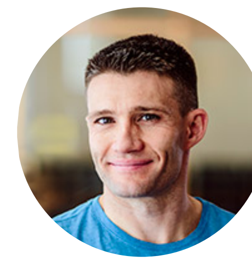
Panoptic Segmentation: Unifying Semantic and Instance Segmentation







Alex Kirillov   Carsten Rother   Kaiming He   Ross Girshick   Piotr Dollár

UNIVERSITÄT HEIDELBERG

FACEBOOK AI RESEARCH

Unifying Semantic and Instance Segmentation

Unifying Semantic and Instance Segmentation

Semantic Segmentation

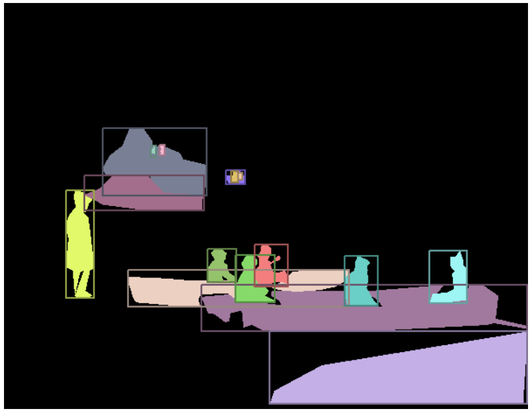
Unifying Semantic and Instance Segmentation

Semantic Segmentation

Object Detection

Unifying Semantic and Instance Segmentation

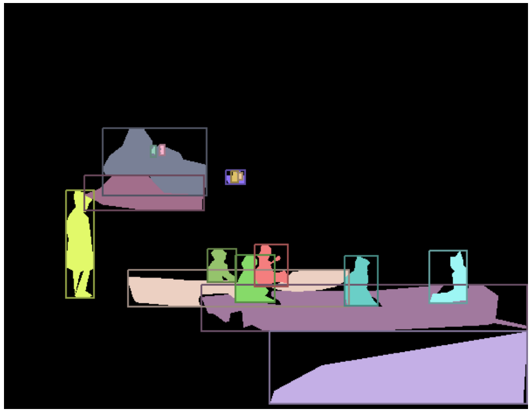
Semantic Segmentation



Object Detection/Seg

Unifying Semantic and Instance Segmentation

Semantic Segmentation



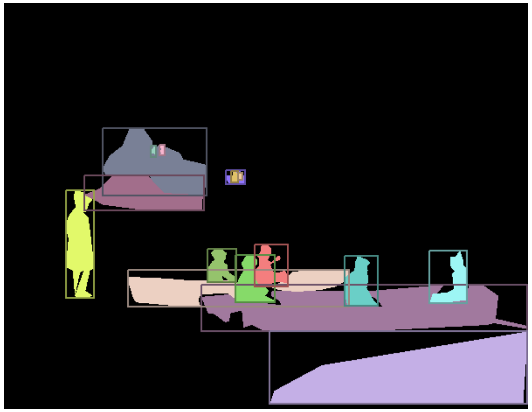
Object Detection/Seg

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

Unifying Semantic and Instance Segmentation

Semantic Segmentation

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



Object Detection/Seg

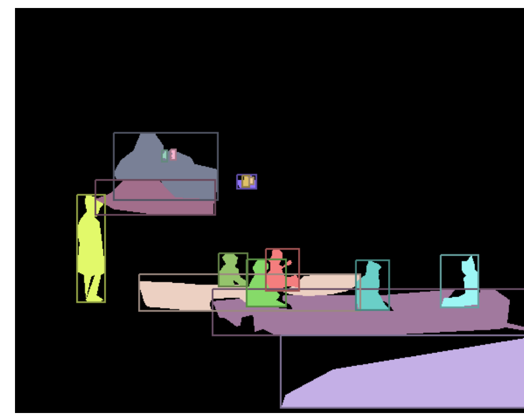
- each object detected and segmented separately
- "stuff" is not segmented

Unifying Semantic and Instance Segmentation

Semantic Segmentation

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

?



Object Detection/Seg

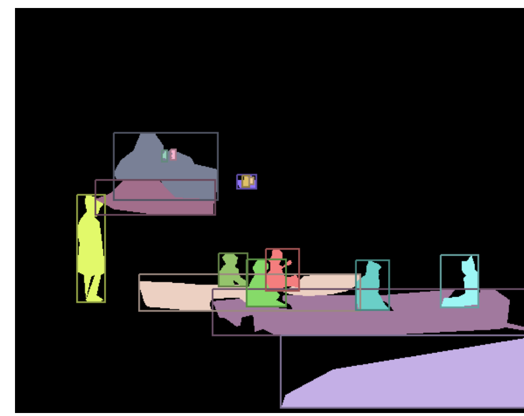
- each object detected and segmented separately
- "stuff" is not segmented

Unifying Semantic and Instance Segmentation

Semantic Segmentation

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

Panoptic Segmentation



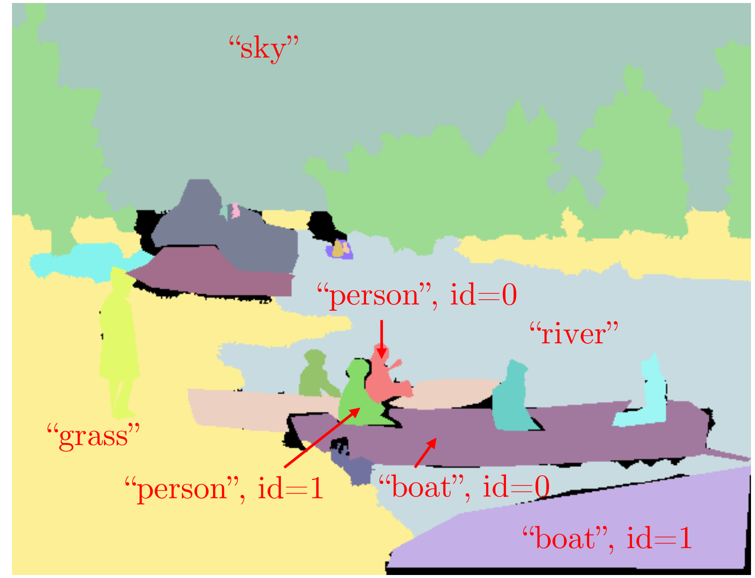
Object Detection/Seg

- each object detected and segmented separately
- "stuff" is not segmented

Outline

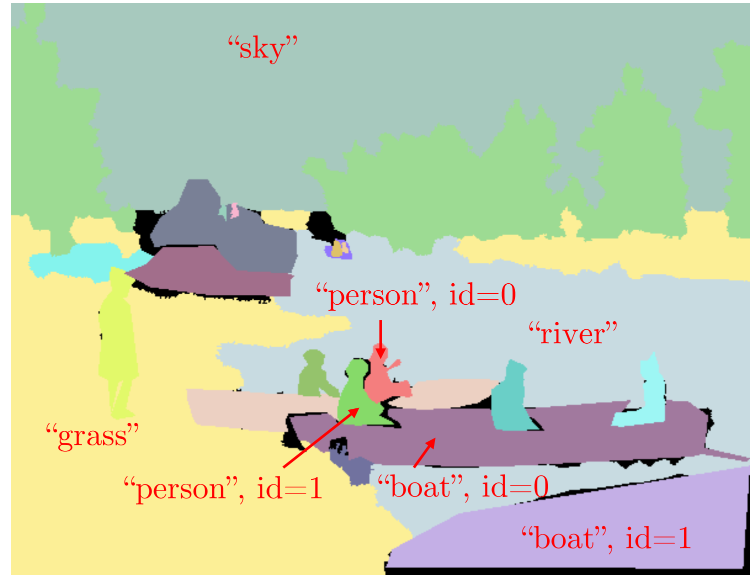
- ➢Motivation
- ➢**Problem Definition**
- ➢Quality Evaluation
- ➢Human Performance
- ➢Humans vs Computers
- ➢Perspectives

