Part 1: COCO 2018 Update
Part 2: Panoptic Segmentation

Piotr Dollár

facebook
Artificial Intelligence

Part 1: COCO 2018 Update

Part 2: Panoptic Segmentation

Piotr Dollár

COCO + Mapillary

Joint Recognition Challenge Workshop at ECCV 2018

http://cocodataset.org/workshop/coco-mapillary-eccv-2018.html
(or just browse to http://cocodataset.org/ and go from there)

COCO + Mapillary

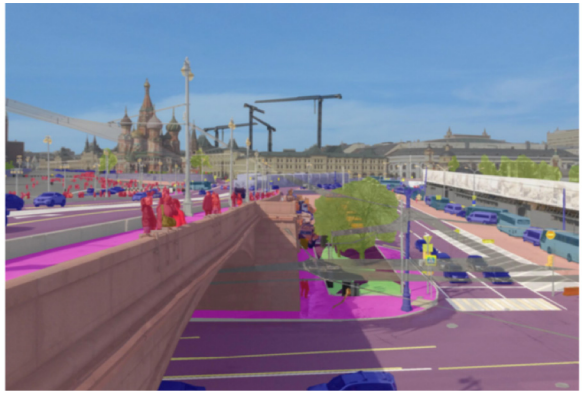
COCO: Recognition in natural scenes

Mapillary: Recognition in street-view scenes

COCO + Mapillary

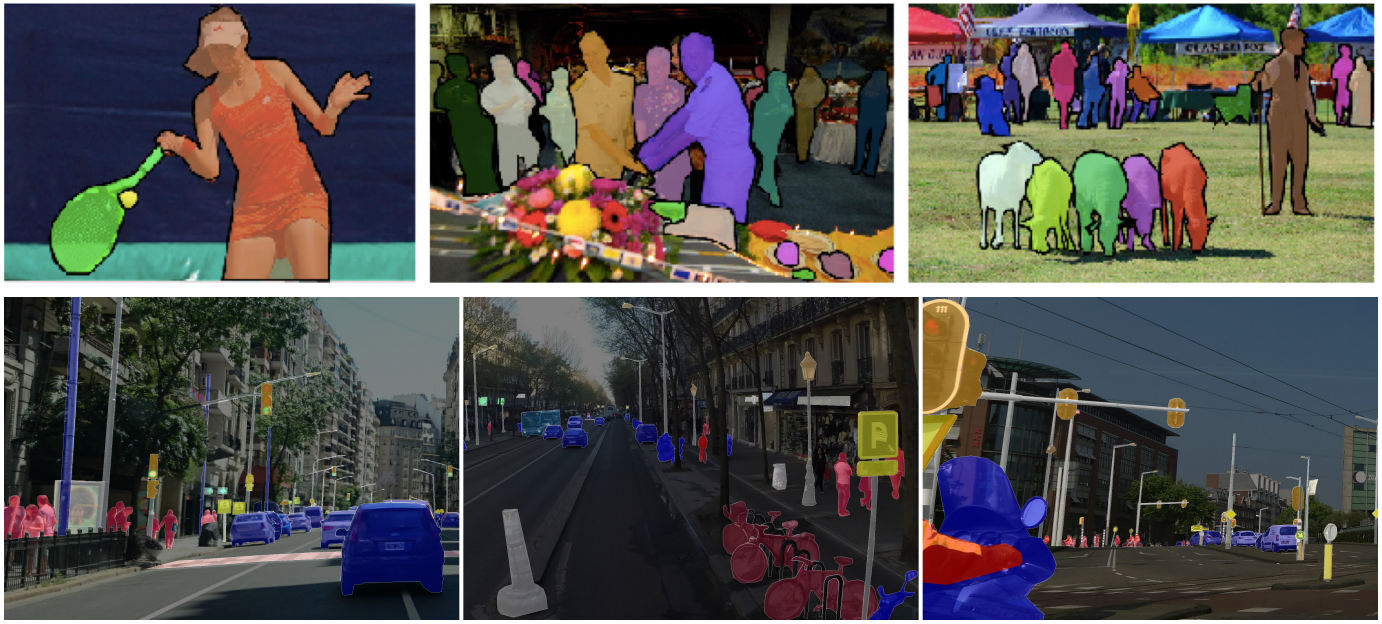
COCO: Recognition in natural scenes


Mapillary: Recognition in street-view scenes



→ **Unified tasks, metrics, and code**

Task #1: Object Detection



- Both COCO + Mapillary
- Same data/metrics as in 2017
- New in 2018:
  - **instance segmentation only!**
  - bounding box task **not** a challenge
  - bounding box test-dev remains open

Task #2: Panoptic Segmentation

- Both COCO + Mapillary
- **New challenge for 2018!**
  - Joint stuff + things recognition
  - Details to be described later in talk
  - Aim to push recognition in new direction

Task #3: Keypoint Detection

- COCO only
- Same data/metrics as in 2017

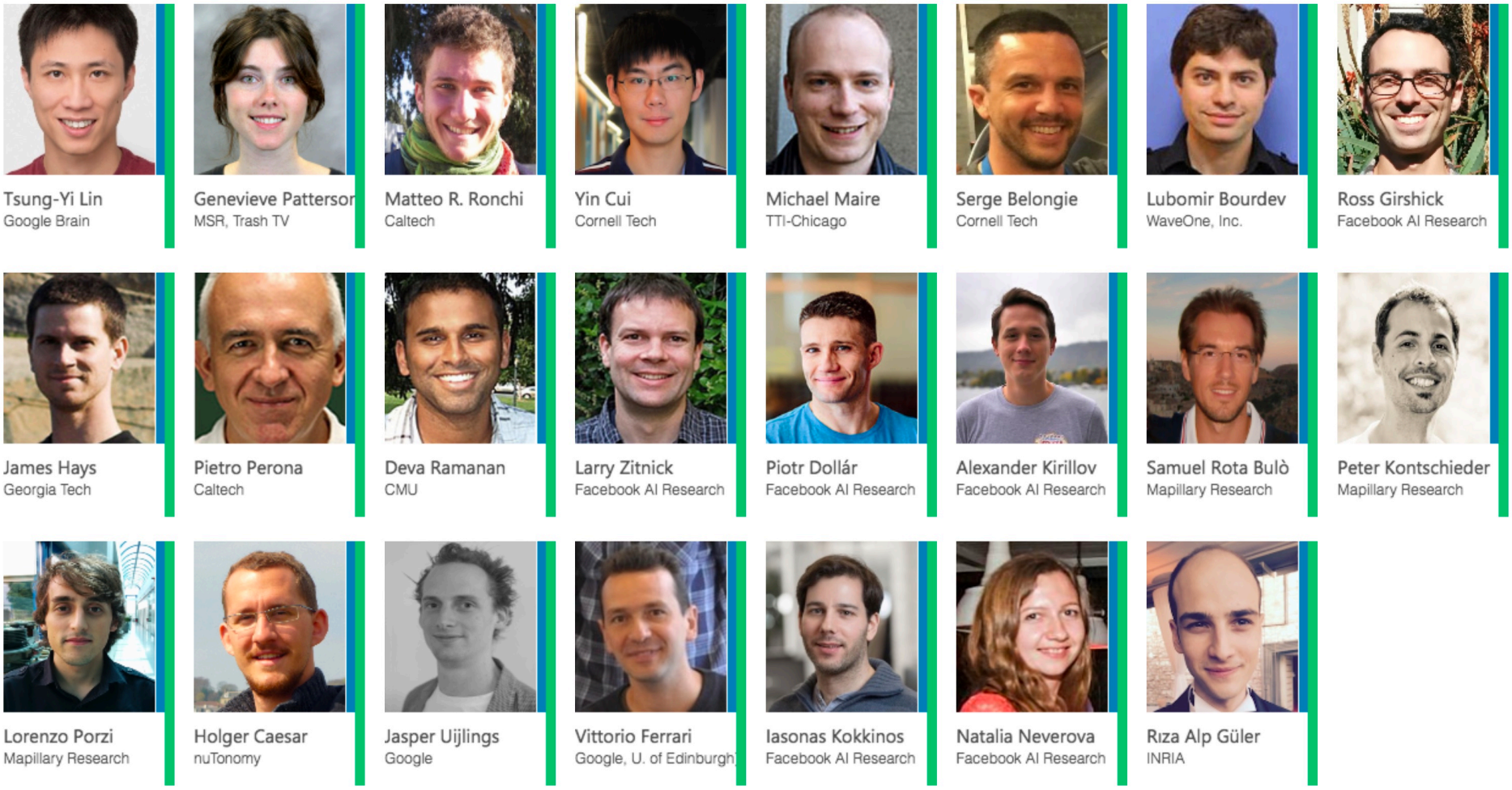
Task #4: DensePose

- COCO only
- **New challenge for 2018!**
  - Dense human pose estimation
  - Aim to push keypoints in new direction
  - http://densepose.org/

