Photogrammetry & Robotics Lab

Machine Learning for Robotics and Computer Vision Tutorial

Segmentation with CNNs

Jens Behley

Exam Dates

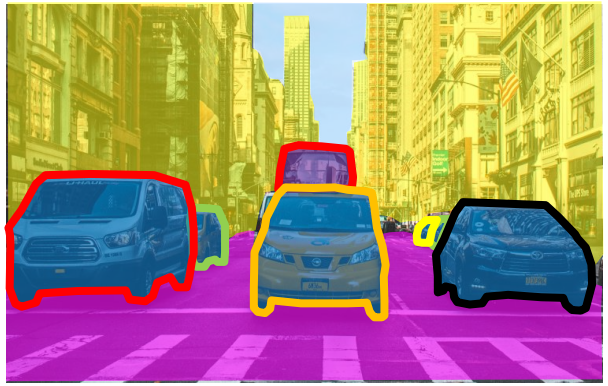
- Oral Exam via Zoom in English
- Webcam must be on all the time and alone in room
- No other windows besides Zoom open.

- Possible dates (we make also a Doodle):
  1. Wed, 11.08.2021
  2. Wed, 25.08.2021
  3. Wed, 01.09.2021

- Are there any hard constraints that make one of the dates unsuitable?

This week's lecture





- Fine-grained scene understanding:
  - Semantic Segmentation
  - Instance Segmentation
  - Panoptic Segmentation
- Discussed common, popular approaches for segmentation in these domains.

Semantic Segmentation

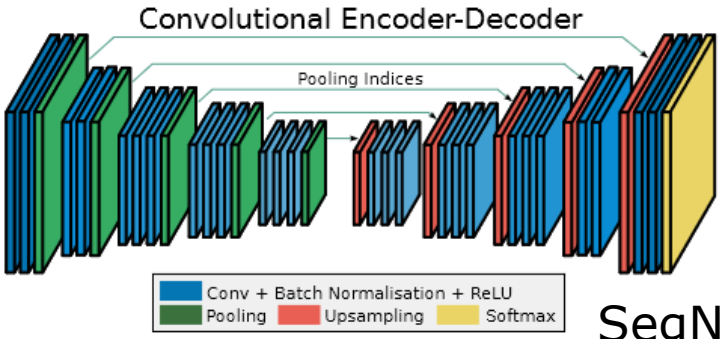
CNN

Building Road Car Bus

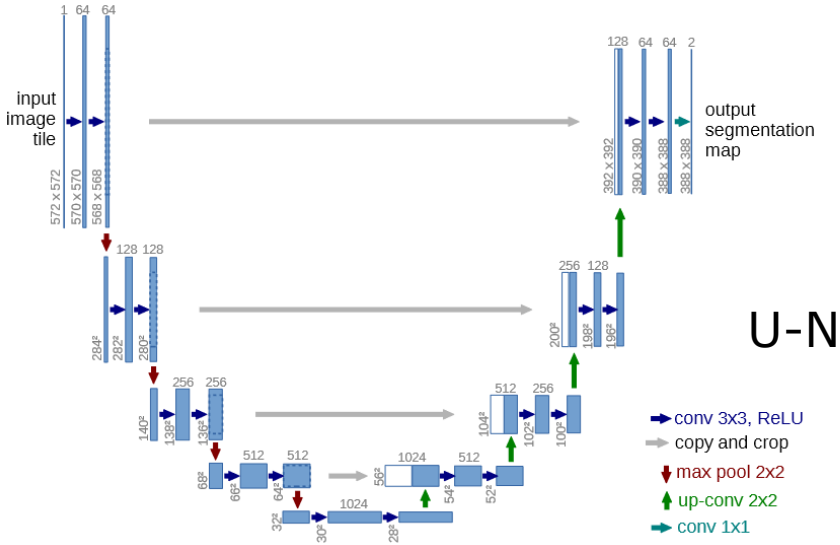
- **Goal:** Provide label $y_{i,j} \in \{1, \ldots, K\}$ for each pixel in the image.