Panoptic Segmentation



For each pixel i predict semantic label l **and instance id** $\boldsymbol{z}$

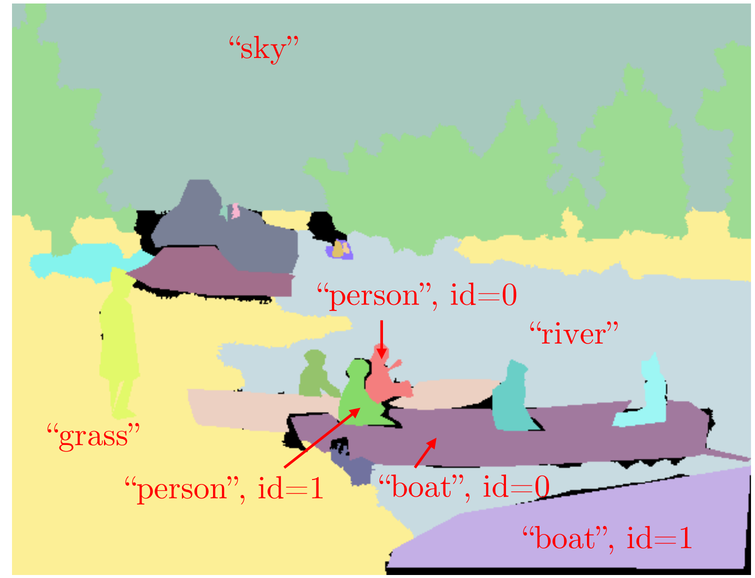
Panoptic Segmentation



For each pixel i predict semantic label l **and instance id** $\boldsymbol{z}$

- no overlaps between segments

Panoptic Segmentation

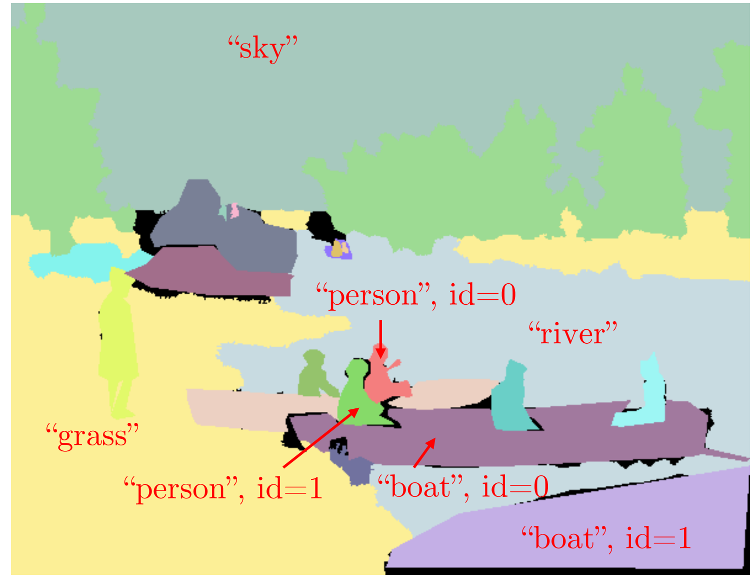


For each pixel i predict semantic label l **and instance id** $\boldsymbol{z}$

- ➢ no overlaps between segments

- Popular datasets can be used
- We introduce simple, intuitive metric
- Drive novel algorithmic ideas

Popular datasets can be used



For each pixel i predict semantic label l **and instance id z**

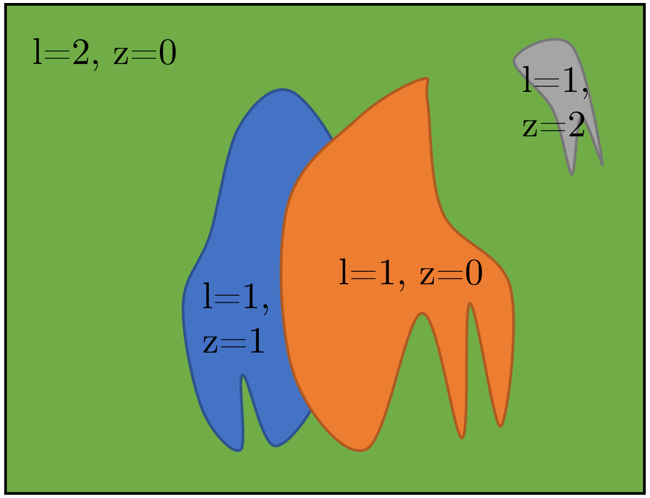
Datasets	Instance Segmentation	Semantic Segmentation
COCO*	+	+
ADE20k/Places	+	+
CityScapes	+	+
Mapillary Vistas	+	+

*COCO has overlaps (no depth order)

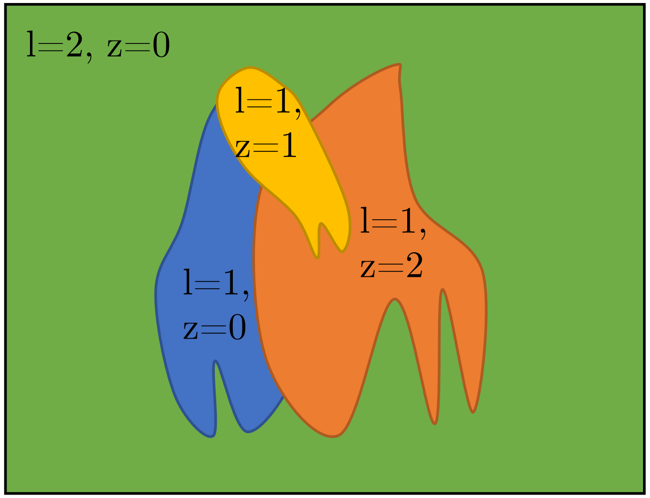
Outline

- ➢Motivation
- ➢Problem Definition
- ➢**Quality Evaluation**
- ➢Human Performance
- ➢Humans vs Computers
- ➢Perspectives

Quality Evaluation

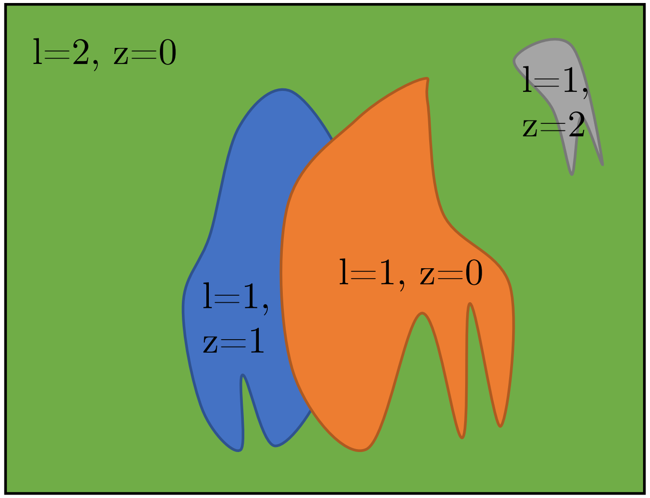


Ground Truth

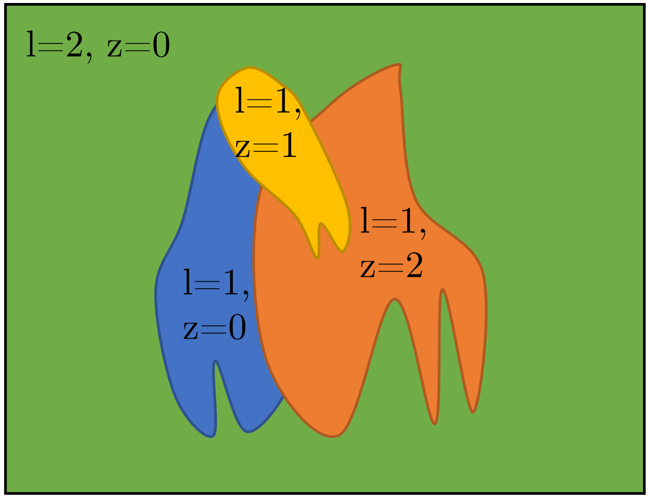


Prediction

Quality Evaluation



Ground Truth

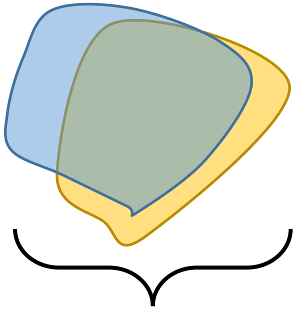


Prediction

Theorem: Matching is unique if overlapping threshold **>** 0.5 IoU and both ground truth and prediction have no overlaps.

