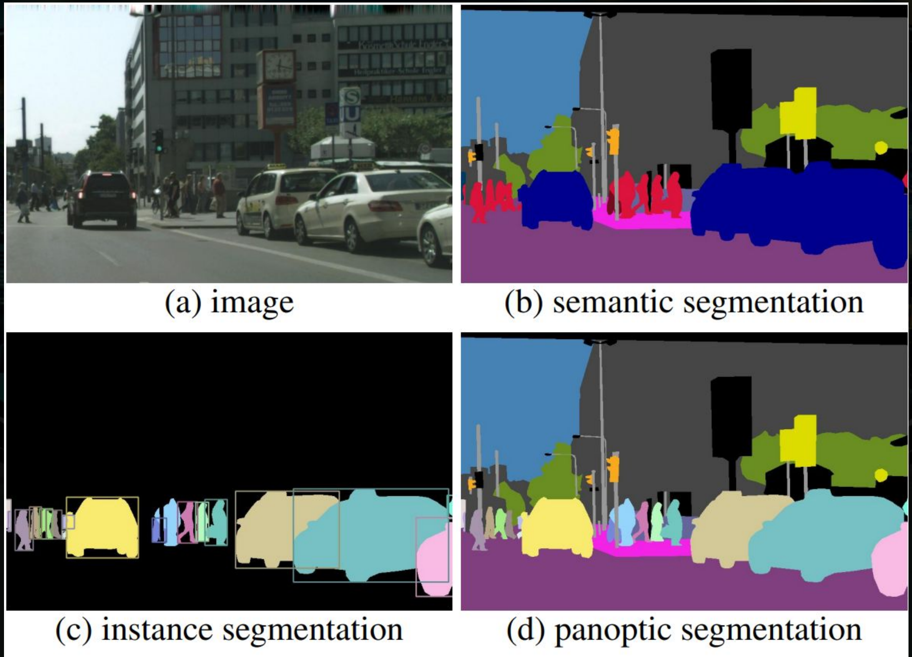
Panoptic FPN

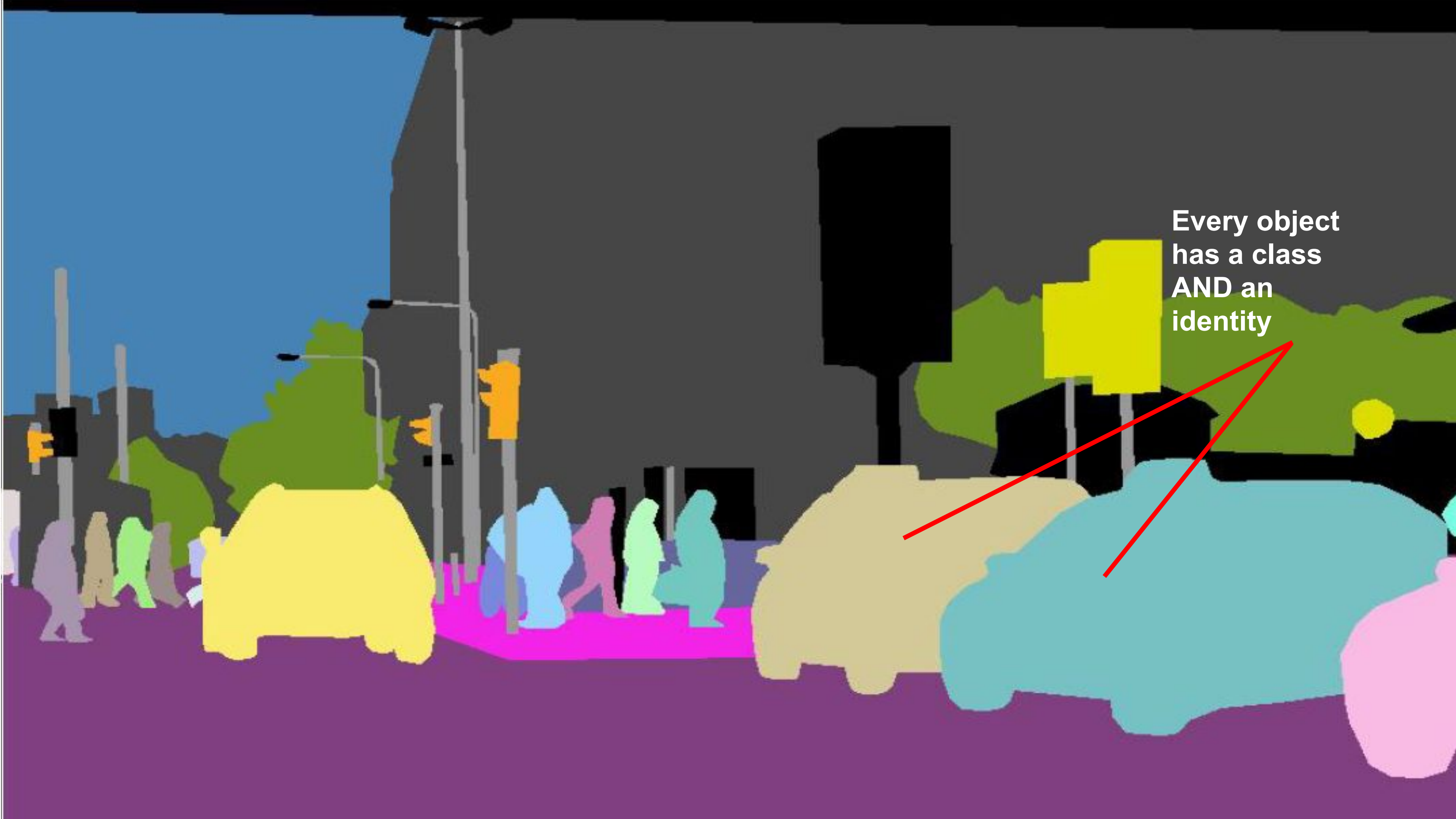
Presented By:
Audun Wigum Arbo, Even Dalen, Christian Echtermeyer