Photo by Marisol Benitez on Unsplash

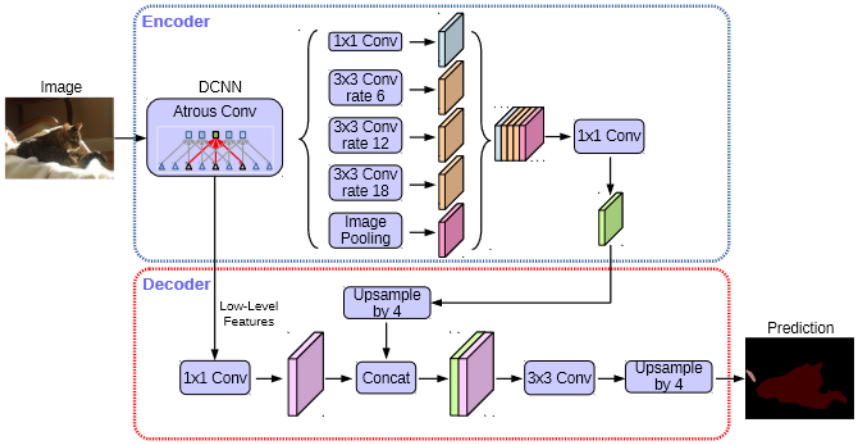
Common Strategies



SegNet

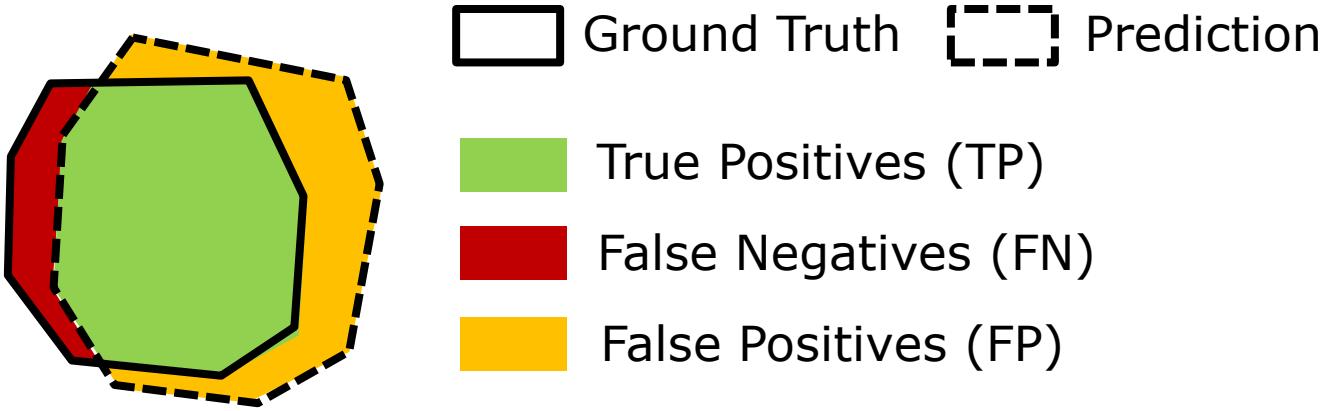


U-Net



- Main paradigm for semantic segmentation: Encoder-Decoder architecture
- Encoder: Capture context, Decoder: Upsampling

Evaluation Metric: mIoU



$$\mathrm{IoU}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c + \mathrm{FN}_c} \qquad \mathrm{mIoU} = \frac{1}{K} \sum_c \mathrm{IoU}_c$$

- For each class, determine pixe-wise intersection-over-union (IoU)
- Mean over class-wise IoUs gives mean Intersection-over-union **(mIoU)**