Challenge Tracks

- COCO Challenge Tasks:
  - Object Detection (Instance Segmentation)
  - Panoptic Segmentation
  - Keypoint Detection
  - DensePose
- Mapillary Challenge Tasks:
  - Object Detection (Instance Segmentation)
  - Panoptic Segmentation
- Tasks **not** in 2018 challenge:
  - Bounding-Box Object Detection
  - Stuff Segmentation

COCO + Mapillary



Part 1: COCO 2018 Update

Part 2: Panoptic Segmentation



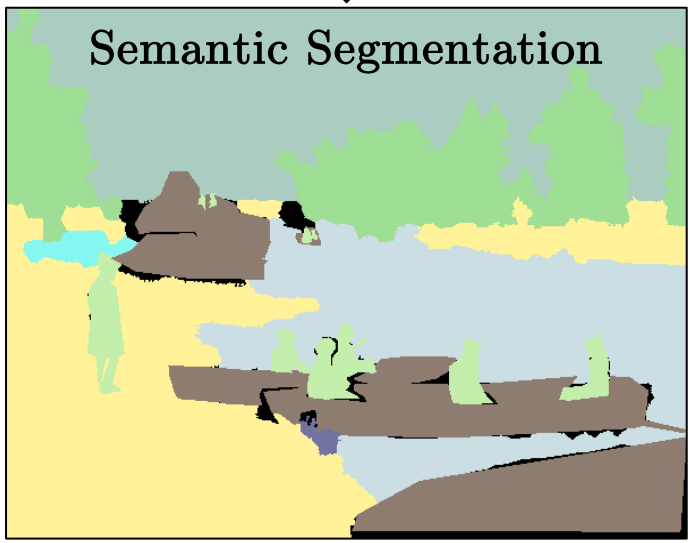


Alex Kirillov Kaiming He Ross Girshick Carsten Rother **Piotr Dollár**

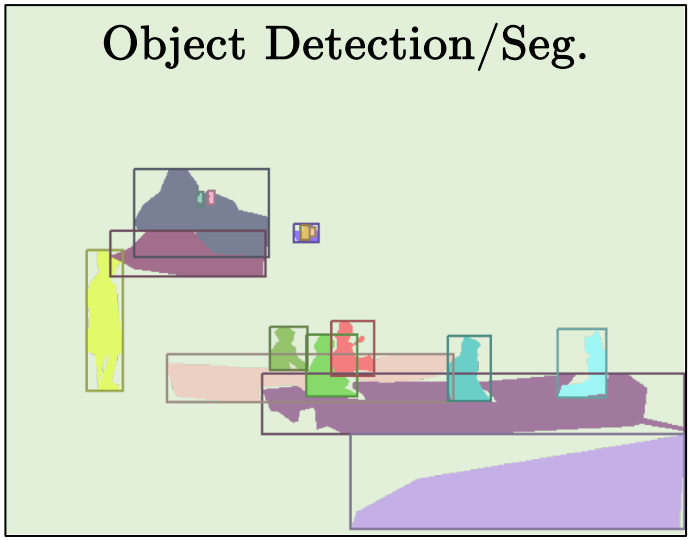
Outline

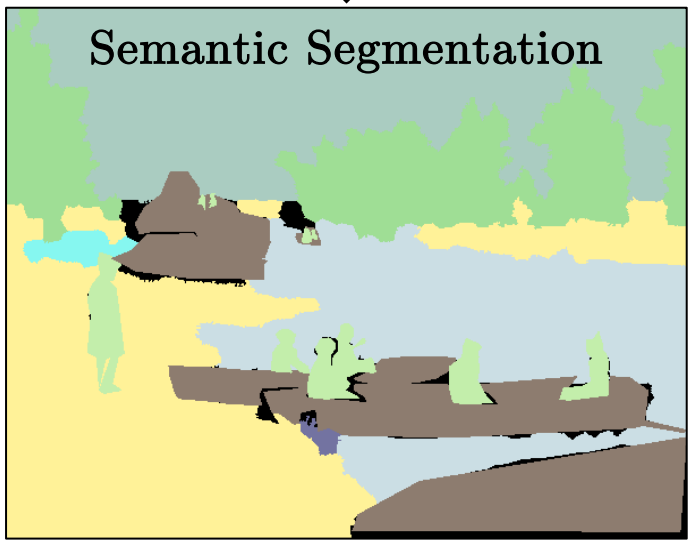
- **Motivation & Definition**
- Quality Evaluation
- Human Performance
- Machine Performance
- Perspectives

Unifying Semantic and Instance Segmentation

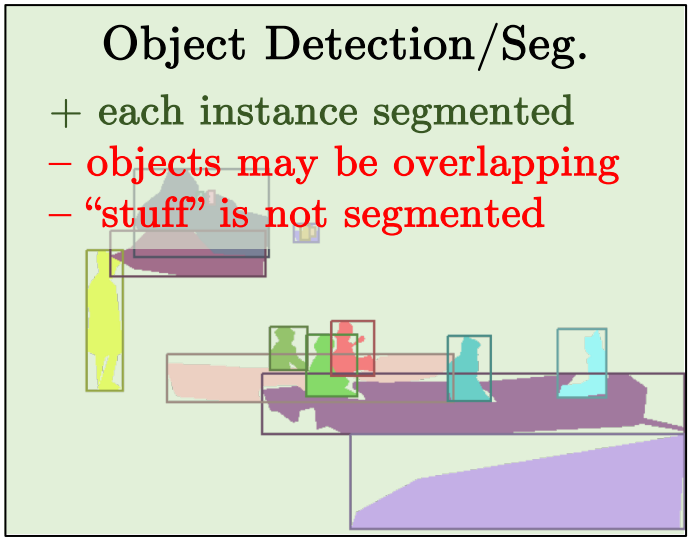


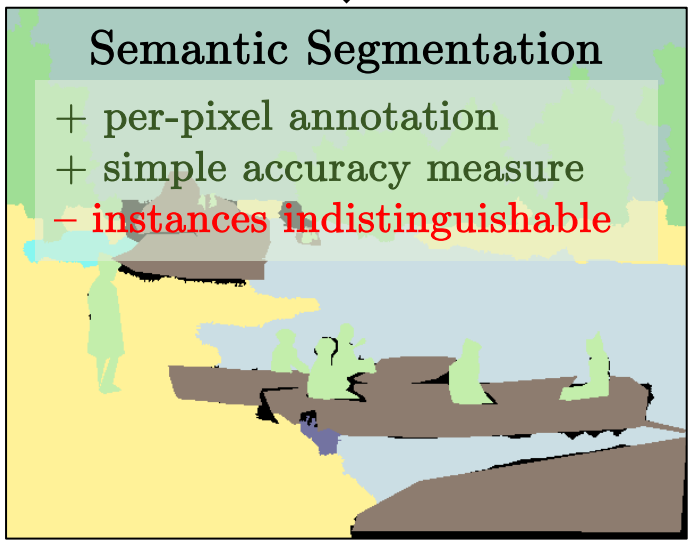
Unifying Semantic and Instance Segmentation



