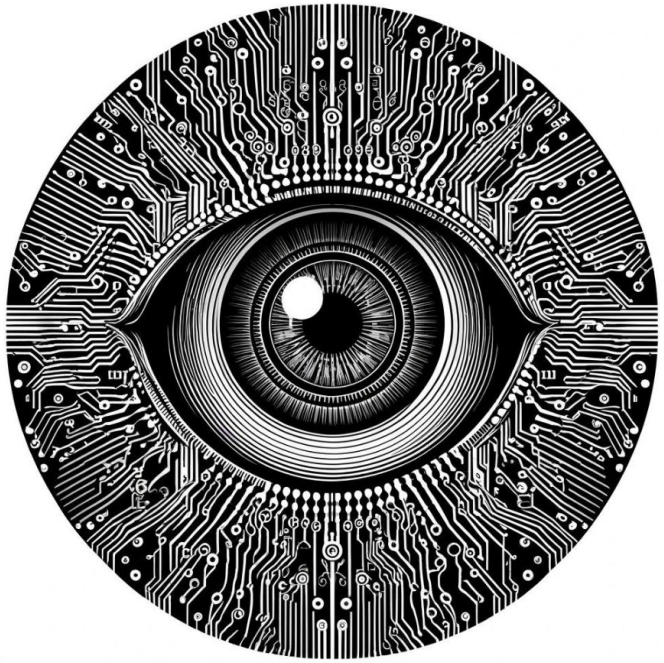
KI Service Zentrum by Hasso-Plattner-Institut HPI