What is panoptic segmentation?

“including **everything** visible in one view”



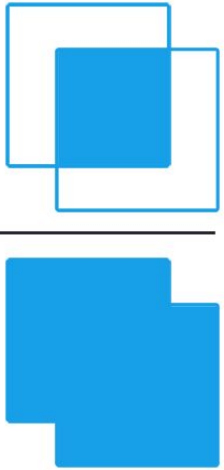
(a) image
(b) semantic segmentation
(c) instance segmentation
(d) panoptic segmentation


Every object
has a class
AND an
identity

Panoptic Segmentation Metric

"...existing metrics are specialized for **either** semantic or instance segmentation and **cannot** be used to evaluate the joint task involving both..."

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



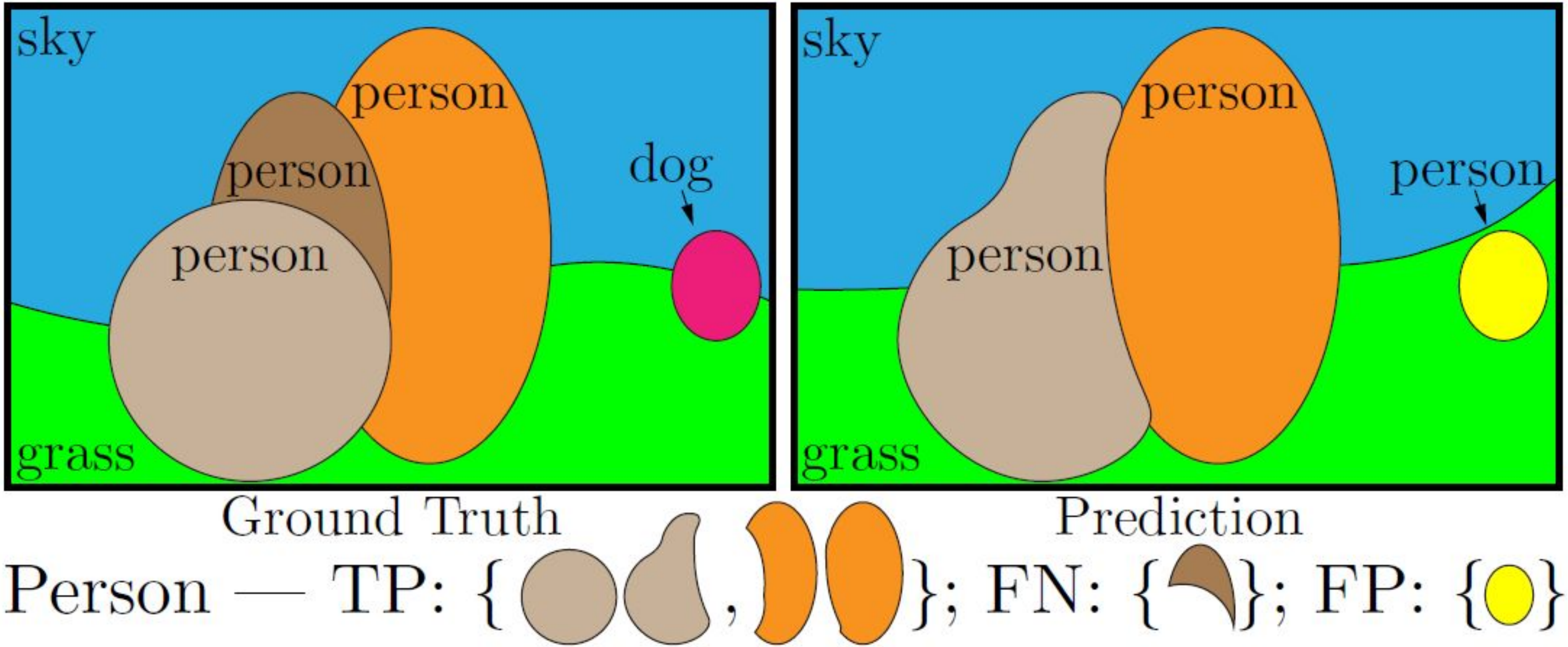
		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP	FN
	NEGATIVE (0)	FP	TN