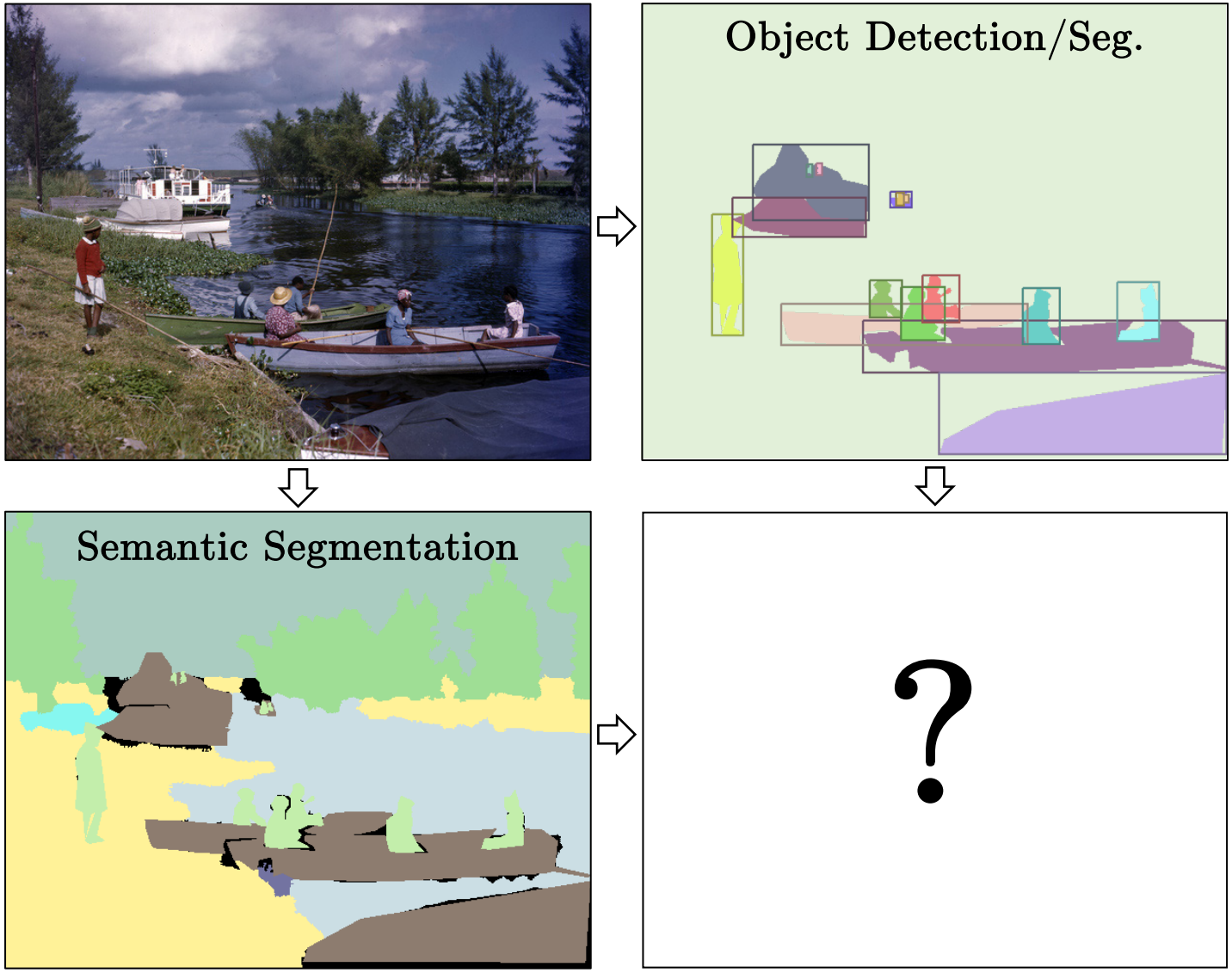


Unifying Semantic and Instance Segmentation

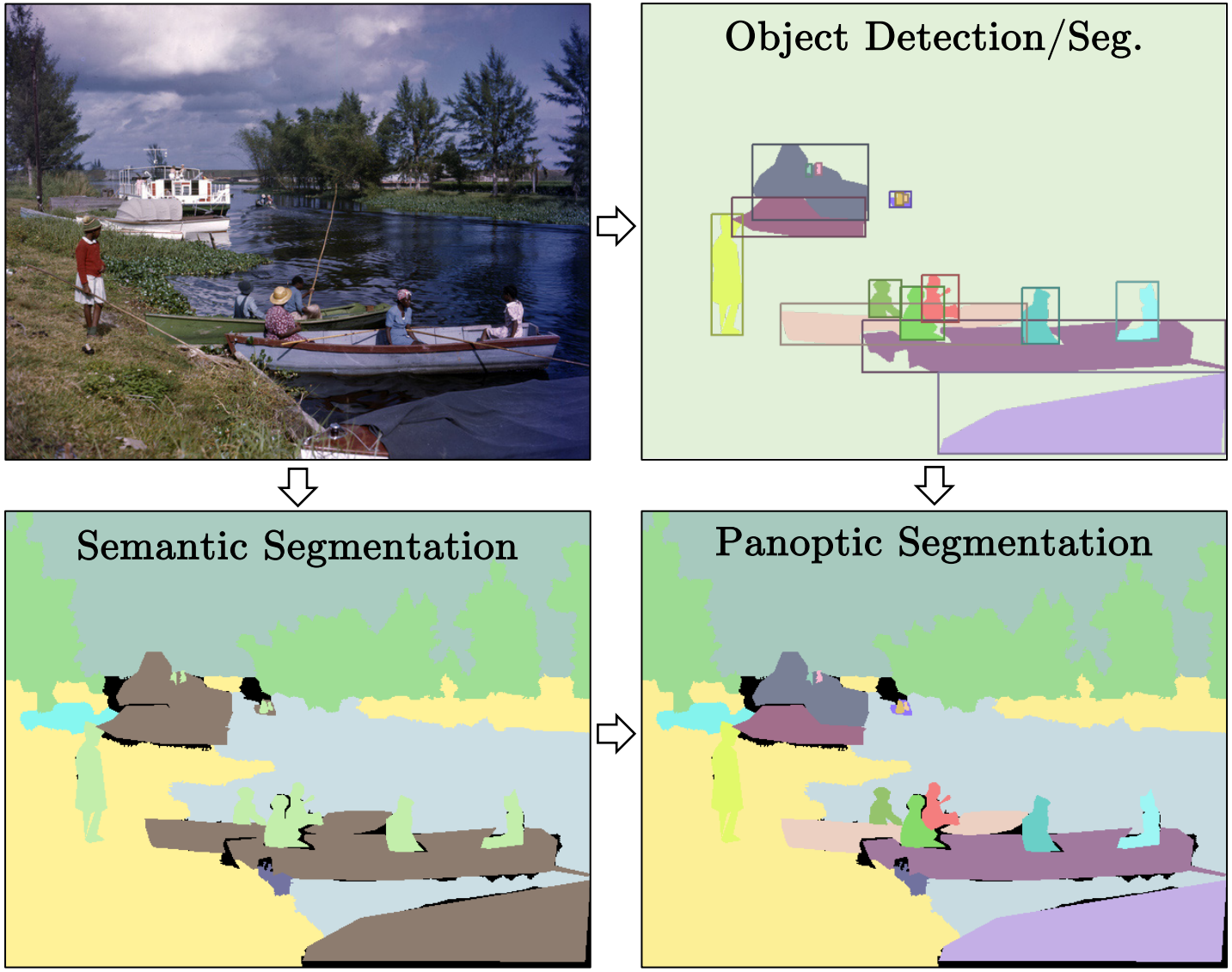




Unifying Semantic and Instance Segmentation



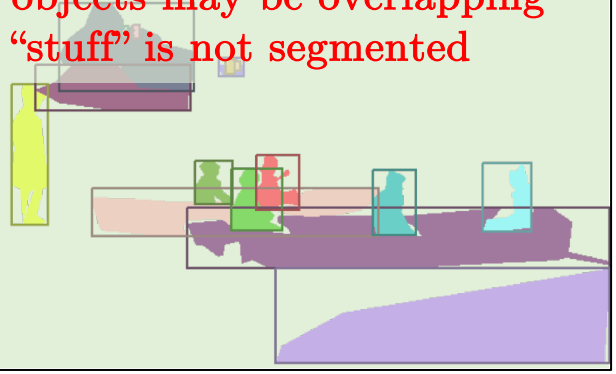
Unifying Semantic and Instance Segmentation



Unifying Semantic and Instance Segmentation

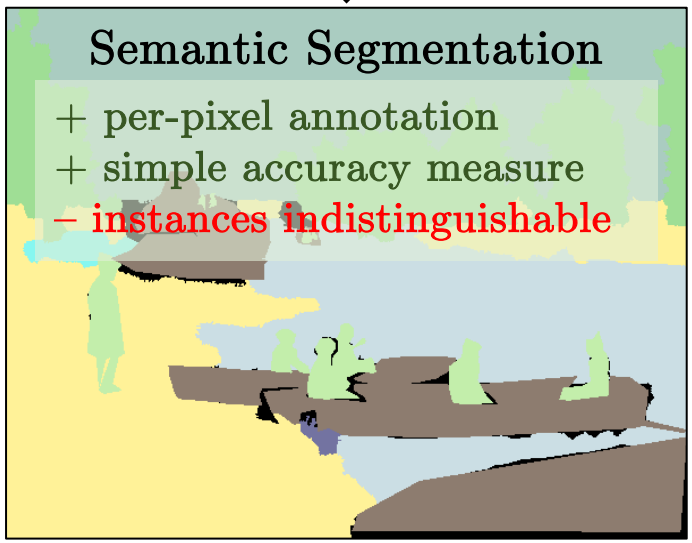
Object Detection/Seg.

+ each instance segmented
– objects may be overlapping
– "stuff" is not segmented