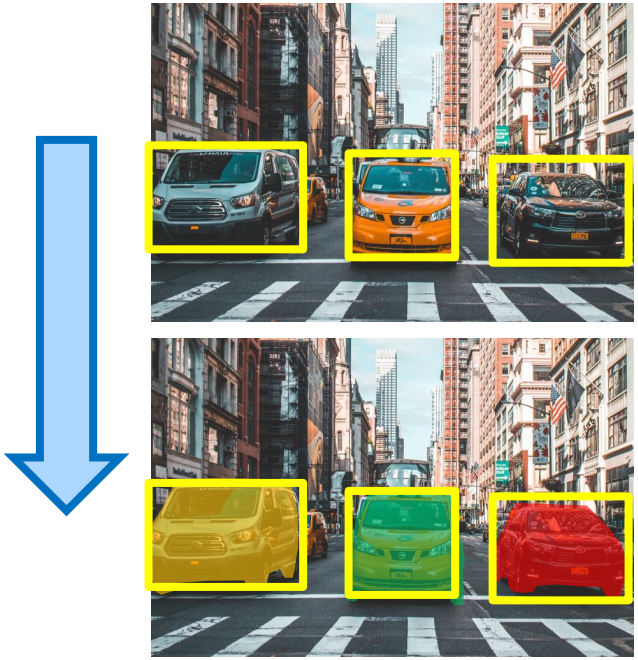
Instance Segmentation

Semantic Segmentation

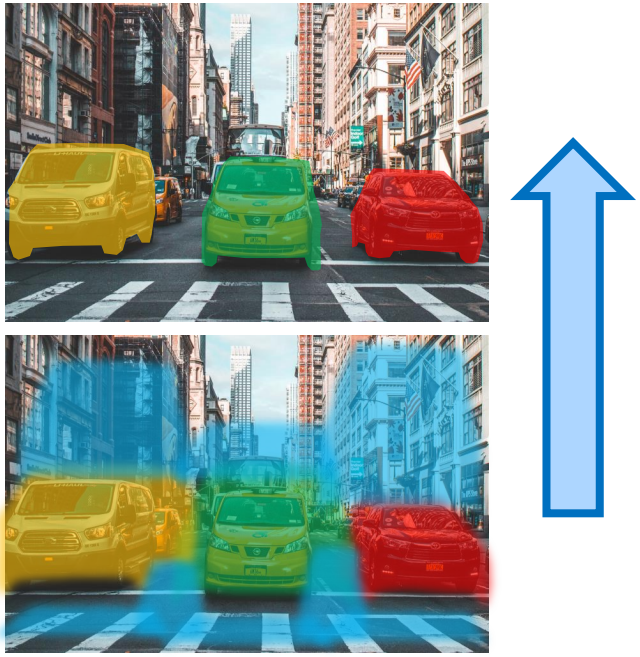
Instance Segmentation

- Semantic segmentation provides only class labels
- Instance segmentation aims at distinguishing different instances of the same object class

Top-down vs. Bottom-up Instance Segmentation



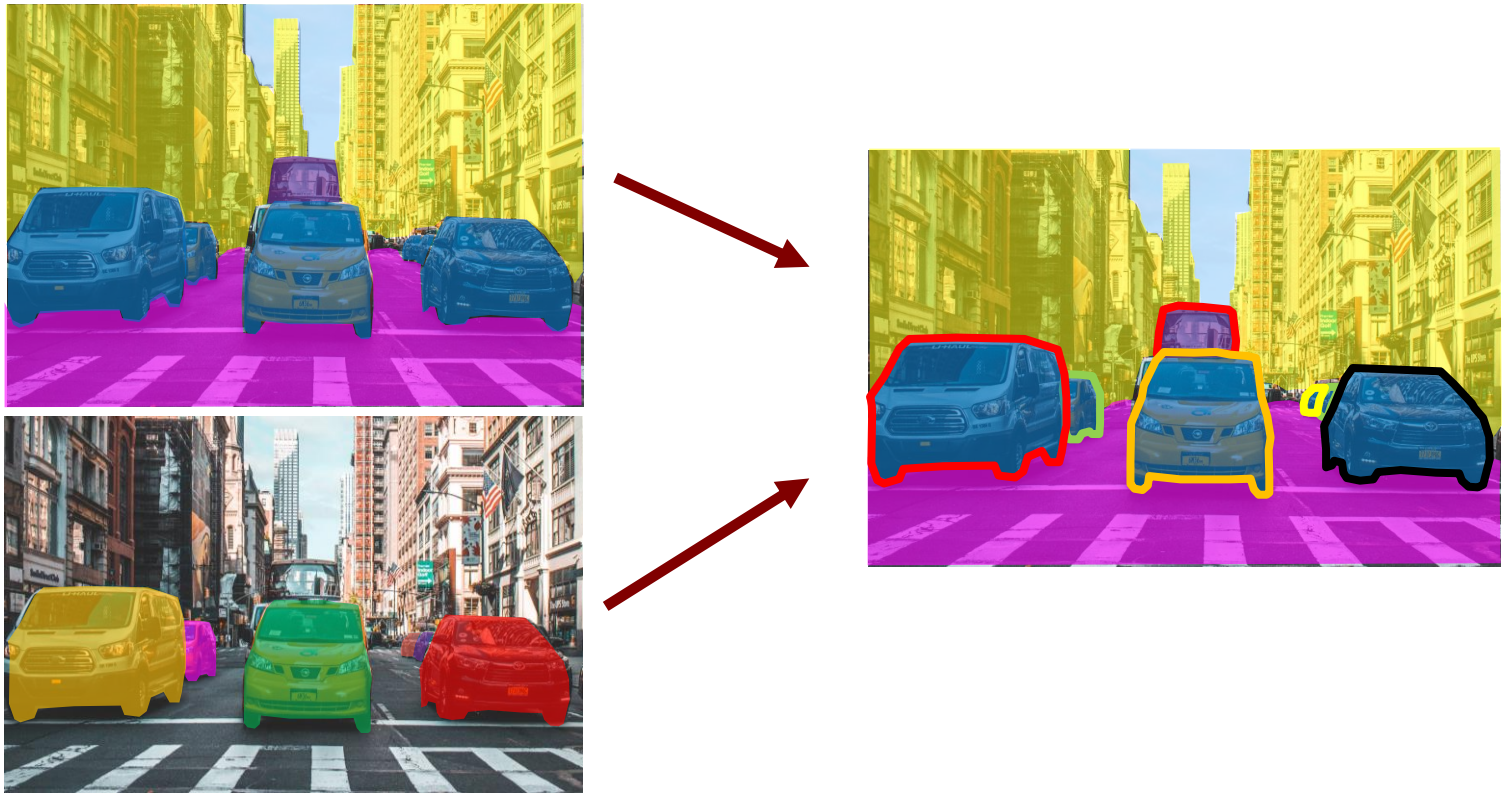
Top-down Approach
(**Mask R-CNN**)