IoU > 0.5 = Match

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

p = predicted segment

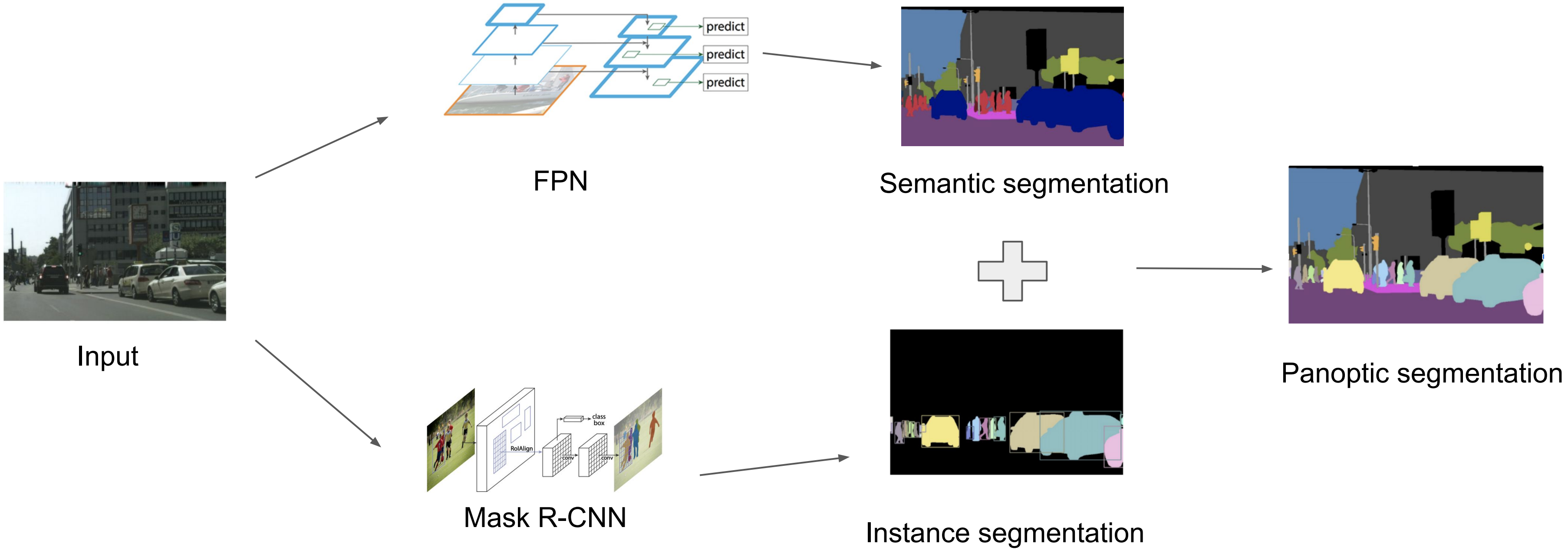
g = ground truth segment


sky
person
person
person
dog
grass
Ground Truth
sky
person
person
person
grass
Prediction
Person — TP: { , }; FN: { }; FP: { }

Contribution of Panoptic FPN

- Previous approaches used separate networks:
  - One for instance segmentation
  - One for semantic segmentation.
- Panoptic FPN uses a single network.
- Increased efficiency and memory footprint.
- Established a baseline performance for future work.