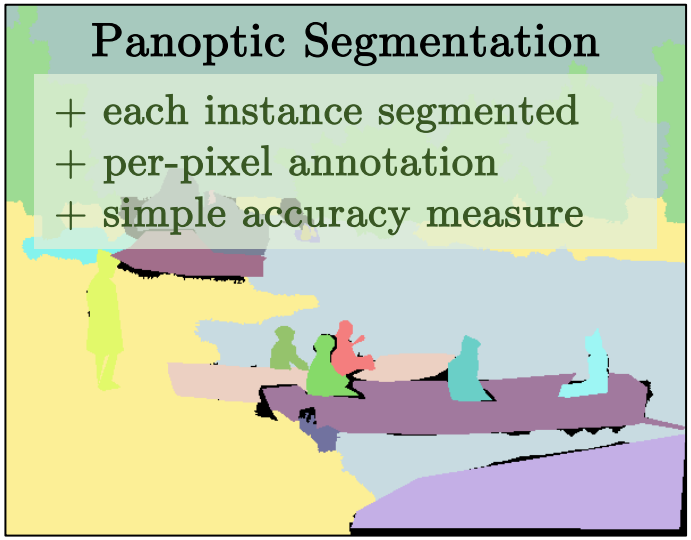
Semantic Segmentation

+ per-pixel annotation
+ simple accuracy measure
– instances indistinguishable

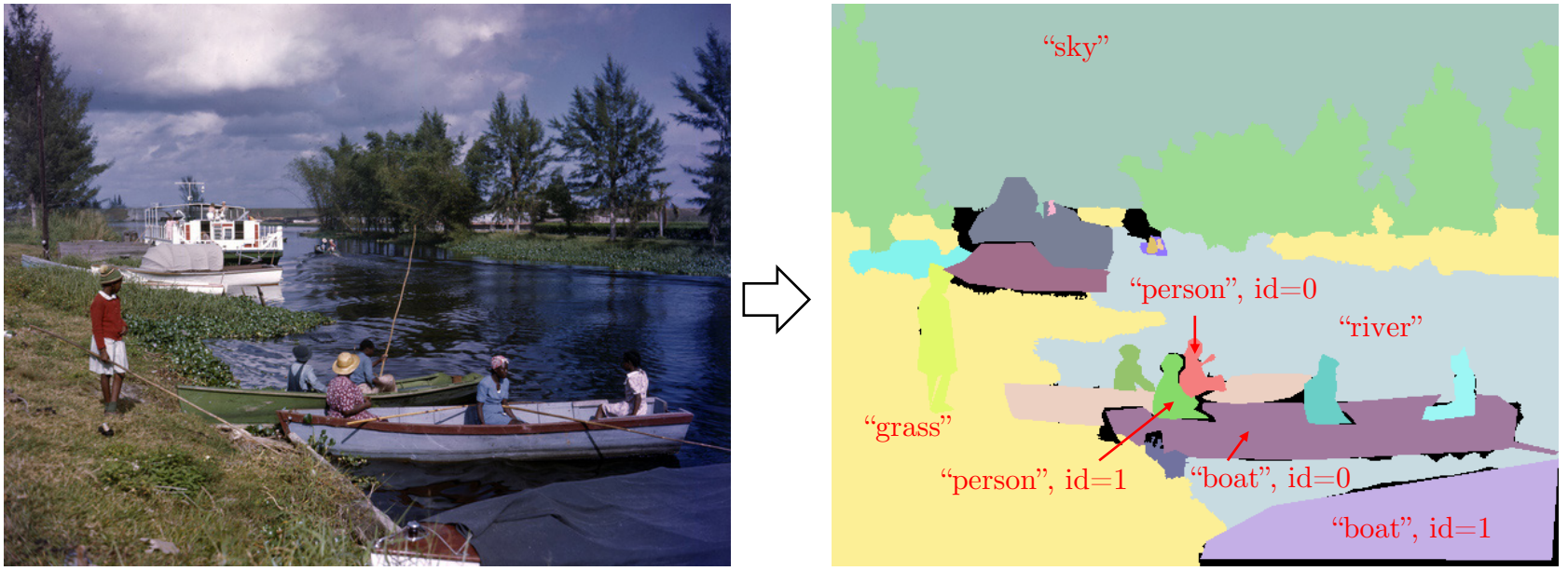


Panoptic Segmentation

+ each instance segmented
+ per-pixel annotation
+ simple accuracy measure

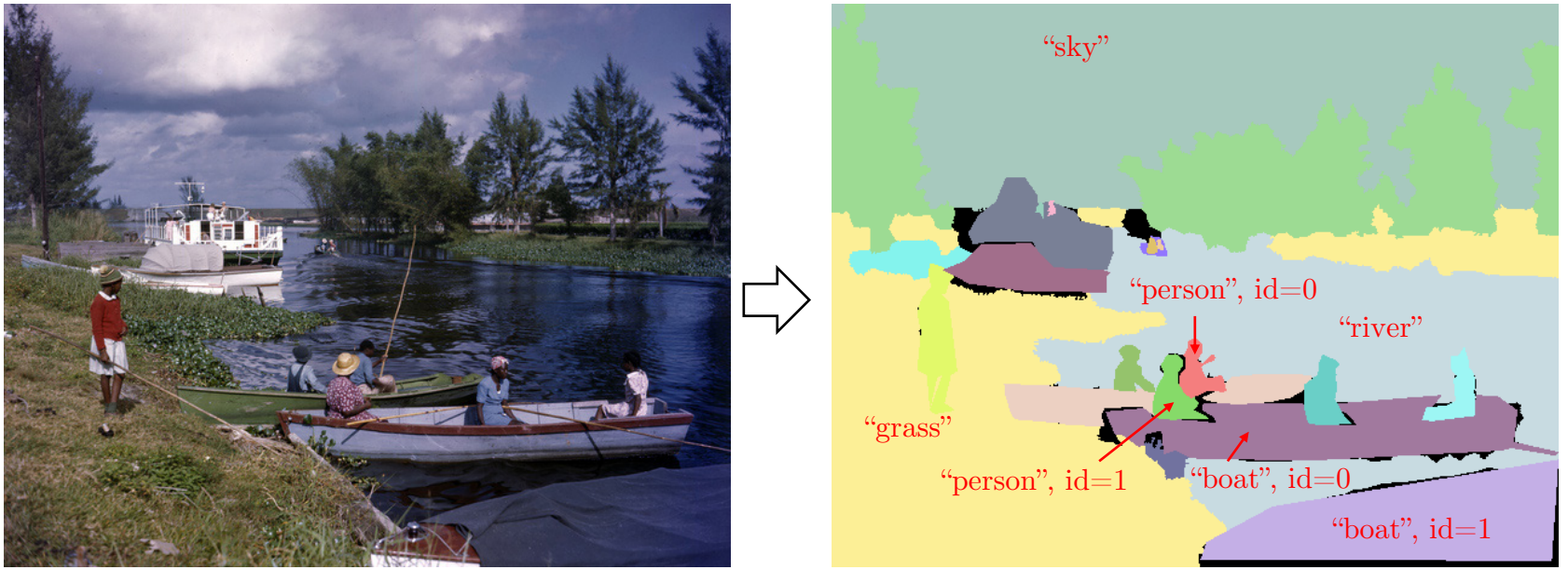


Panoptic Segmentation



- For each pixel i predict semantic label l and instance id z
- Note no overlaps between segments by design
- We introduce a simple and intuitive metric
- Popular dataset are panoptic ready
- Simple! Obvious! Natural!
  - **And yet not currently a studied task...**
- Note: does not address parts, hierarchy, amodal segmentation

Panoptic Segmentation



Datasets	Instance Segmentation	Semantic Segmentation
ADE20k/Places	+	+
CityScapes	+	+
Mapillary Vistas	+	+
COCO*	+	+

Panoptic Segmentation

Previous attempts to explore the combined task:

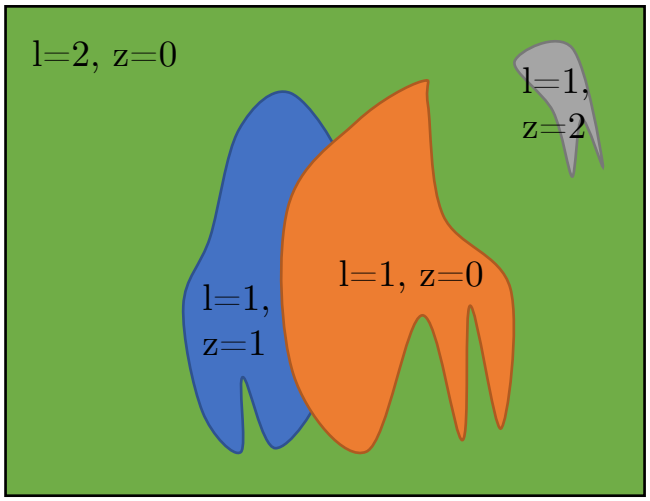
- Scene Parsing
- Image Parsing
- Holistic Scene Understanding

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.

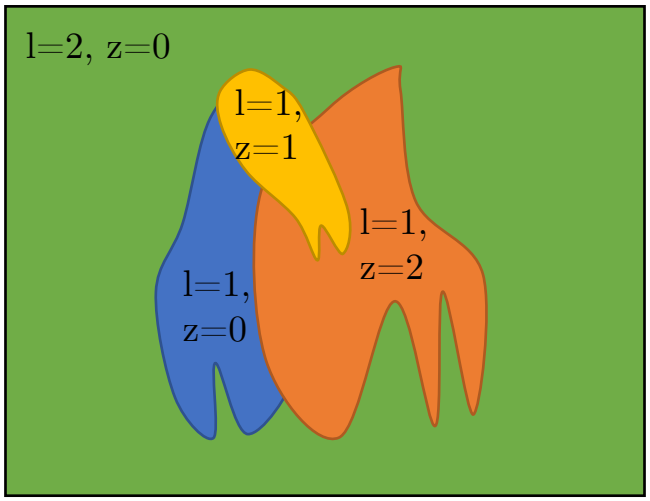
Outline

- Motivation & Definition
- **Quality Evaluation**
- Human Performance
- Machine Performance
- Perspectives

Panoptic Quality (PQ)