Deep Learning Approaches to Image Segmentation



Antonio Rueda-Toicen

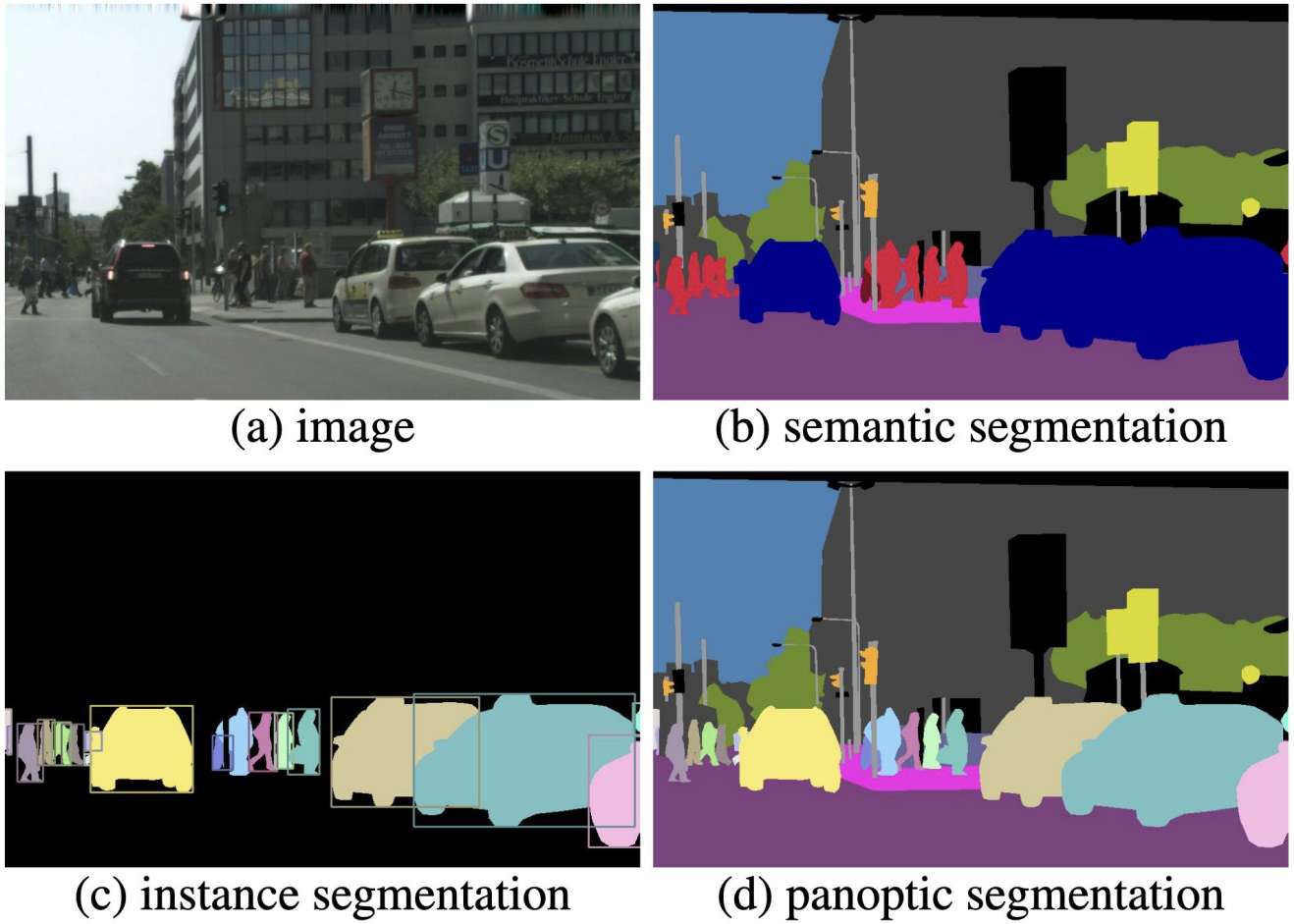
SPONSORED BY THE

Federal Ministry of Education and Research

Learning goals

- Understand the deep learning solutions for labeled image segmentation: semantic, instance, panoptic
- Describe class-agnostic and zero-shot segmentation with Segment Anything (SAM)

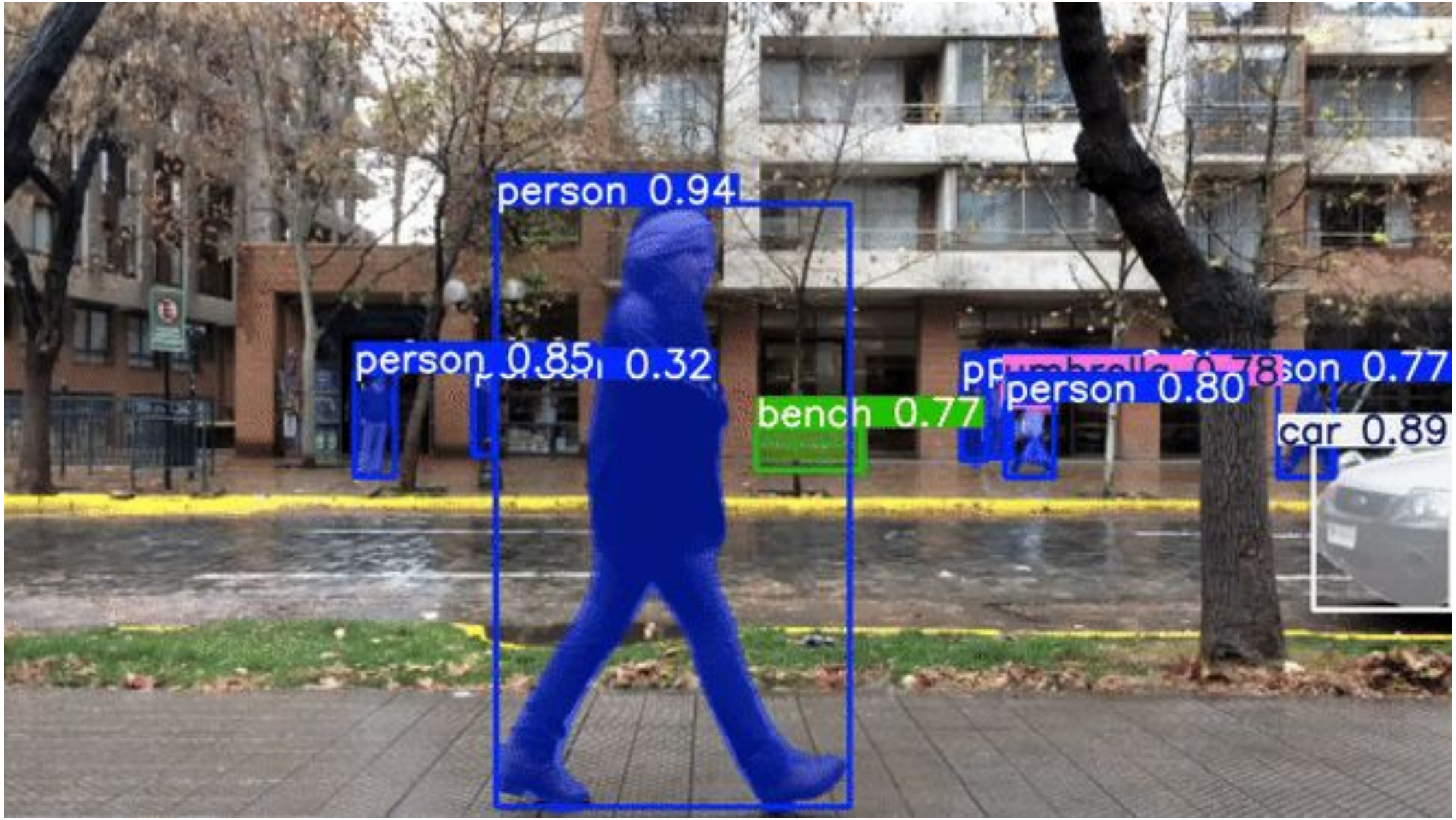
Semantic, Instance, and Panoptic Segmentation