Bottom-up Approach
(**Embeddings**)

- **Top-down**: Instances are first determined and then foreground/background mask estimated
- **Bottom-up**: Determine per-pixel properties that are then used to cluster instances

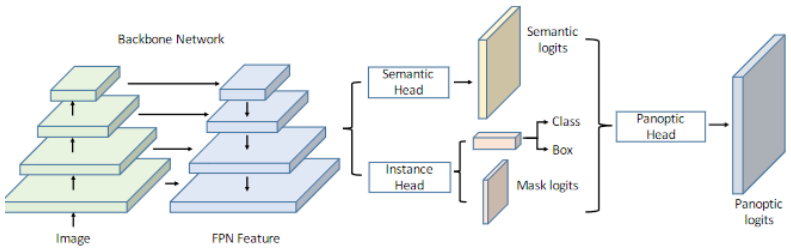
Panoptic Segmentation



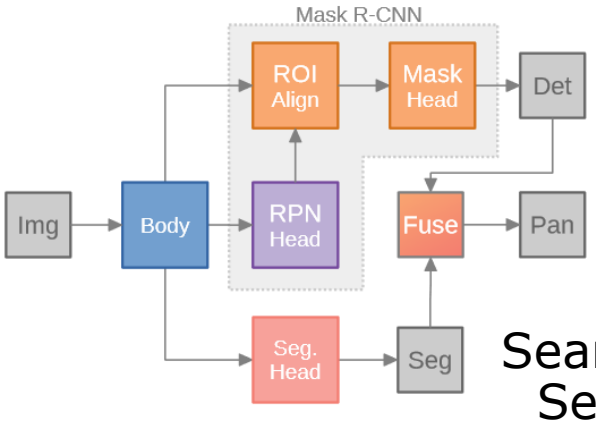
- Panoptic Segmentation unifies semantic and instance segmentation
- Distinguish **stuff** (e.g., vegetation, road, …) and **thing** classes (e.g., car, pedestrian, …)

[Kirillov, 2019]