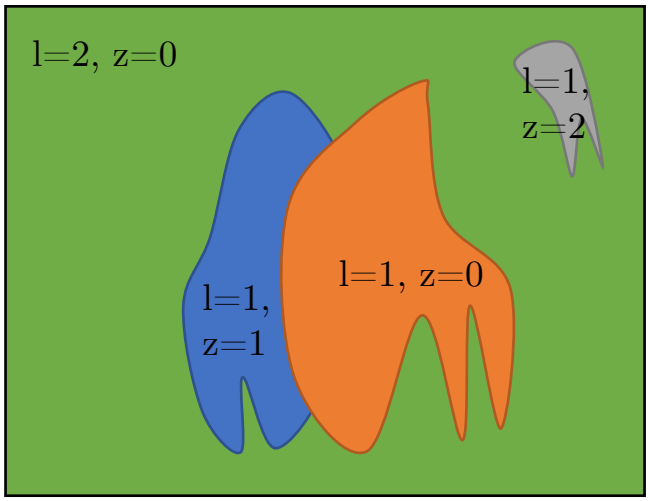


Ground Truth

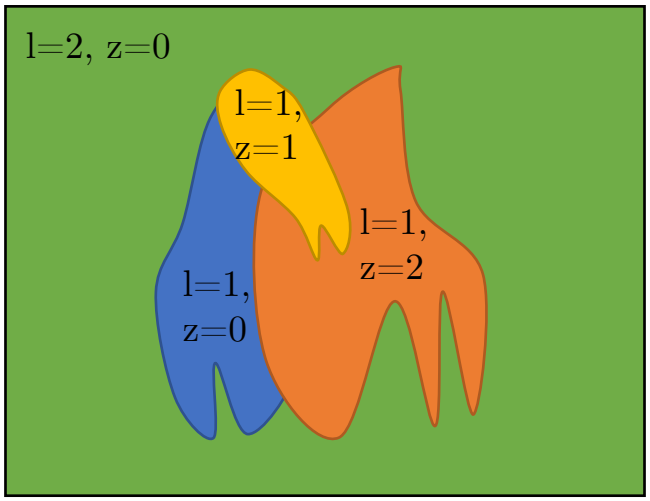


Prediction

Panoptic Quality (PQ)



Ground Truth



Prediction

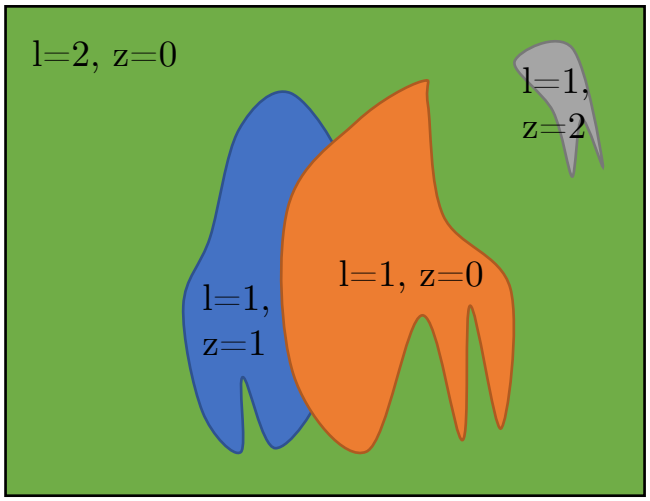
PQ Desiderata:

- Completeness
- Interpretability
- Simplicity

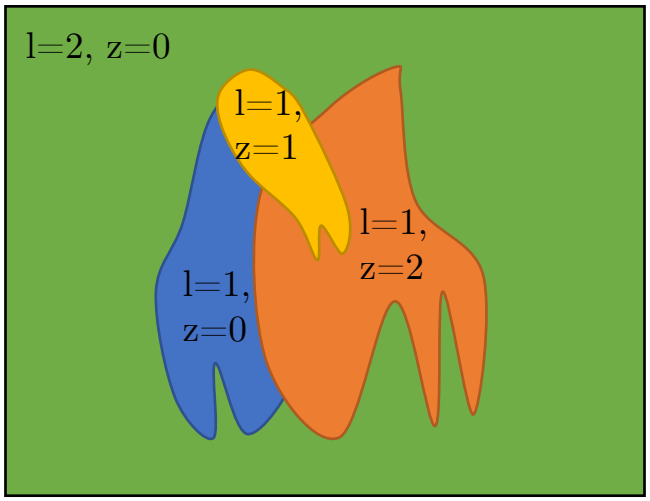
PQ Computation:

- Step 1: Matching
- Step 2: Calculation

Panoptic Quality (PQ): Matching



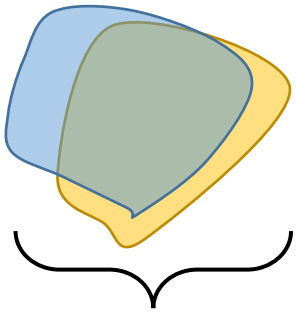
Ground Truth



Prediction

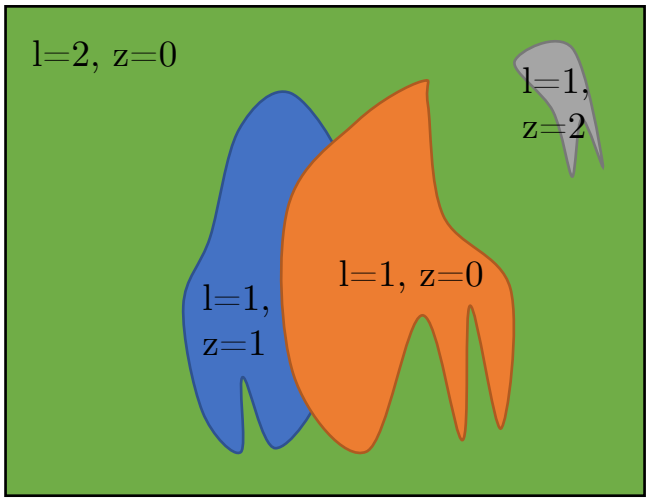
Theorem: For panoptic segmentation problem each ground truth segment can have at most one corresponding predicted segment with IoU greater than 0.5

Proof sketch:

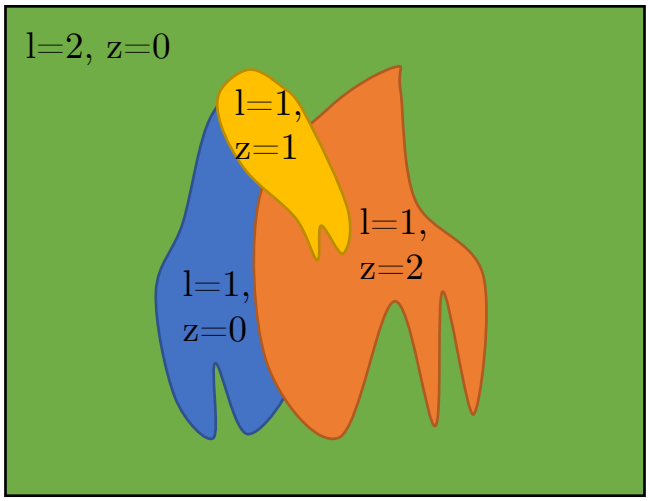
if 

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

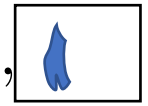
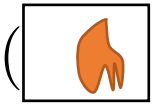
Panoptic Quality (PQ): Matching