Proof sketch:

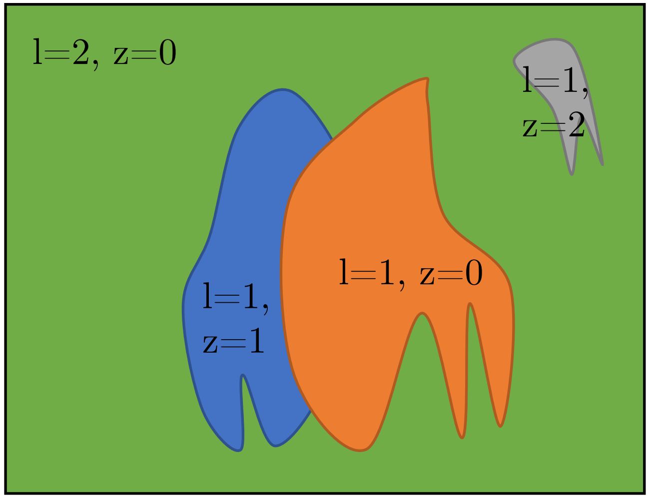
if



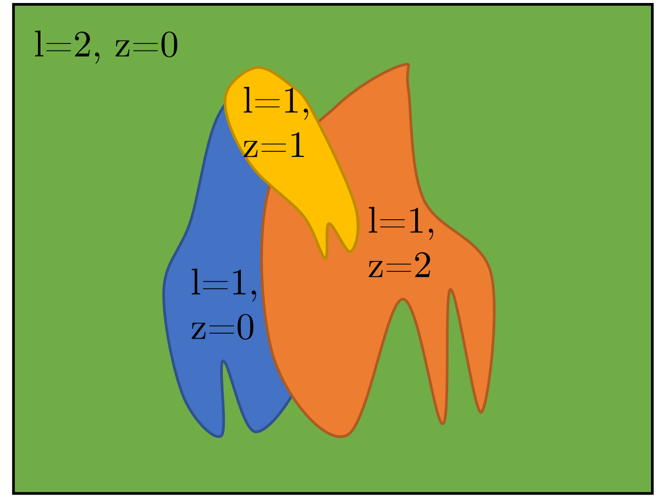
$\text{IoU} > 0.5$

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

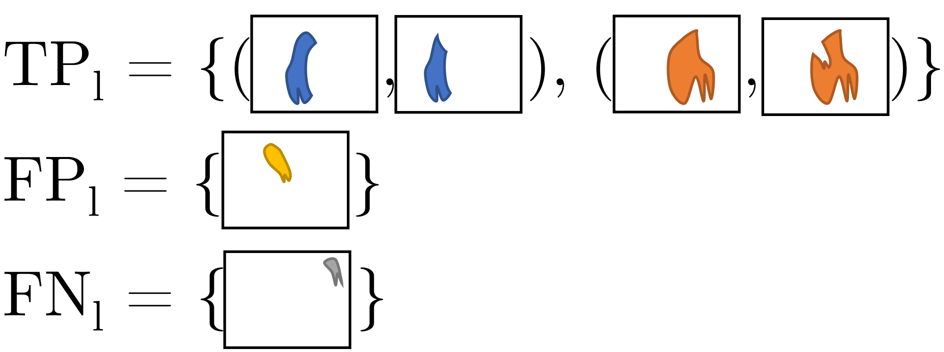
Quality Evaluation



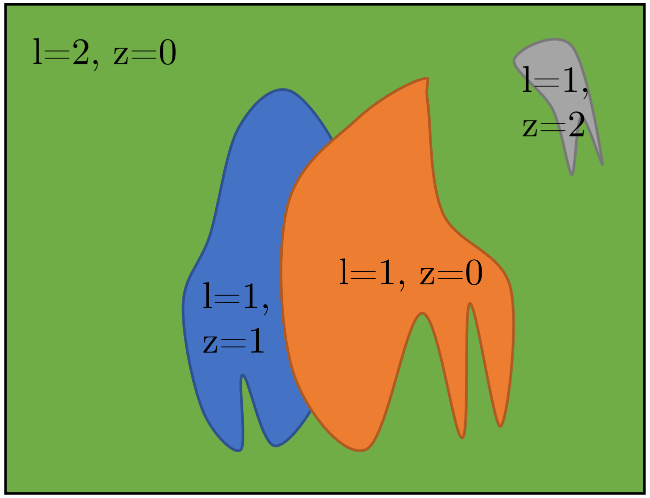
Ground Truth



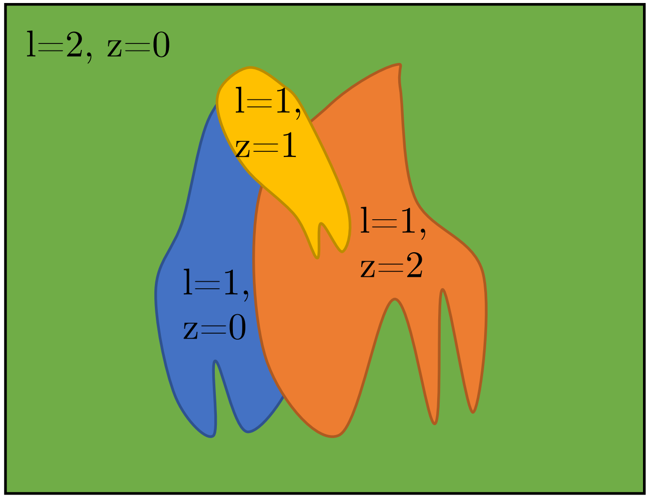
Prediction



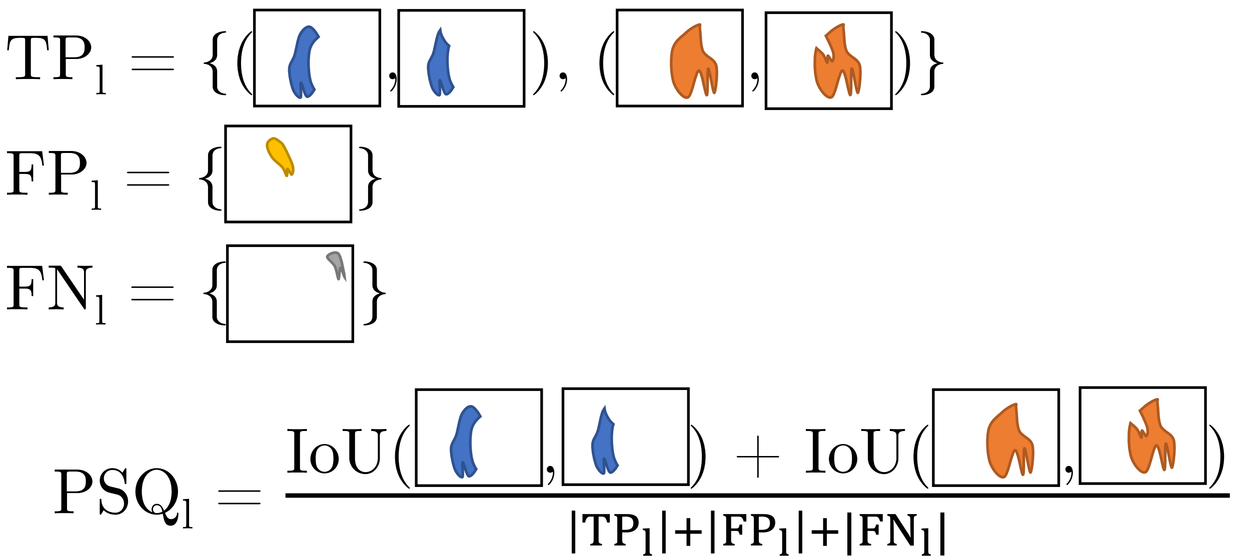
Quality Evaluation



Ground Truth

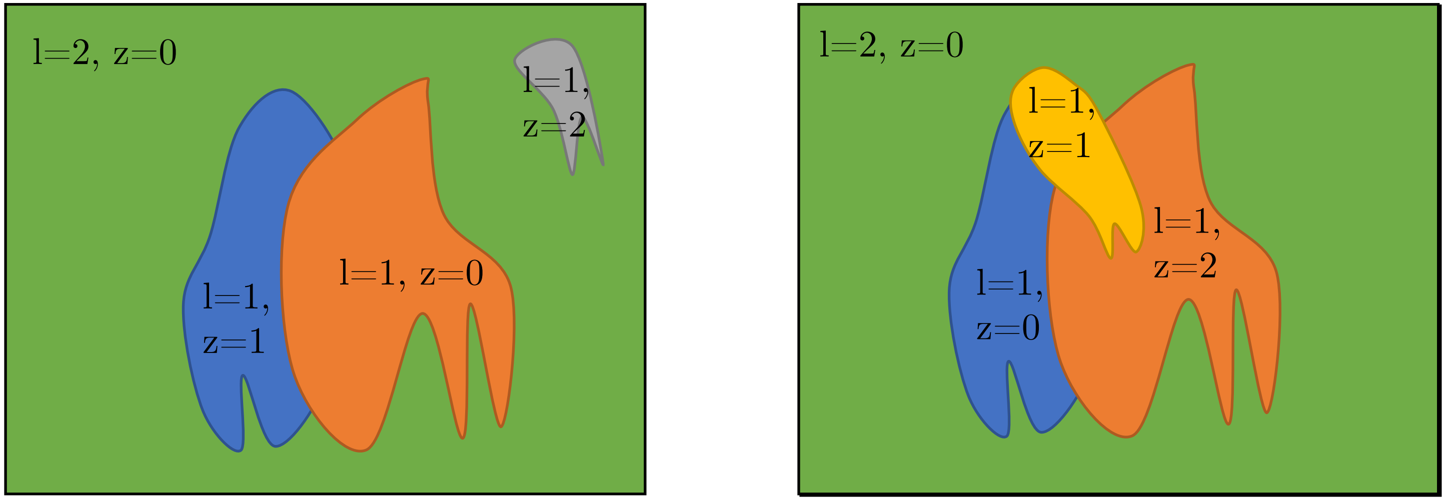


Prediction



$$= \frac{\sum_{(g,p)\in TP_l} \mathrm{IoU}(g,p)}{|TP_l|+|FP_l|+|FN_l|}$$

Quality Evaluation



$$\mathrm{PSQ}_l = \frac{\sum_{(g,p)\in TP_l} \mathrm{IoU}(g,p)}{|TP_l|+|FP_l|+|FN_l|} = \underbrace{\frac{\sum_{(g,p)\in TP_l} \mathrm{IoU}(g,p)}{|TP_l|}}_{\text{Segmentation Quality}} \cdot \underbrace{\frac{|TP_l|}{|TP_l|+|FP_l|+|FN_l|}}_{\text{Detection Quality}}$$

Outline

- Motivation
- Problem Definition
- Quality Evaluation
- **Human Performance**
- Humans vs Computers
- Perspectives

Panoptic Segmentation Quality (PSQ)

$$\mathrm{PSQ_l} = \frac{\sum_{(g,p)\in \mathrm{TP_l}} \mathrm{IoU}(g,p)}{|\mathrm{TP_l}|+|\mathrm{FP_l}|+|\mathrm{FN_l}|} = \underbrace{\frac{\sum_{(g,p)\in \mathrm{TP_l}} \mathrm{IoU}(g,p)}{|\mathrm{TP_l}|}}_{\text{Seg Quality}} \cdot \underbrace{\frac{|\mathrm{TP_{l=1}}|}{|\mathrm{TP_l}|+|\mathrm{FP_l}|+|\mathrm{FN_l}|}}_{\text{Det Quality}}$$

Panoptic Segmentation Quality (PSQ)

$$\mathrm{PSQ}_l = \frac{\sum_{(\mathrm{g},\mathrm{p})\in \mathrm{TP}_l} \mathrm{IoU}(\mathrm{g},\mathrm{p})}{|\mathrm{TP}_l|+|\mathrm{FP}_l|+|\mathrm{FN}_l|} = \underbrace{\frac{\sum_{(\mathrm{g},\mathrm{p})\in \mathrm{TP}_l} \mathrm{IoU}(\mathrm{g},\mathrm{p})}{|\mathrm{TP}_l|}}_{\text{Seg Quality}} \cdot \underbrace{\frac{|\mathrm{TP}_{l=1}|}{|\mathrm{TP}_l|+|\mathrm{FP}_l|+|\mathrm{FN}_l|}}_{\text{Det Quality}}$$