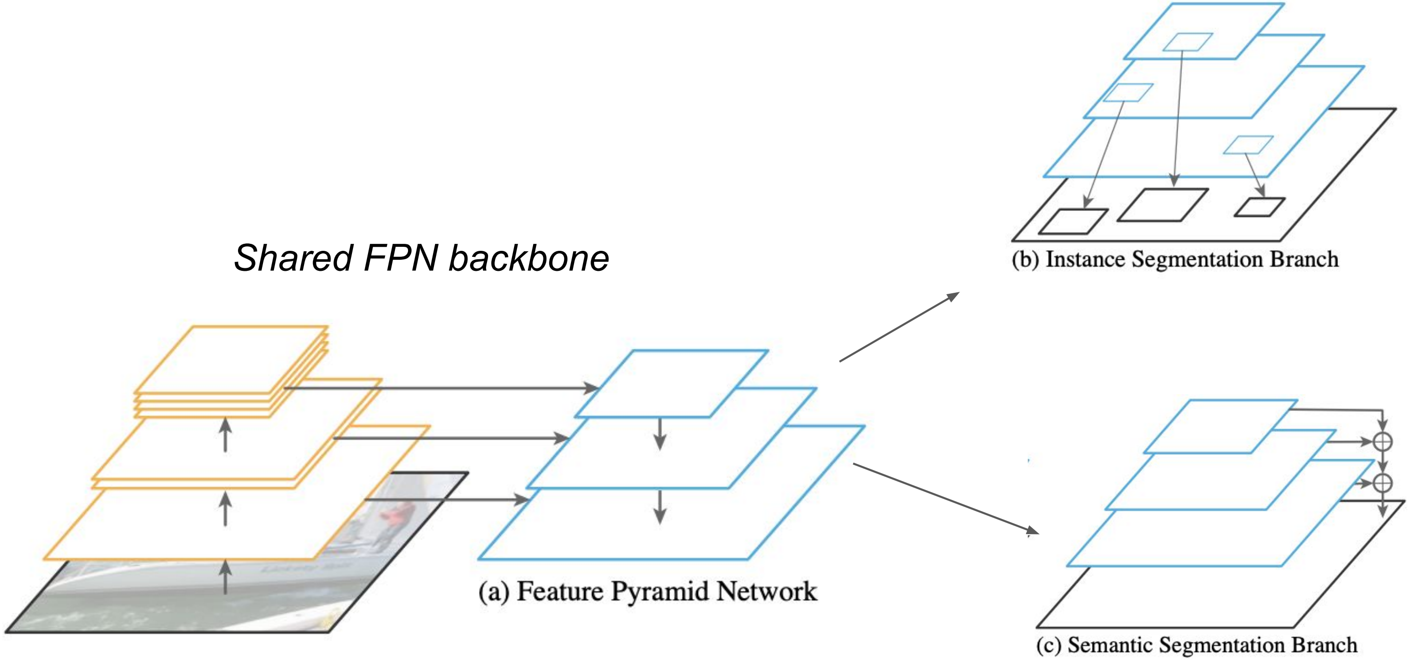
Inefficient approach: Semantic + instance = Panoptic