Ground Truth



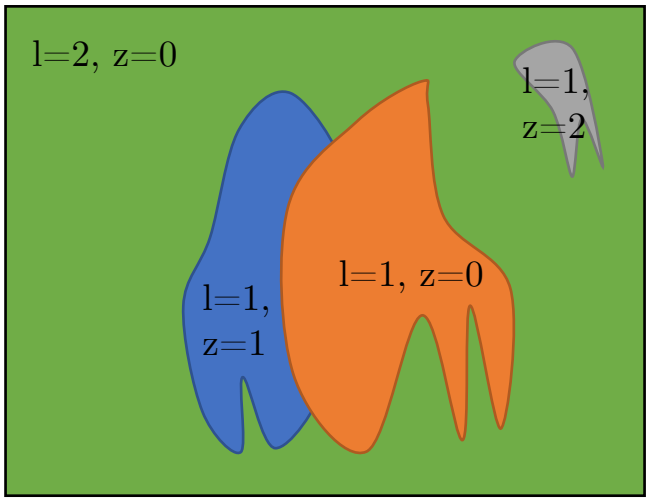
Prediction

$\text{TP} = \{($  ,  $),\ ($  ,  $)\}$

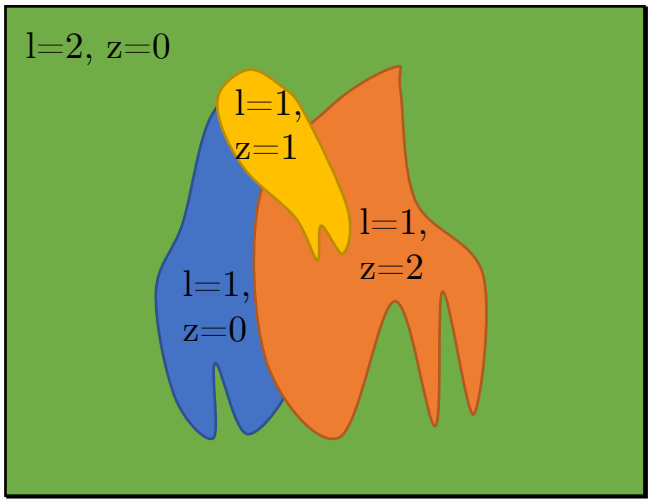
$\text{FP} = \{$  $\}$

$\text{FN} = \{$  $\}$

Panoptic Quality (PQ): Calculation



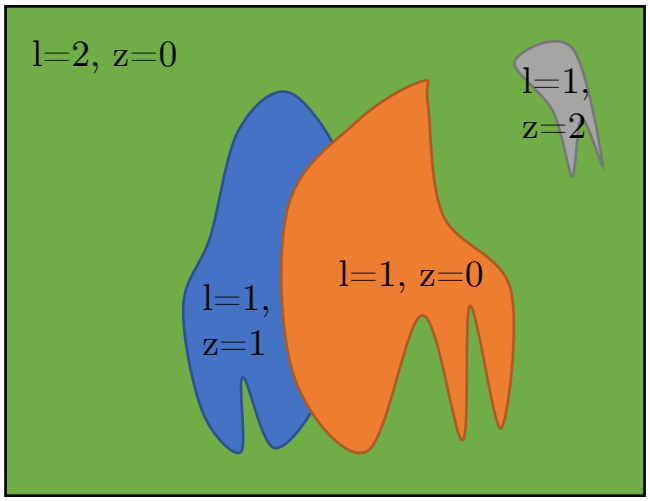
Ground Truth



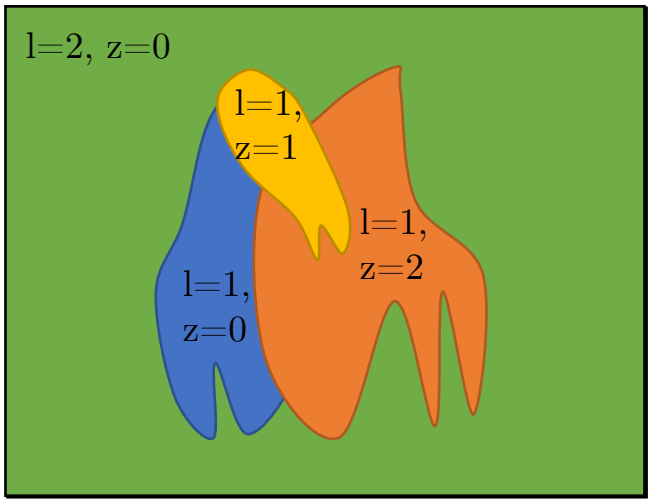
Prediction

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

Panoptic Quality (PQ): Calculation



Ground Truth



Prediction

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

Outline

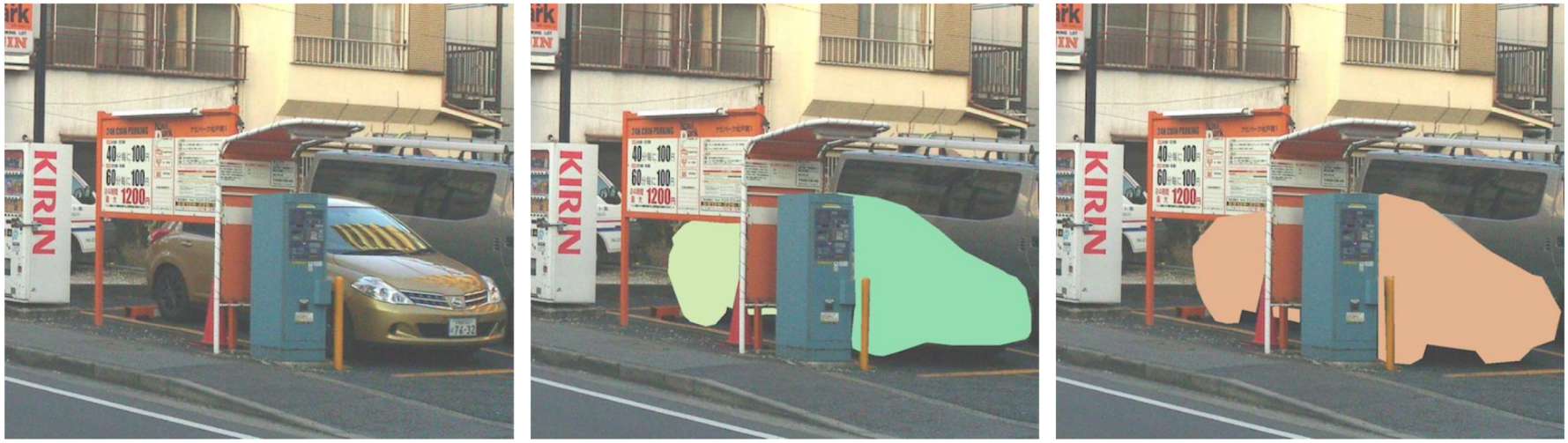
- Motivation & Definition
- Quality Evaluation
- **Human Performance**
- Machine Performance
- Perspectives

Human Performance: images annotated twice

Cityscapes	Mapillary Vistas	ADE20k	COCO
30 images	46 images	64 images	5000 images

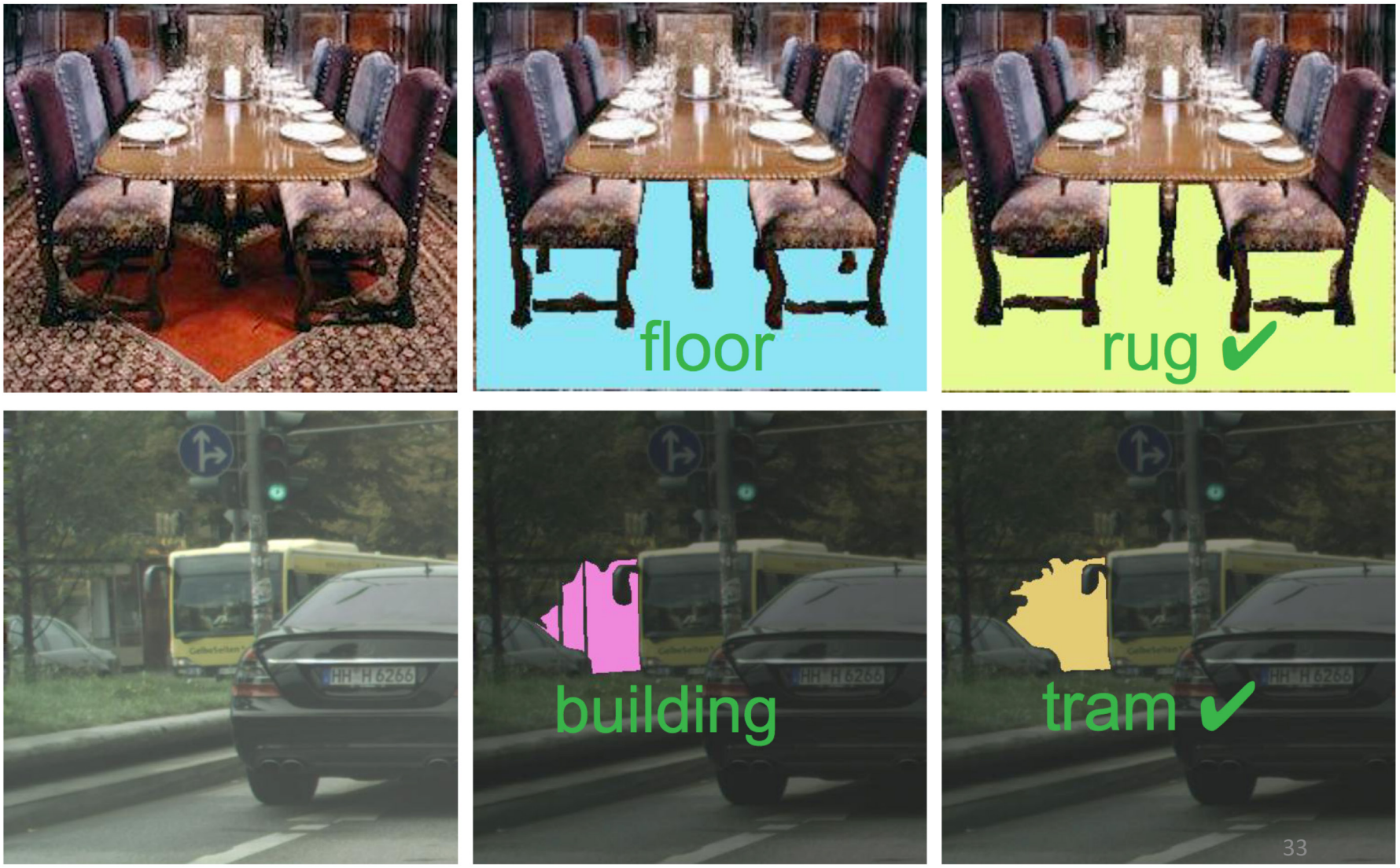
Human Performance: Segmentation Flaws

Mapillary Vistas



Cityscapes

Human Performance: Classification Flaws



Human Performance: PQ

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

Human Performance: PQ

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

	PQ
Cityscapes	69.7
ADE20k	67.1
Vistas	57.5
COCO	53.5

Human Performance: PQ = SQ x RQ

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

	PQ	SQ	RQ
Cityscapes	69.7	84.2	82.1
ADE20k	67.1	85.8	78.0
Vistas	57.5	79.5	71.4
COCO	53.5	82.6	63.9