no confidence scores

$\Downarrow$

human performance can be measured

Panoptic Segmentation Quality (PSQ)

$$\mathrm{PSQ}_l = \frac{\sum_{(g,p)\in \mathrm{TP}_l} \mathrm{IoU}(g,p)}{|\mathrm{TP}_l|+|\mathrm{FP}_l|+|\mathrm{FN}_l|} = \underbrace{\frac{\sum_{(g,p)\in \mathrm{TP}_l} \mathrm{IoU}(g,p)}{|\mathrm{TP}_l|}}_{\text{Seg Quality}} \cdot \underbrace{\frac{|\mathrm{TP}_{l=1}|}{|\mathrm{TP}_l|+|\mathrm{FP}_l|+|\mathrm{FN}_l|}}_{\text{Det Quality}}$$

no confidence scores

$\Downarrow$

human performance can be measured

CityScapes: 30 images were annotated independently twice.

Panoptic Segmentation Quality (PSQ)

$$\mathrm{PSQ}_l = \frac{\sum_{(g,p)\in TP_l} \mathrm{IoU}(g,p)}{|TP_l|+|FP_l|+|FN_l|} = \underbrace{\frac{\sum_{(g,p)\in TP_l} \mathrm{IoU}(g,p)}{|TP_l|}}_{\text{Seg Quality}} \cdot \underbrace{\frac{|TP_{l=1}|}{|TP_l|+|FP_l|+|FN_l|}}_{\text{Det Quality}}$$

no confidence scores
⇩
human performance can be measured

CityScapes: 30 images were annotated independently twice.

class	PSQ	Seg Quality	Det Quality
car	66.6%	87.5%	76.2%
person	61.8%	80.8%	76.4%
motorcycle	51.8%	77.8%	66.7%
pole	46.9%	70.3%	66.7%
road	98.0%	98.0%	100.0%
traffic sign	67.1%	79.5%	84.4%
average	**62.6%**	**83.9%**	**73.43%**

All Objects

Panoptic Segmentation Quality (PSQ)

$$\mathrm{PSQ}_l = \frac{\sum_{(g,p)\in TP_l} \mathrm{IoU}(g,p)}{|TP_l|+|FP_l|+|FN_l|} = \underbrace{\frac{\sum_{(g,p)\in TP_l} \mathrm{IoU}(g,p)}{|TP_l|}}_{\text{Seg Quality}} \cdot \underbrace{\frac{|TP_{l=1}|}{|TP_l|+|FP_l|+|FN_l|}}_{\text{Det Quality}}$$

