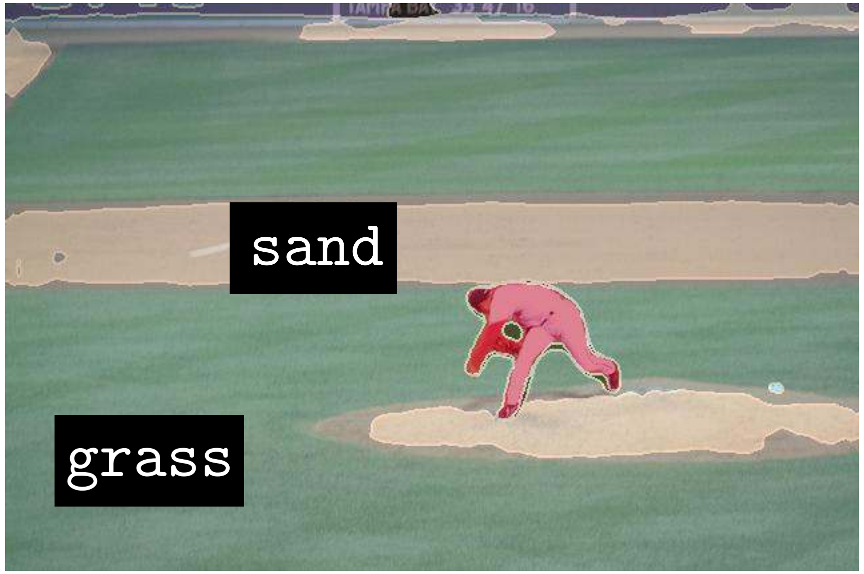
prediction



ground truth

suboptimal overlaps resolution

Is panoptic segmentation solved?



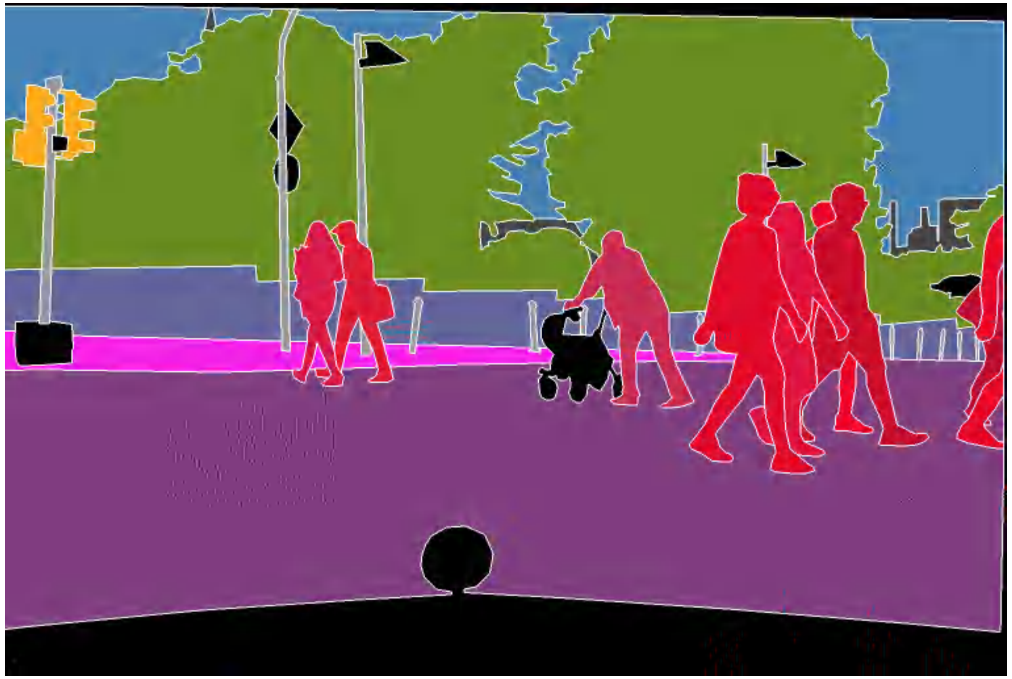
prediction

ground truth

missing context

Is panoptic segmentation solved?