Notes:

- Mean(SQ) > Mean(RQ)
- Var(SQ) < Var(RQ)
- RQ depends on # categories

Human Performance: Stuff *vs.* Things

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

	PQ	$\mathrm{PQ}^{\mathrm{St}}$	$\mathrm{PQ}^{\mathrm{Th}}$
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
		⇧	⇧

Notes:

- $\mathrm{PQ}^{\mathrm{ST}}$ similar to $\mathrm{PQ}^{\mathrm{TH}}$

Human Performance: All

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

	PQ	PQ^{St}	PQ^{Th}	SQ	SQ^{St}	SQ^{Th}	RQ	RQ^{St}	RQ^{Th}
Cityscapes	69.7	71.3	67.4	84.2	84.4	83.9	82.1	83.4	80.2
ADE20k	67.1	70.3	65.9	85.8	85.5	85.9	78.0	82.4	76.4
Vistas	57.5	62.6	53.4	79.5	81.6	77.9	71.4	76.0	67.7
COCO	53.5	47.1	57.8	82.6	84.3	81.4	63.9	55.2	69.7

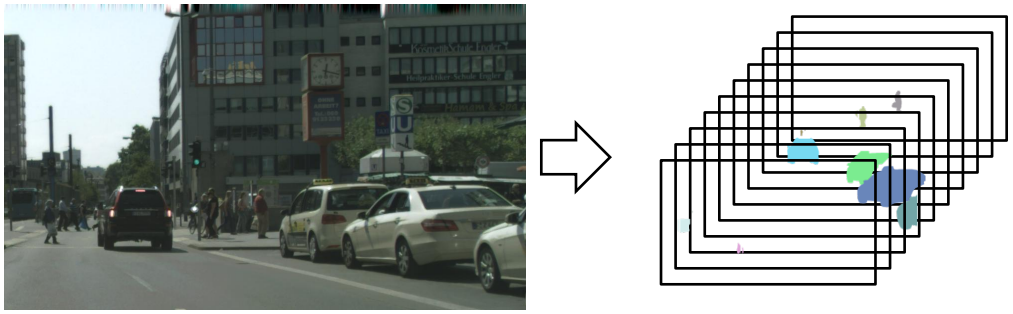
Notes:

- Mean(SQ) > Mean(RQ)
- Var(SQ) < Var(RQ)
- RQ depends on # categories

Outline

- Motivation & Definition
- Quality Evaluation
- Human Performance
- **Machine Performance**
- Perspectives

Machine Performance: Baselines

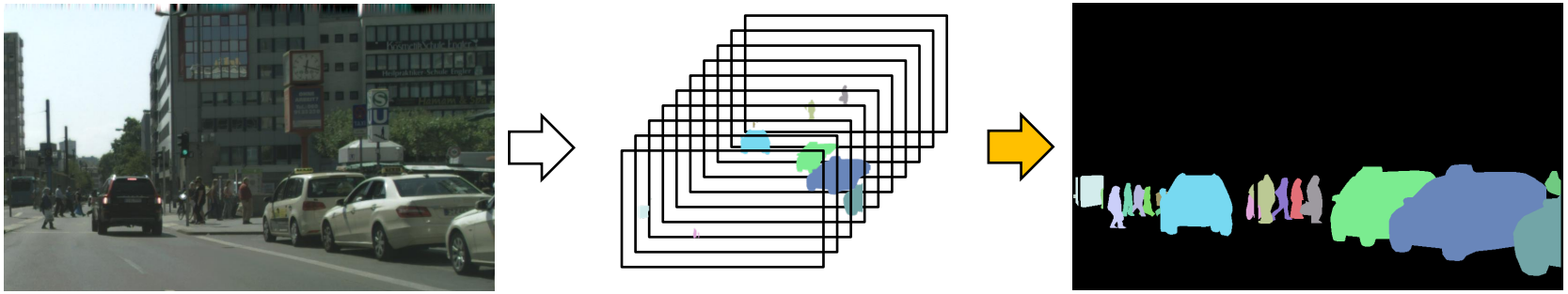


overlapping instances

Notes:

- Typical instance segmenters produce overlapping outputs

Machine Performance: Baselines



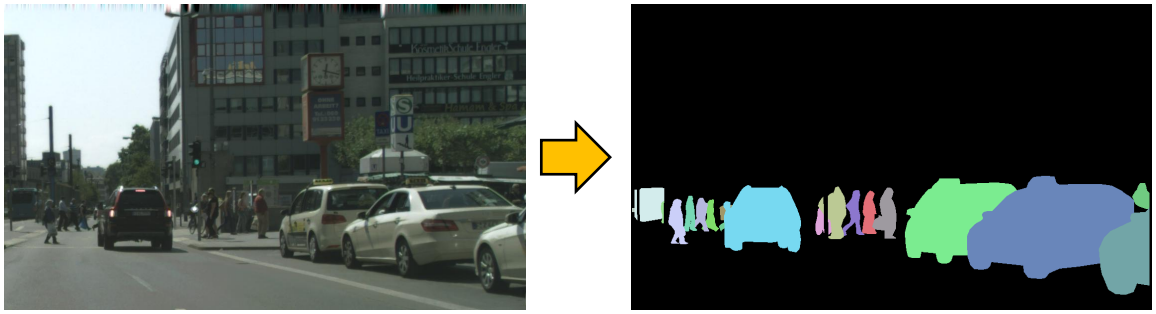
overlapping instances

non-overlapping instances

Notes:

- Typical instance segmenters produce overlapping outputs
- Post-processing can remove overlaps (NMS++)

Machine Performance: Baselines

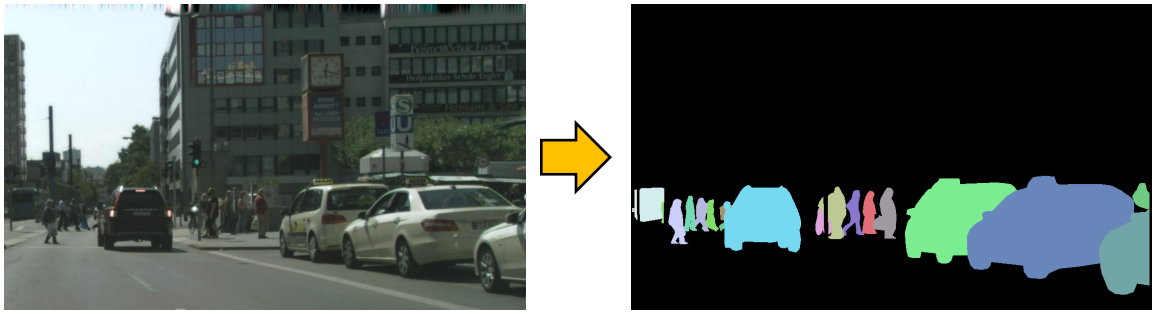


non-overlapping instances

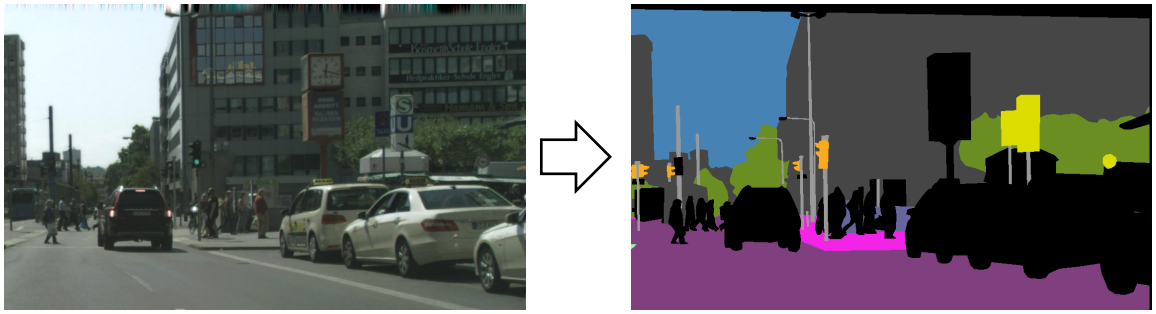
Notes:

- Typical instance segmenters produce overlapping outputs
- Post-processing can remove overlaps (NMS++)
- For simplicity assume always apply this post-processing