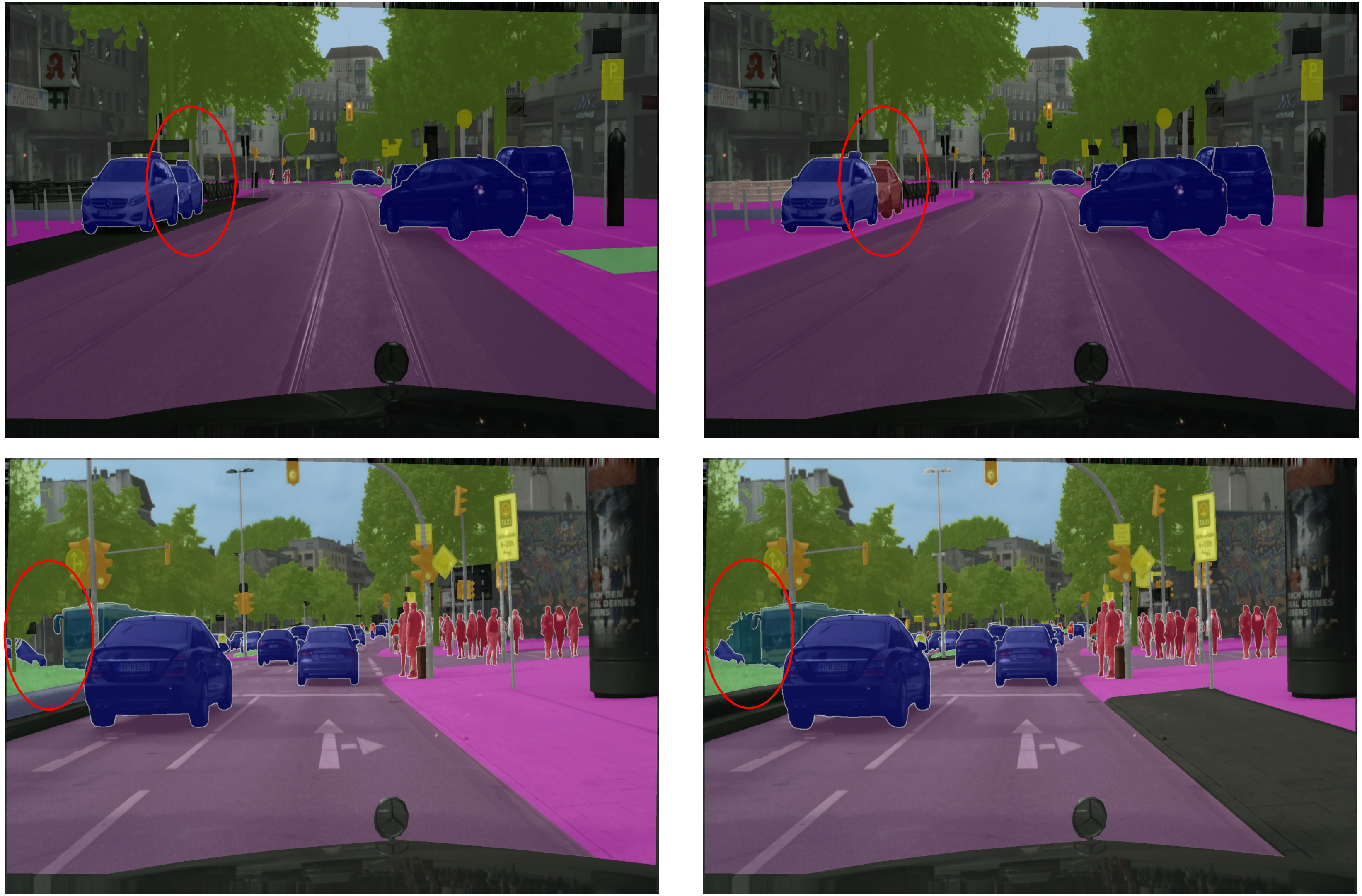
no confidence scores
⇩
human performance can be measured

CityScapes: 30 images were annotated independently twice.

class	PSQ	Seg Quality	Det Quality
car	89.4%	91.3%	97.9%
person	82.0%	78.1%	94.1%
motorcycle	68.8%	79.4%	86.7%
pole	48.2%	70.3%	68.6%
road	98.0%	98.0%	100.0%
traffic sign	74.0%	79.5%	93.1%
average	**68.7%**	**85.1%**	**80.1%**

Objects $> 32^2$

Human Annotation Flaws



Classification Flaws

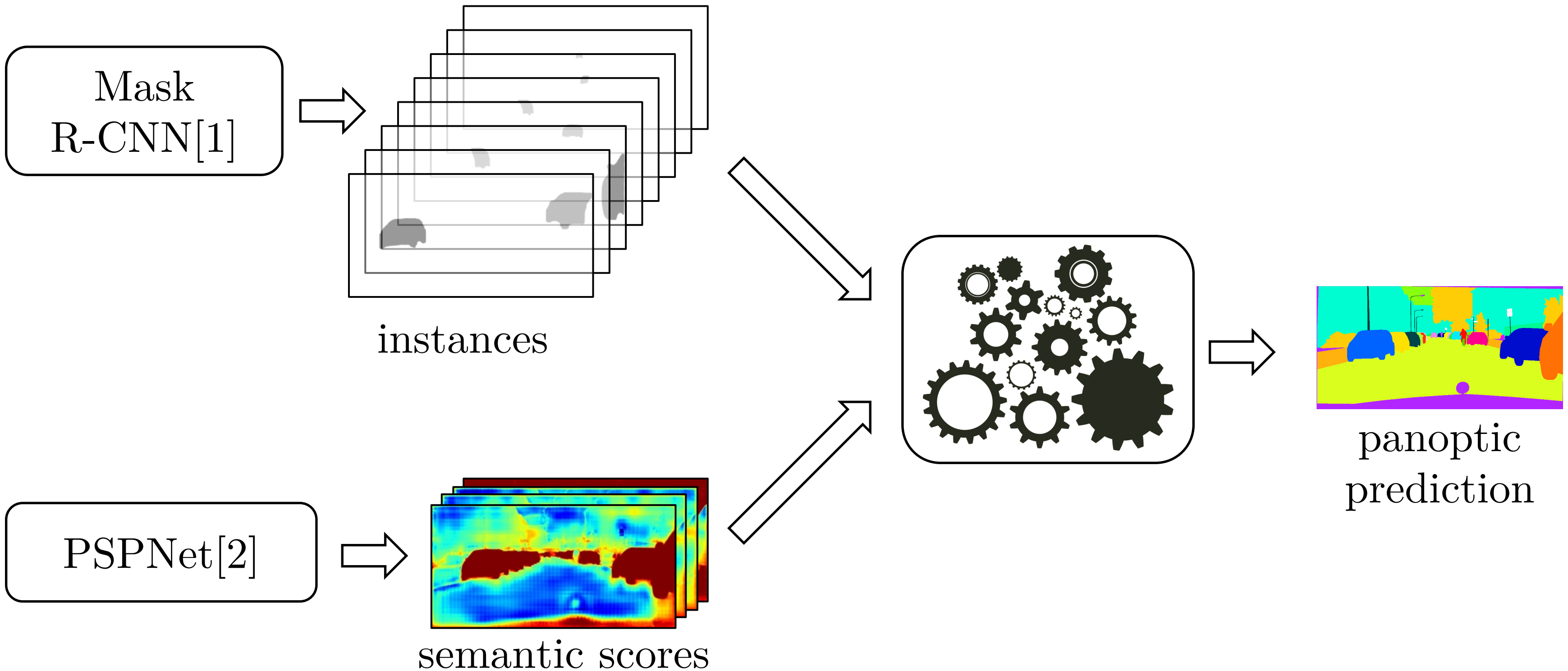
Human Annotation Flaws

Segmentation Flaws

Outline

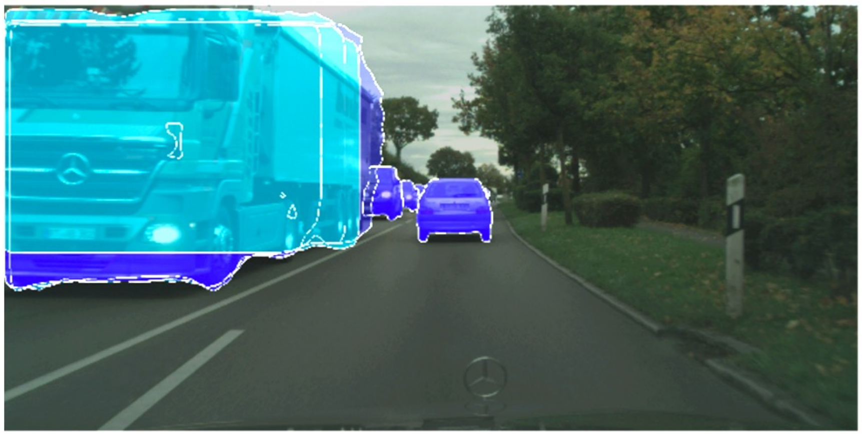
- Motivation
- Problem Definition
- Quality Evaluation
- Human Performance
- **Humans vs Computers**
- Perspectives