prediction

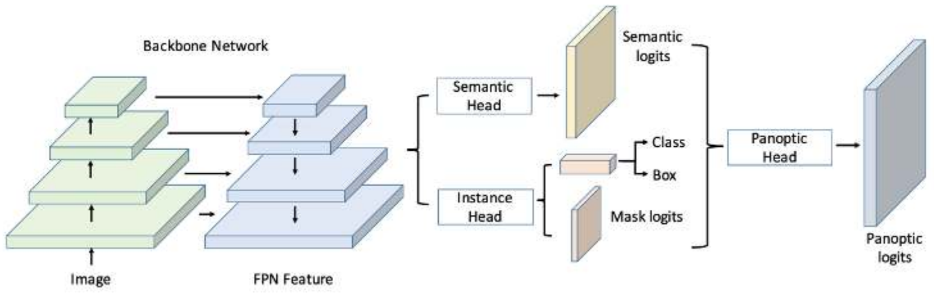


ground truth

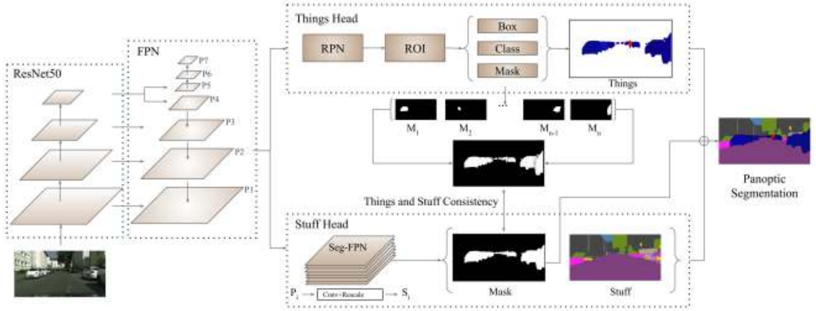
poor alignment

Recent development

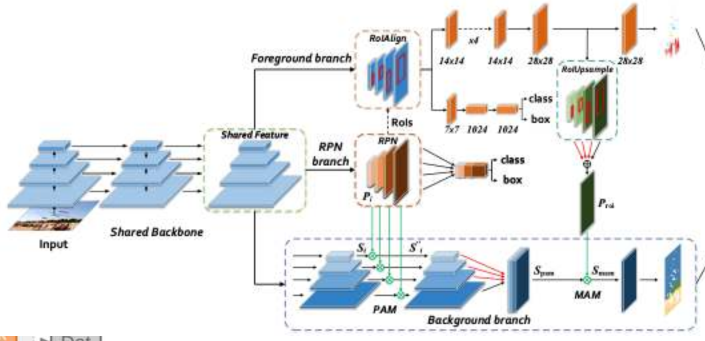
1. Li et al. Attention-guided Unified Network for Panoptic Segmentation, CVPR 2019
2. Xiong et al. UPSNet: A Unified Panoptic Segmentation Network, CVPR 2019
3. Liu et al. An End-to-End Network for Panoptic Segmentation, CVPR 2019
4. Li et al. Learning to Fuse Things and Stuff, arXiv 2018
5. Porzi et al. Seamless Scene Segmentation, arXiv 2019

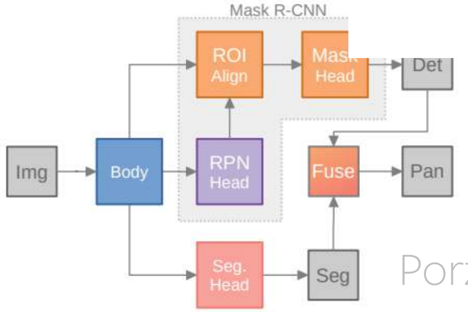


Xiong et al.

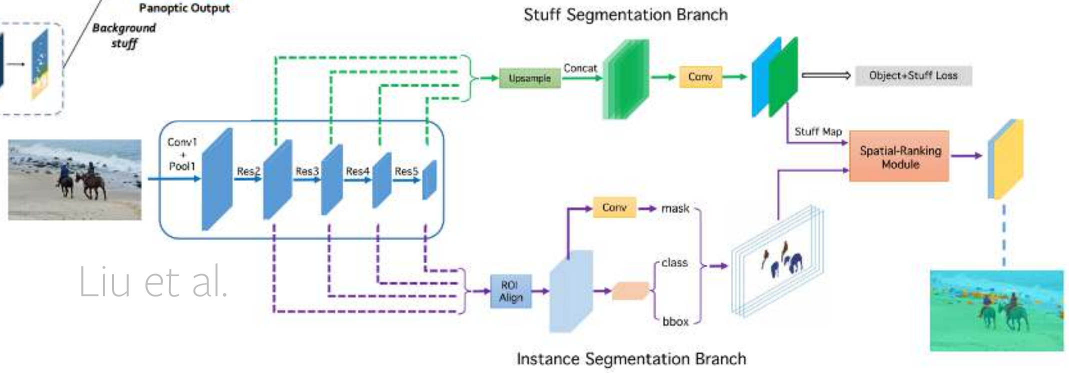


Li et al.





Porzi et al.



Liu et al.

Schemes credits:
papers on the slide

Takeaway

- **Panoptic segmentation** – practically important task with a lot of room for improvement
- **Panoptic FPN** – simple baseline for the task that your method should beat
- **Panoptic segmentation challenges** – COCO & Vistas (ICCV`19), Cityscapes leaderboard

ICCV 2019
Seoul, Korea

FAIR Research Engineer

Menlo Park, CA
Seattle, WA

ACCELERATE AND SCALE CV RESEARCH

Familiarity with CV and ML
Ability to write high-quality and performance-critical code

wlo@fb.com