Machine Performance: Baselines



non-overlapping instances

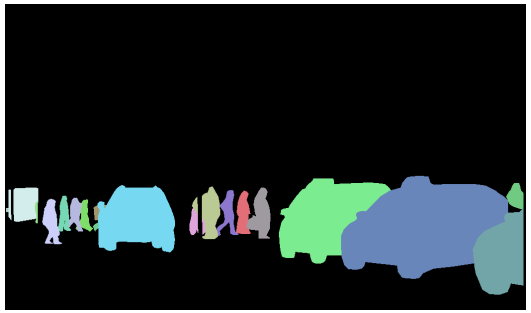


semantic segmentation

Notes:

- Run instance segmentation independently
- Run semantic segmentation independently

Machine Performance: Baselines



non-overlapping instances

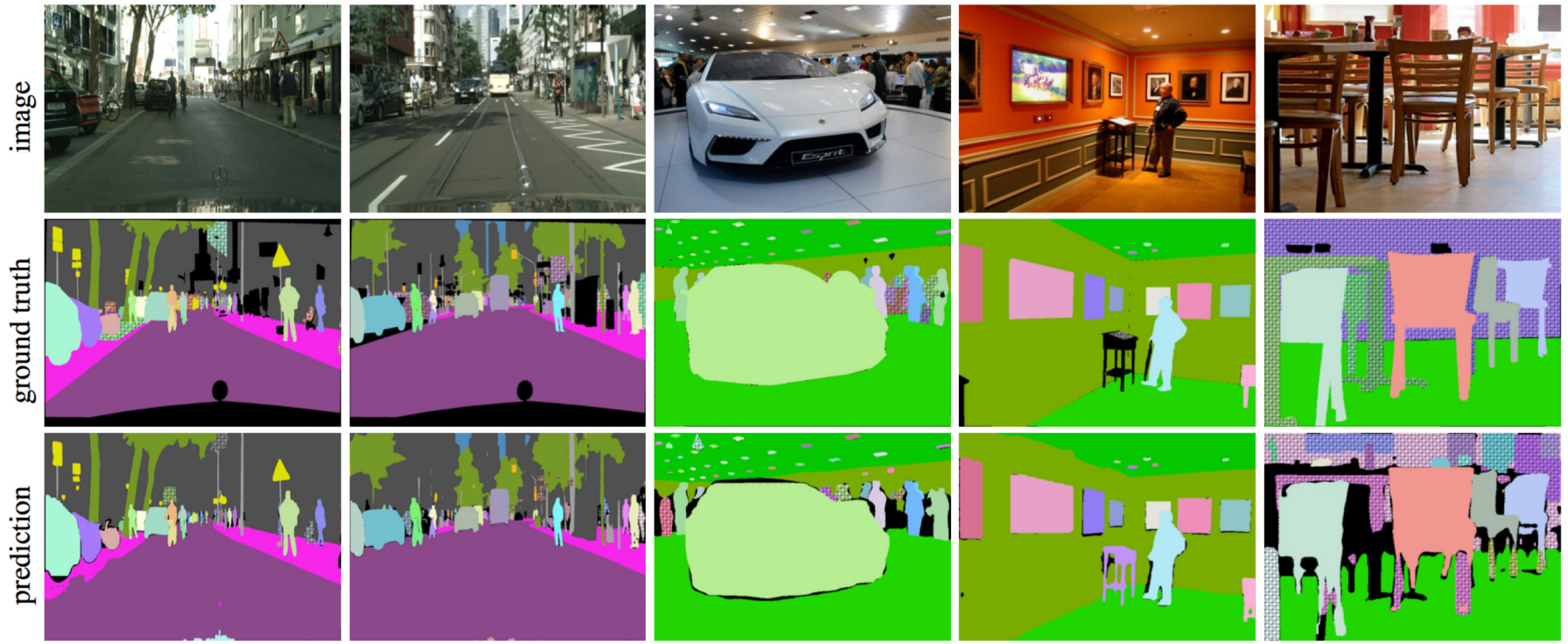
panoptic segmentation

semantic segmentation

Notes:

- Run instance segmentation independently
- Run semantic segmentation independently
- Simple post-processing to merge

Machine Performance: Baselines



Machine Performance: Instance Segmentation

Cityscapes	AP	AP^{NO}	PQ^{Th}	SQ^{Th}	RQ^{Th}
Mask R-CNN+COCO [He et al. 2017]	**36.4**	**33.1**	**54.0**	**79.4**	**67.8**
Mask R-CNN [He et al. 2017]	31.5	28.0	49.6	78.7	63.0

ADE20k	AP	AP^{NO}	PQ^{Th}	SQ^{Th}	RQ^{Th}
Megvii [Xiao et al. 2017]	**30.1**	**24.8**	**41.1**	**81.6**	**49.6**
G-RMI [Fathi et al. 2017]	24.6	20.6	35.3	79.3	43.2

Notes:

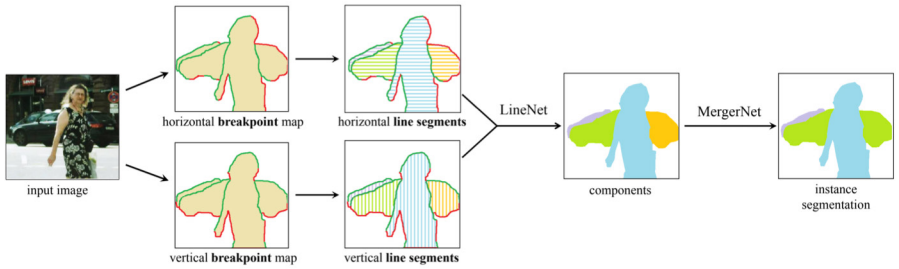
- Bigger AP => bigger PQ
- AP^{NO}: AP for prediction without overlaps

Machine Performance: Instance Segmentation

Cityscapes	AP	AP^{NO}	PQ^{Th}	SQ^{Th}	RQ^{Th}
Mask R-CNN+COCO [He et al. 2017]	**36.4**	**33.1**	**54.0**	**79.4**	**67.8**
Mask R-CNN [He et al. 2017]	31.5	28.0	49.6	78.7	63.0
SGN [Liu et al. 2017]	29.2	29.2	-	-	-
ADE20k	AP	AP^{NO}	PQ^{Th}	SQ^{Th}	RQ^{Th}
Megvii [Xiao et al. 2017]	**30.1**	**24.8**	**41.1**	**81.6**	**49.6**
G-RMI [Fathi et al. 2017]	24.6	20.6	35.3	79.3	43.2

Notes:

- Bigger AP => bigger PQ
- AP^{NO}: AP for prediction without overlaps
- PQ favors methods that don't produce overlaps (Liu et al.)



Machine Performance: Semantic Segmentation

Cityscapes	IoU	PQ^{St}	SQ^{St}	RQ^{St}
PSPNet multi-scale [Zhao et al. 2017]	**80.6**	**66.6**	**82.2**	**79.3**
PSPNet single-scale [Zhao et al. 2017]	79.6	65.2	81.6	78.0

ADE20k	IoU	PQ^{St}	SQ^{St}	RQ^{St}
CASIA_IVA_JD [Fu et al. 2017]	**32.3**	**27.4**	**61.9**	**33.7**
G-RMI [Fathi et al. 2017]	30.6	19.3	58.7	24.3

Notes:

- Bigger IOU => bigger PQ
- PQ Penalizes spurious small segments

Machine Performance: Panoptic Segmentation

Cityscapes	PQ	SQ	RQ	PQ^{St}	PQ^{Th}
human	$69.6^{+2.5}_{-2.7}$	$84.1^{+0.8}_{-0.8}$	$82.0^{+2.7}_{-2.9}$	$71.2^{+2.3}_{-2.5}$	$67.4^{+4.6}_{-4.9}$
machine	61.2	80.9	74.4	66.4	54.0
ADE20k	PQ	SQ	RQ	PQ^{St}	PQ^{Th}
human	$67.6^{+2.0}_{-2.0}$	$85.7^{+0.6}_{-0.6}$	$78.6^{+2.1}_{-2.1}$	$71.0^{+3.7}_{-3.2}$	$66.4^{+2.3}_{-2.4}$
machine	35.6	74.4	43.2	24.5	41.1
Vistas	PQ	SQ	RQ	PQ^{St}	PQ^{Th}
human	$57.7^{+1.9}_{-2.0}$	$79.7^{+0.8}_{-0.7}$	$71.6^{+2.2}_{-2.3}$	$62.7^{+2.8}_{-2.8}$	$53.6^{+2.7}_{-2.8}$
machine	38.3	73.6	47.7	41.8	35.7

Notes:

- There is a healthy gap between Human and Machine PQ
- Comparisons imperfect (different data used)

Outline

- Motivation & Definition
- Quality Evaluation
- Human Performance
- Machine Performance
- **Perspectives**

Why Panoptic Segmentation?

Cityscapes | Mapillary Vistas | ADE20k | COCO

- Unified stuff and things segmentation
- New challenge for our community
- Potential expected and unexpected innovations

COCO + Mapillary

Joint Recognition Challenge Workshop at ECCV 2018

http://cocodataset.org/workshop/coco-mapillary-eccv-2018.html
(or just browse to http://cocodataset.org/ and go from there)