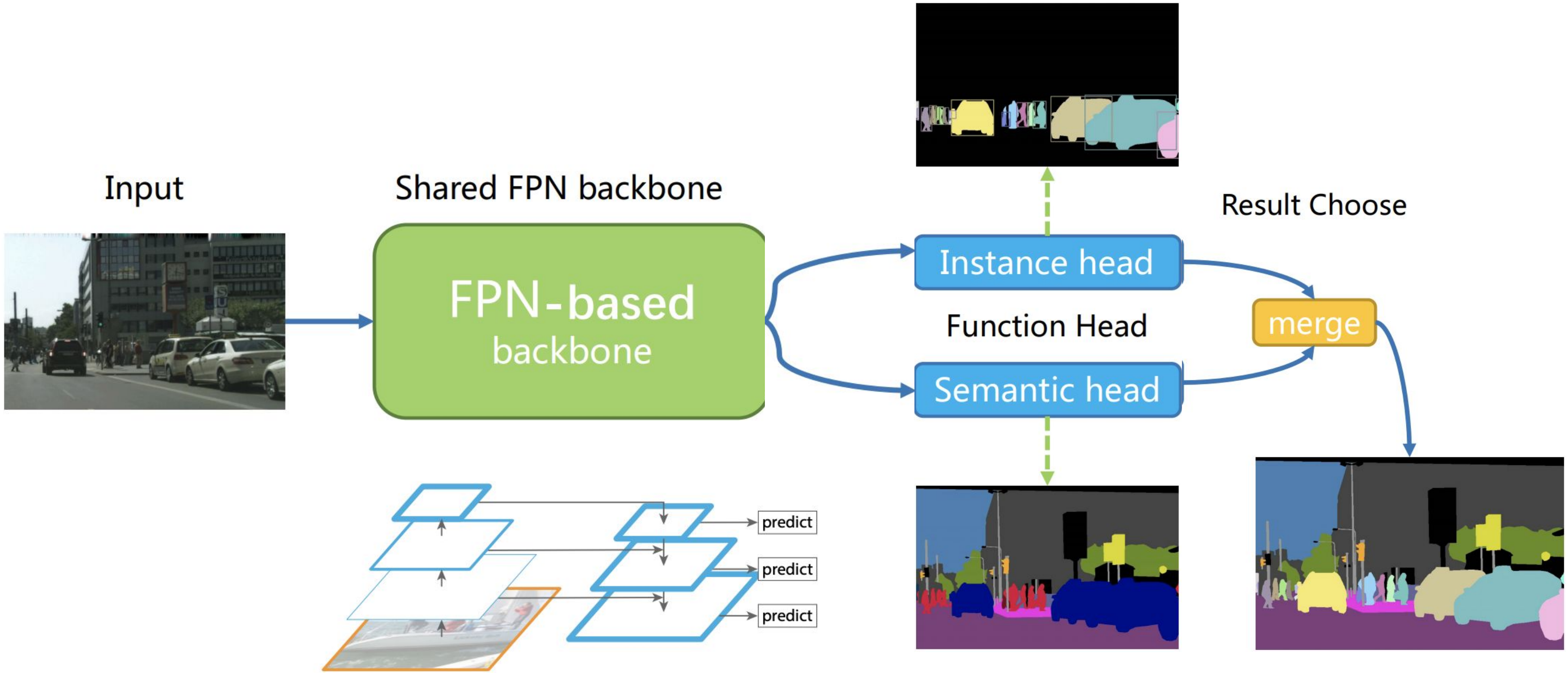
Panoptic FPN architecture

Based on Mask R-CNN head

Shared FPN backbone

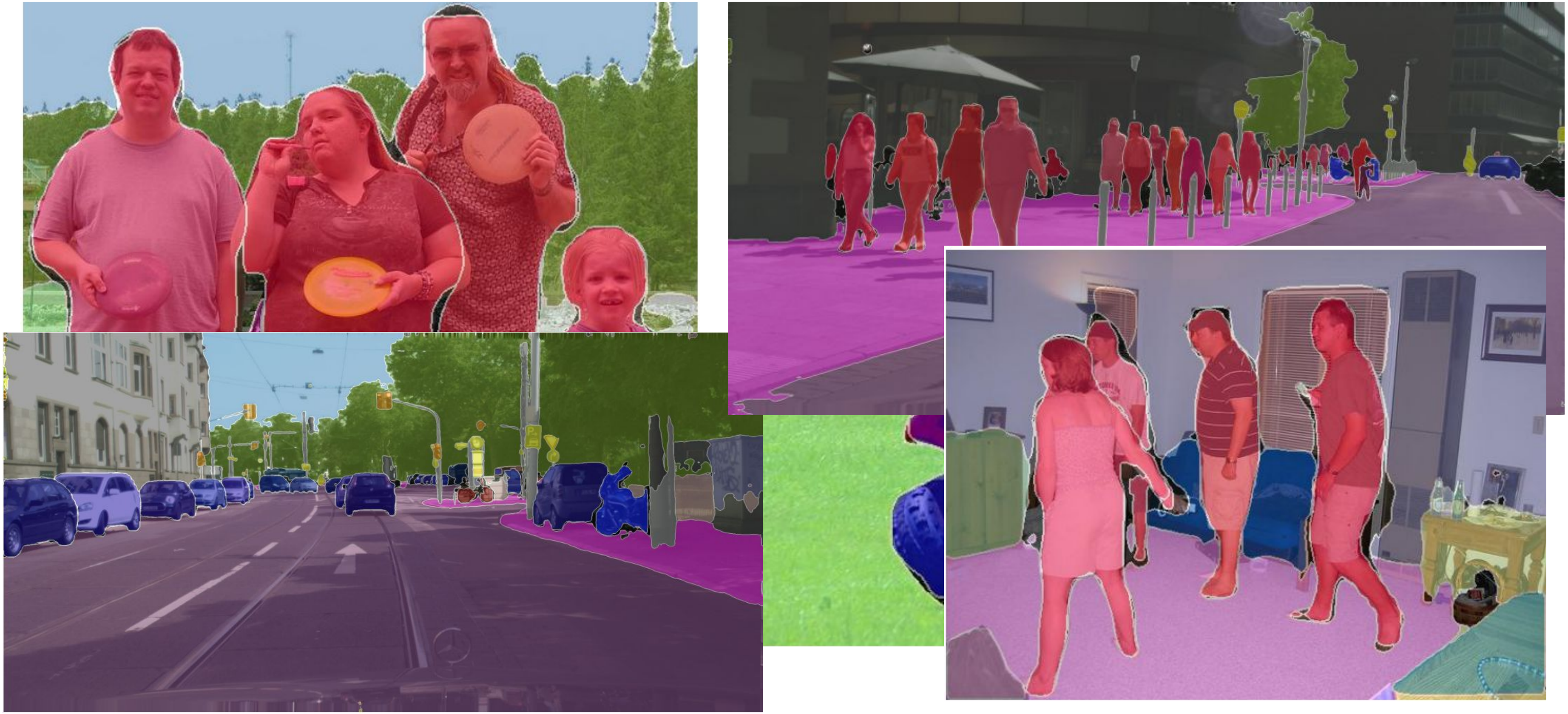


Scale and sum feature maps.



Source: http://presentations.cocodataset.org/ECCV18/COCO18-Panoptic-Caribbean.pdf

Results: some example output



Results: the metric values

Loss: $\lambda_s Loss_S + \lambda_i Loss_i$
Scaling of the individual losses:

- Semantic seg. loss with λ_s
- Instance seg. loss with λ_i

The following results use the optimal λ_s and λ_i from the set $\{0.5, 0.75, 1.0\}$

Results: Two FPN networks VS combined

	backbone	AP	PQ^{Th}	mIoU	PQ^{St}	PQ
COCO	R50-FPN×2	33.9	46.6	40.2	27.9	39.2
	R101-FPN	35.2	47.5	42.1	29.5	40.3
		+1.3	*+0.9*	*+1.9*	*+1.6*	*+1.1*
Cityscapes	R50-FPN×2	32.2	51.3	74.5	62.4	57.7
	R101-FPN	33.0	52.0	75.7	62.5	58.1
		+0.8	*+0.7*	*+1.3*	*+0.1*	*+0.4*

(b) **Panoptic Segmentation: Panoptic *R101-FPN* *vs.* R50-FPN×2.** *Given a roughly* equal *computational budget, a single FPN network for the panoptic task outperforms two independent FPN networks for instance and semantic segmentation by a healthy margin.*

Results: COCO + Cityscapes panoptic (as of 2018)*

	coarse	PQ	PQ^{Th}	PQ^{St}	mIoU	AP
DIN [1, 34]	✓	53.8	42.5	62.1	**80.1**	28.6
Panoptic FPN		**58.1**	**52.0**	**62.5**	75.7	**33.0**

(b) **Panoptic Segmentation on Cityscapes.** For Cityscapes, there is no public leaderboard for panoptic segmentation at this time.

see http://cocodataset.org/#panoptic-leaderboard).

*Single-network entries only

Summarized: Main contributions of Panoptic FPN

- A single network:
  - A common Feature Pyramid Network (FPN)
  - Mask R-CNN (instance segmentation)
    + branch with semantic segmentation
- State-of-the-art performance in both instance- and semantic segmentation, with **only ~0.5x computing resources** compared to multi-network
- Outperforms all *single-model* entries in the 2018 COCO Panoptic Segmentation Challenge
- <u>A good baseline for the panoptic segmentation task</u>

References

- Panoptic Feature Pyramid Networks
- CVPR 2019 Oral Session 2-2A: Recognition
- Feature Pyramid Networks for Object Detection
- Panoptic Segmentation