Mask R-CNN + PSPNet Combination Heuristic

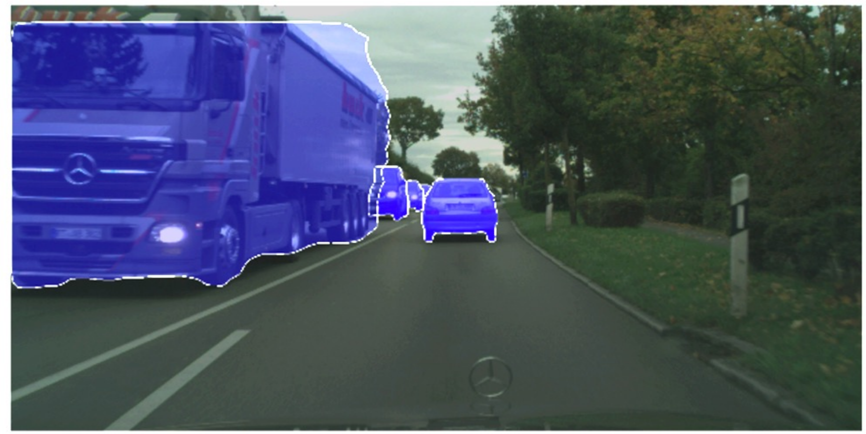


[1] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.
[2] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. Pyramid scene parsing network. CVPR 2017.

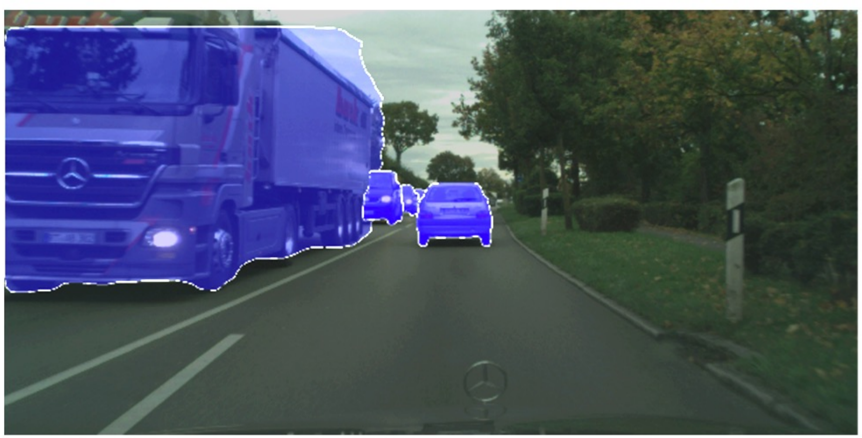
Mask R-CNN Non-overlapping Instances



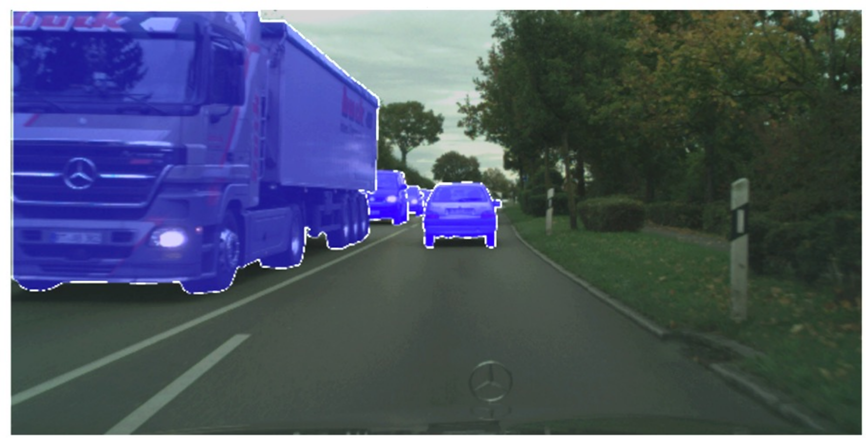
Mask R-CNN output



Mask R-CNN filtered



Non-overlapping Instances



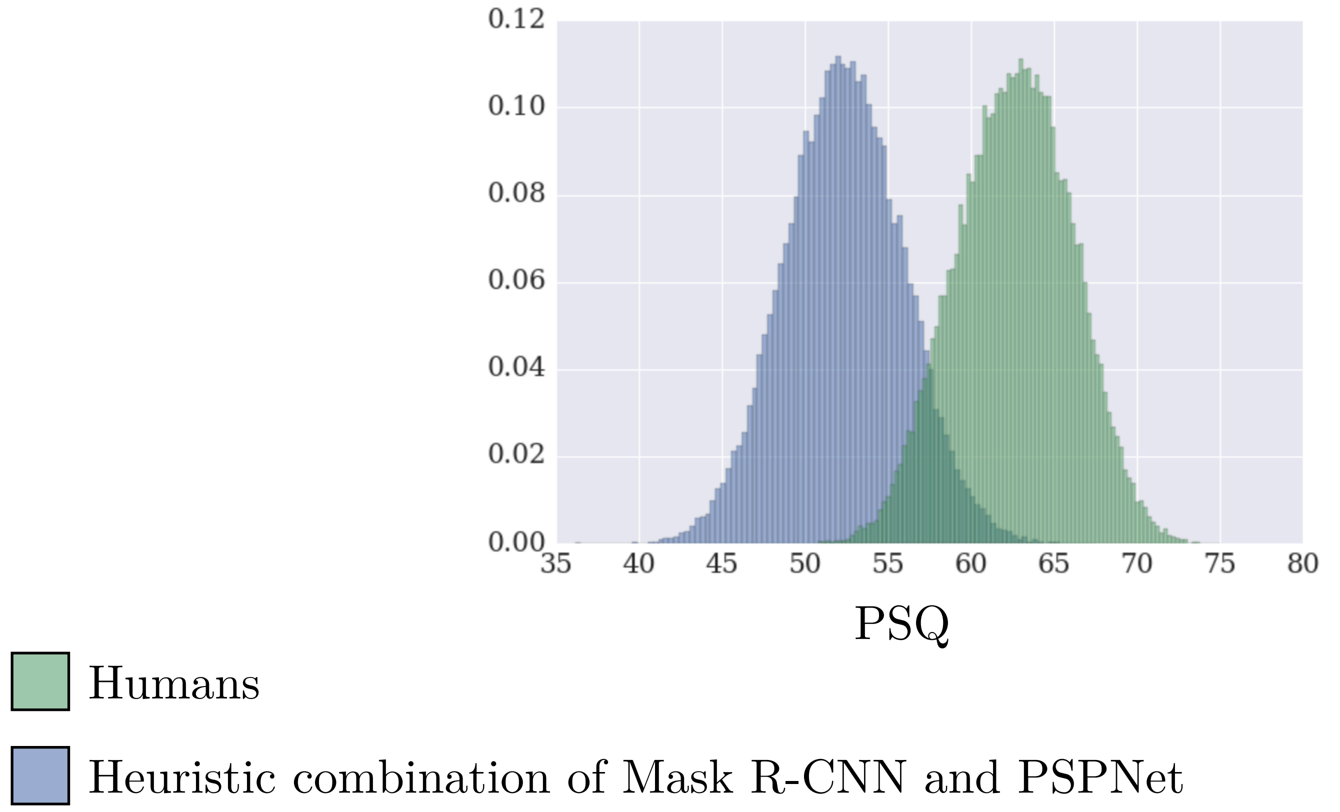
Ground Truth

PSQ – Humans vs Computers

	PSQ avg.	Seg Quality avg.	Det Quality avg.
Humans	**62.6%**	**83.9%**	**73.43%**
Mask R-CNN + PSPNet	**51.7%**	**81.0%**	**62.01%**