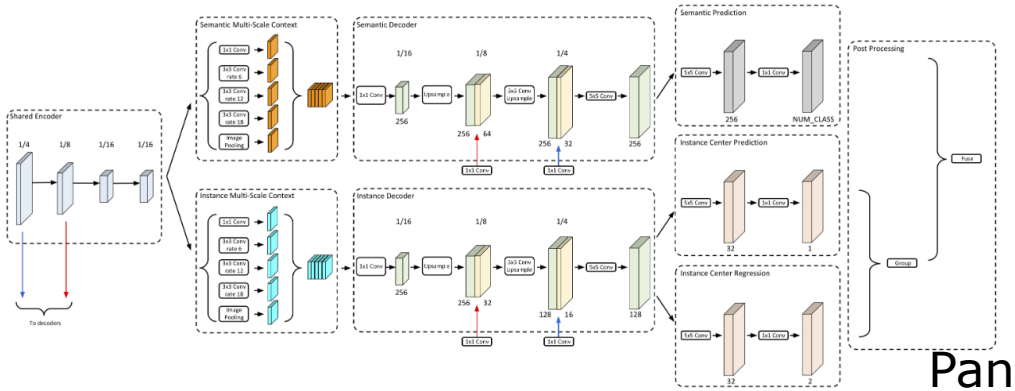
Panoptic Segmentation Approaches



UPSNet



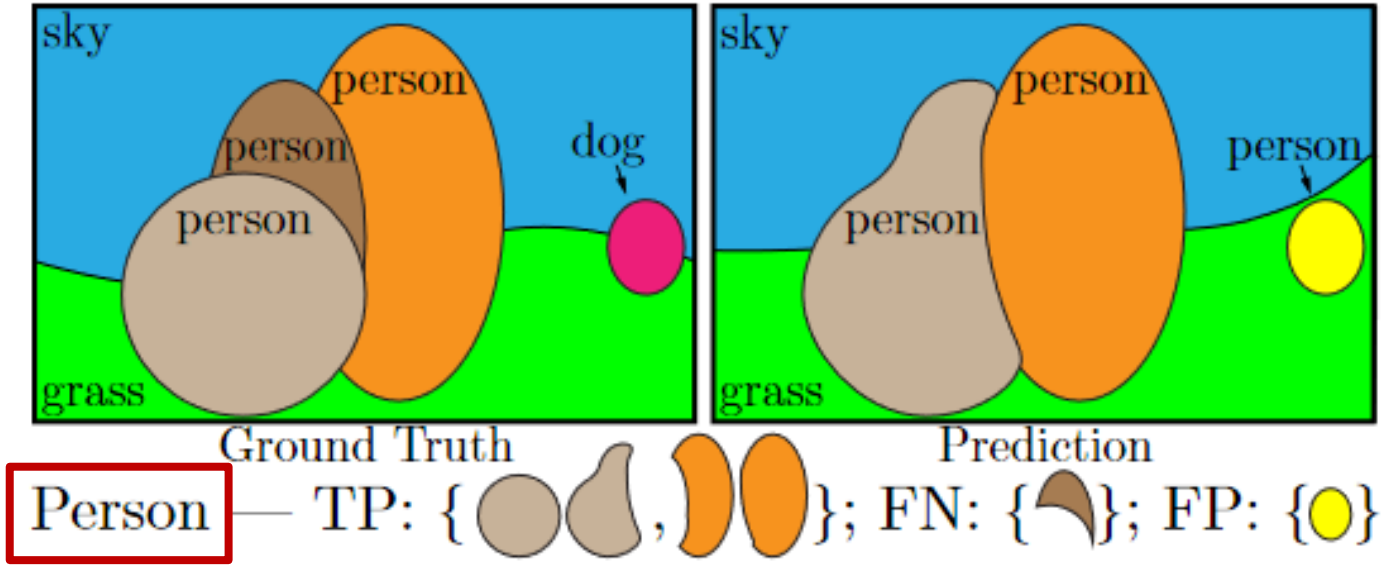
Seamless Scene Segmentation



Panoptic-DeepLab

- Shared Encoder-Separate Decoder Architectures
- 1 decoder produces semantics, the other instances

Metric: Panoptic Quality



- Class-wise PQ is defined as:

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

- Overall panoptic quality (PQ)

$$PQ = \frac{1}{K}\sum_c \mathrm{PQ}_c$$

[Kirillov, 2019]

Figure from [Kirillov, 2019]

Pytorch Vision Models

- A good start for baselines already included in Pytorch (see `torchvision.models`):
  - **Detection:** Faster-RCNN, RetineNet, SSD
  - **Semantic Segmentation:** FCN, DeepLabV3, LR-ASPP with different backbones: ResNet-50, ResNet-101, MobileNetV3
  - **Instance Segmentation:** Mask R-CNN

Other Resources

- Detection
  - https://github.com/facebookresearch/detectron2
  - https://github.com/ultralytics/yolov5
- Semantic Segmentation
  - https://github.com/open-mmlab/mmsegmentation
- Panoptic-Deeplab
  - https://github.com/facebookresearch/detectron2/tree/master/projects/Panoptic-DeepLab

See you next week!