(a) image (b) semantic segmentation

(c) instance segmentation (d) panoptic segmentation

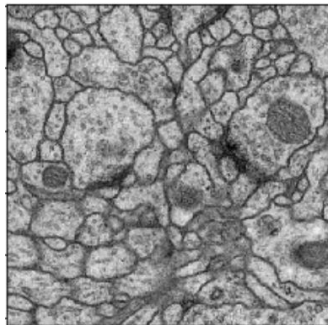
Image from [Panoptic Segmentation]

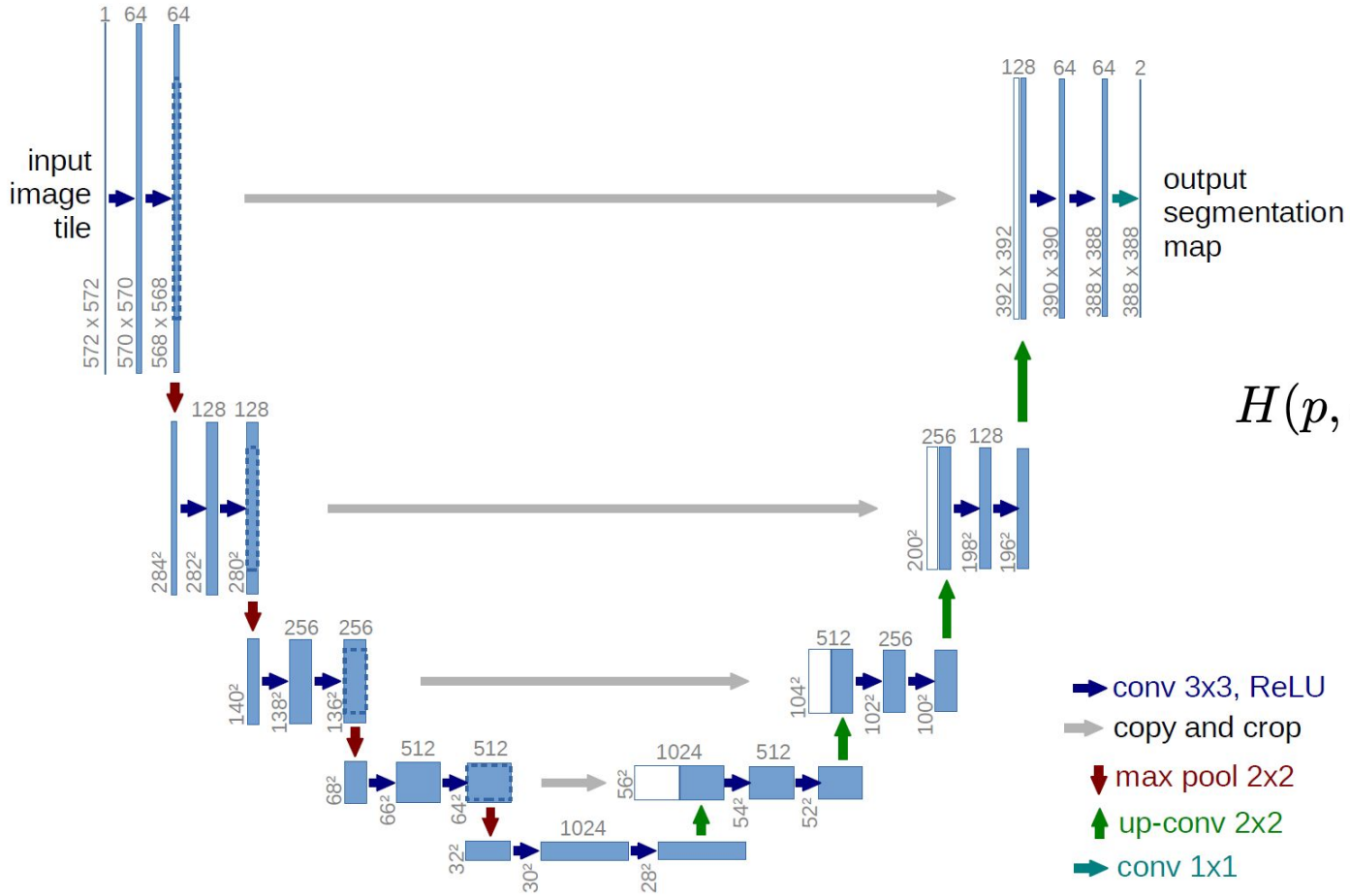


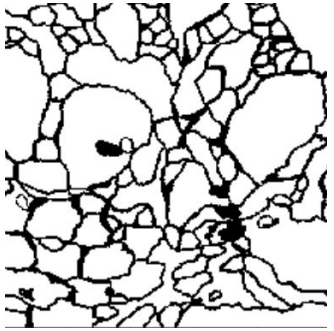
Example of instance segmentation with YOLO11