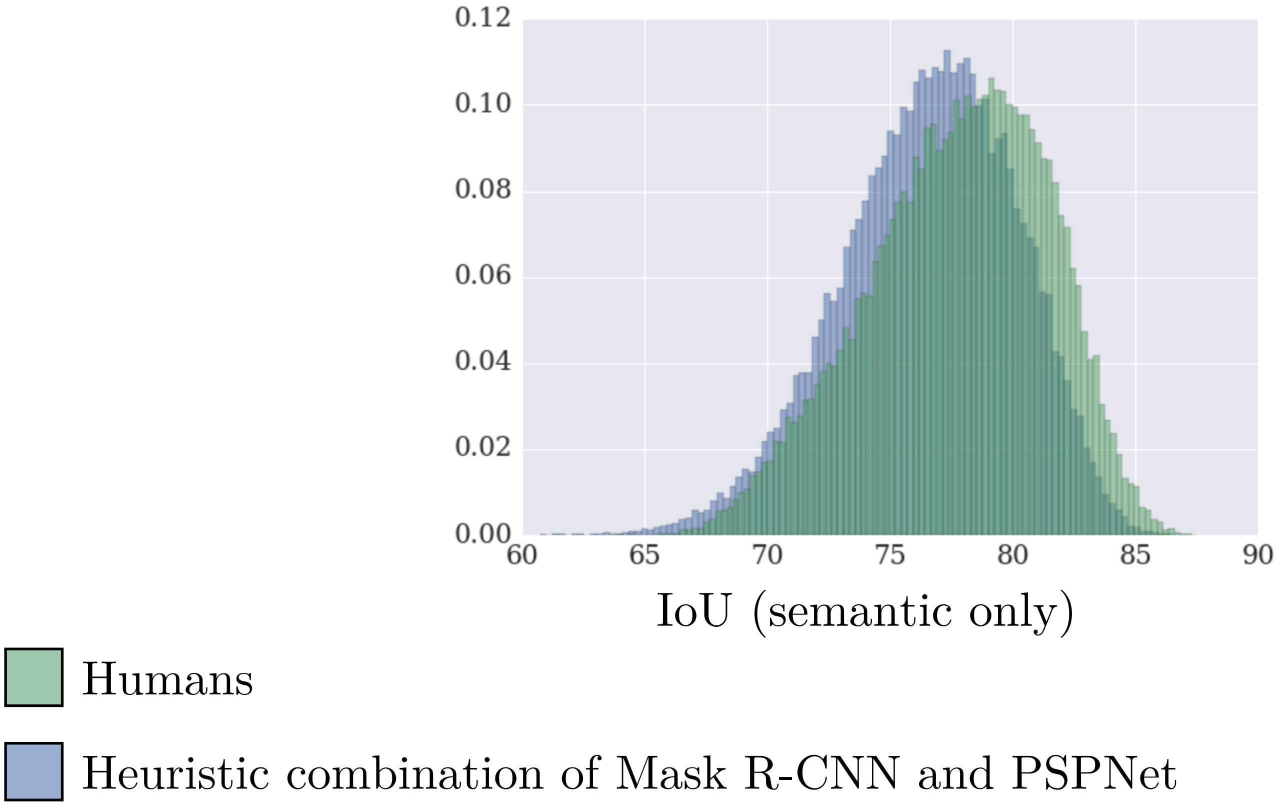
PSQ – Humans vs Computers

	PSQ avg.	Seg Quality avg.	Det Quality avg.
Humans	**62.6%**	**83.9%**	**73.43%**
Mask R-CNN + PSPNet	**51.7%**	**81.0%**	**62.01%**



PSQ – Humans vs Computers

	PSQ avg.	Seg Quality avg.	Det Quality avg.
Humans	**62.6%**	**83.9%**	**73.43%**
Mask R-CNN + PSPNet	**51.7%**	**81.0%**	**62.01%**



Outline

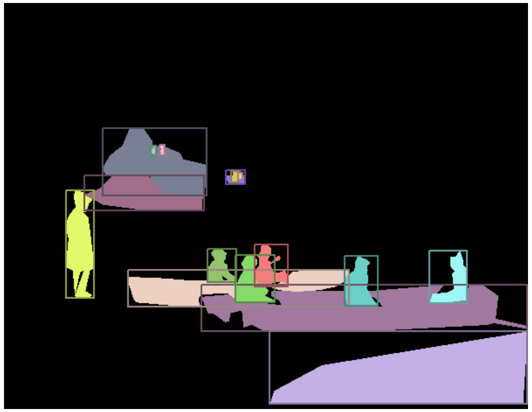
- Motivation
- Problem Definition
- Quality Evaluation
- Human Performance
- Humans vs Computers
- **Perspectives**

Why solve it?

Semantic Segmentation

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

Panoptic Segmentation



Object Detection/Seg

- each object detected and segmented separately
- "stuff" is not segmented

Why solve it?

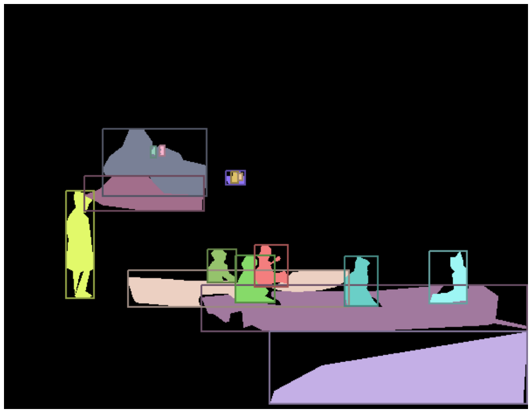
Semantic Segmentation

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

FCN 8s, Dilation8, DeepLab, PSPNet, RefineNet, U-Net, etc.

Panoptic Segmentation

?

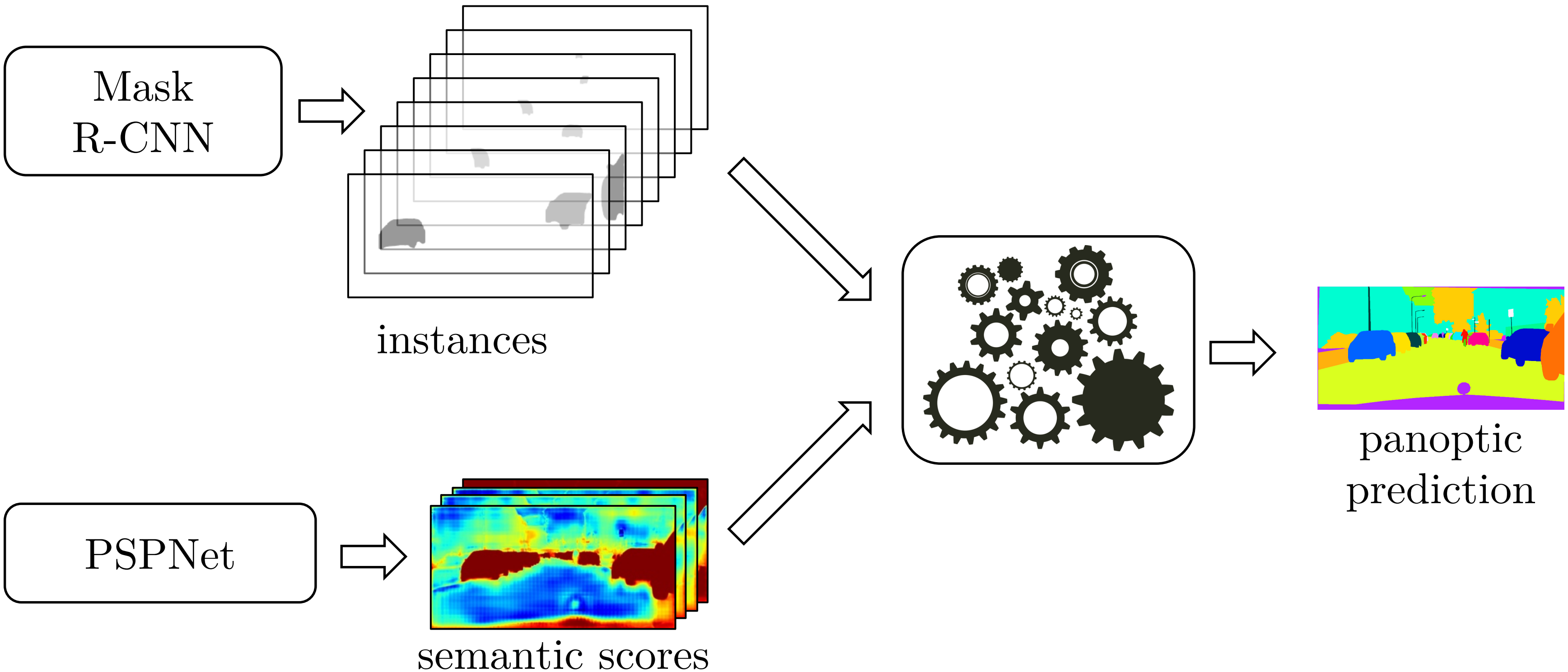


Object Detection/Seg

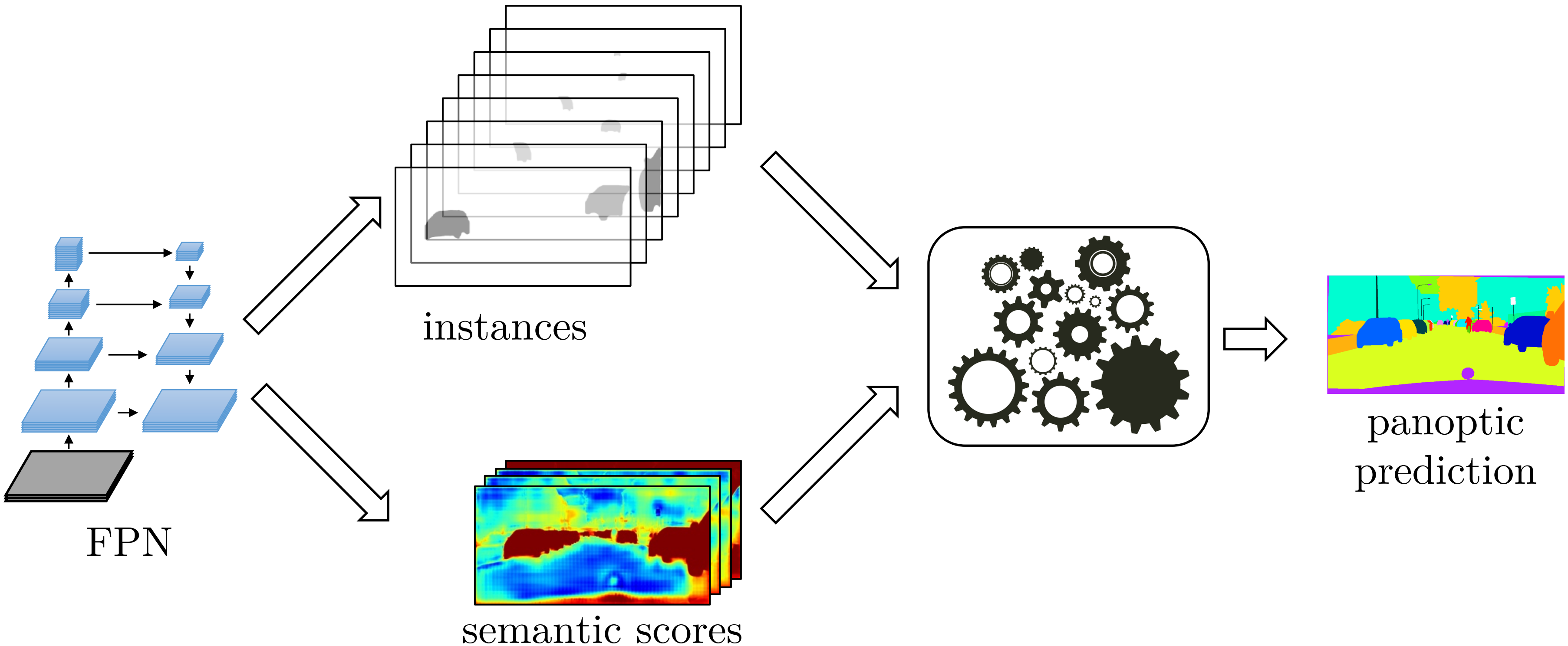
- each object detected and segmented separately
- "stuff" is not segmented

Fast/er R-CNN, DeepMask, SharpMask, Mask R-CNN, FCIS, YOLO, RetinaNet, FPN, etc.

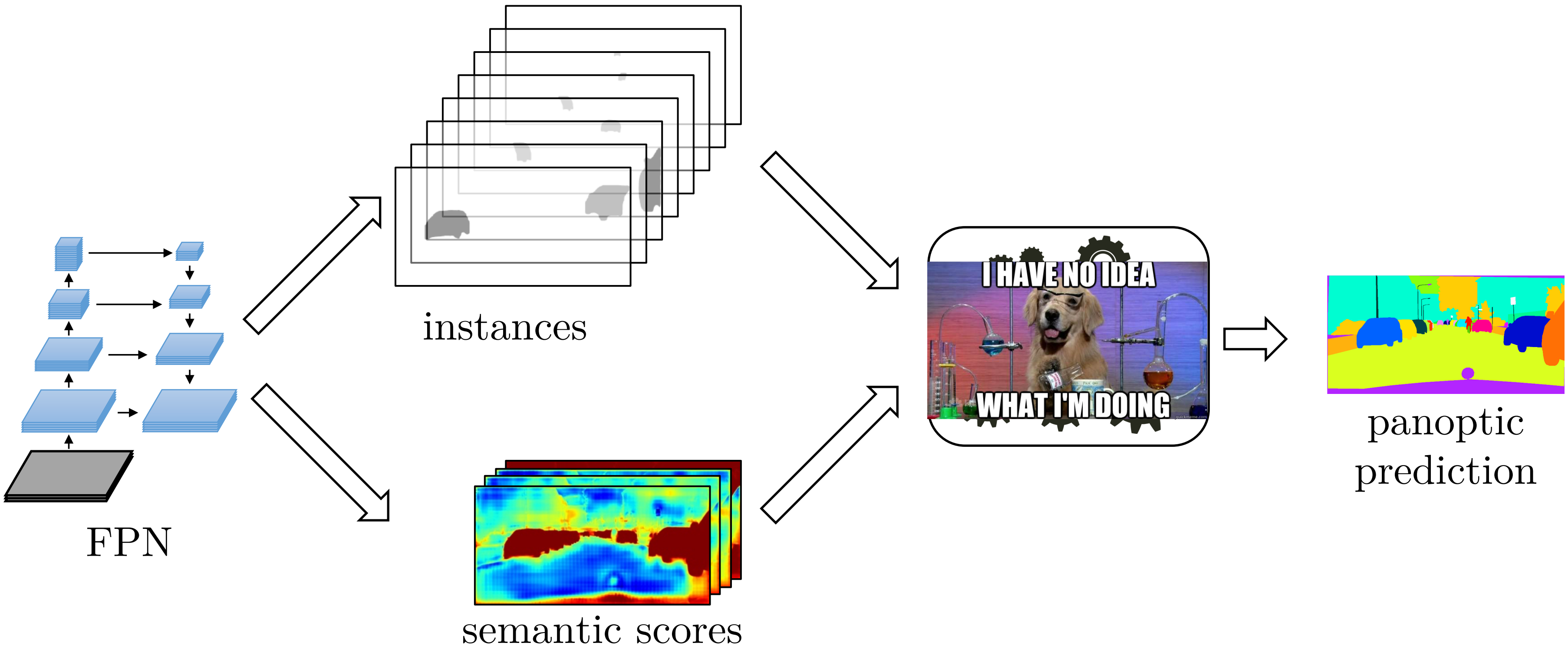
Why solve it?


Mask
R-CNN
instances
PSPNet
semantic scores
panoptic
prediction

Why solve it?



Why solve it?


instances
I HAVE NO IDEA
WHAT I'M DOING
FPN
semantic scores
panoptic prediction

Panoptic Segmentation: Future Plans

- Panoptic Segmentation paper on ArXiv
- Efficient evaluation code on GitHub
- Possible competition(s)

Panoptic COCO

Panoptic CityScapes