Image from https://learnopencv.com/yolo11/

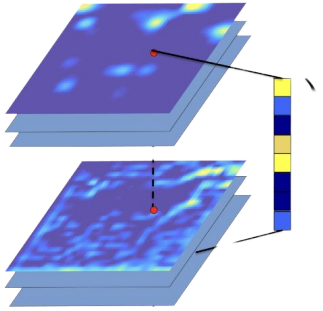
Semantic segmentation with U-Net



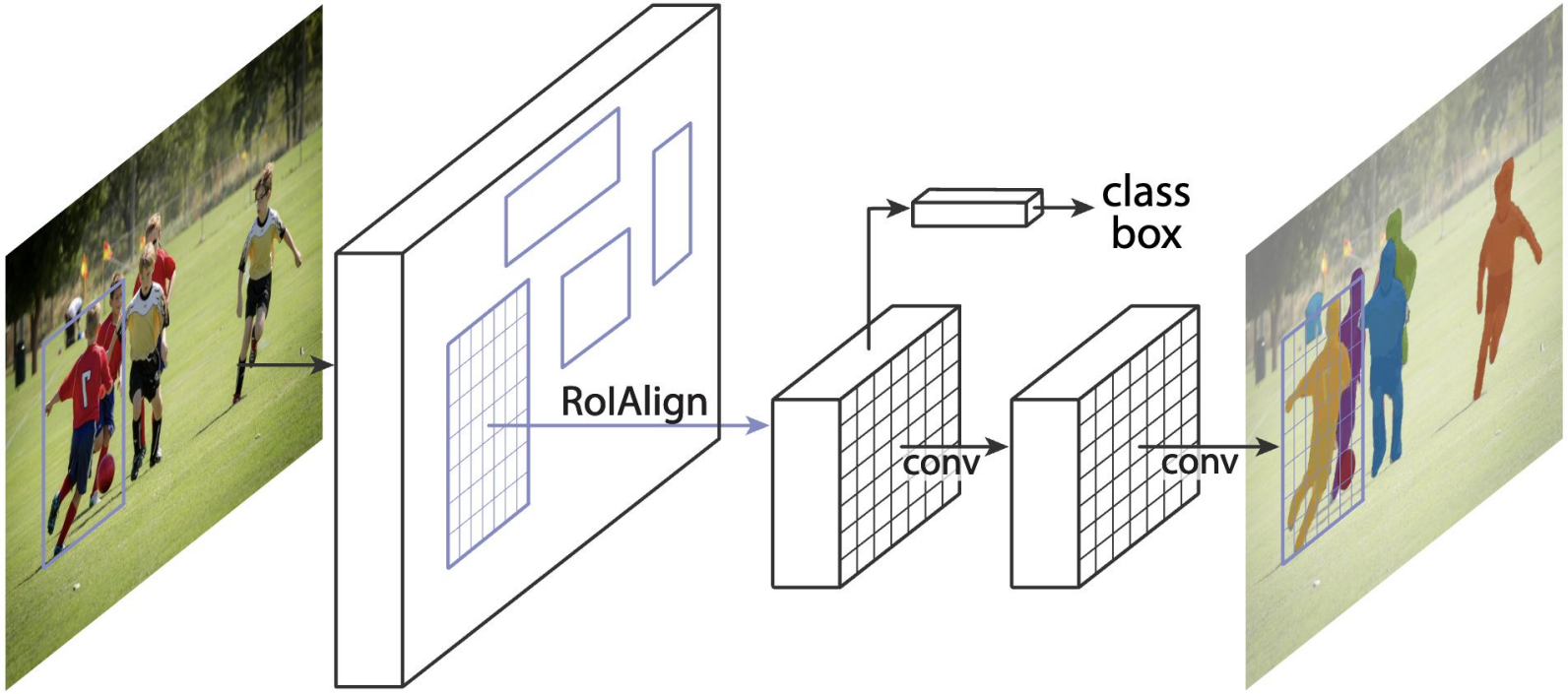




$$H(p,q) = -\sum_i p(i) \log q(i)$$



Instance segmentation with Mask R-CNN



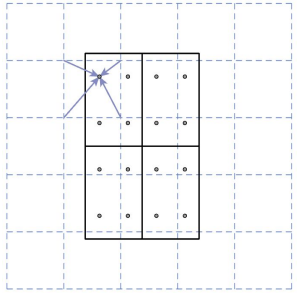
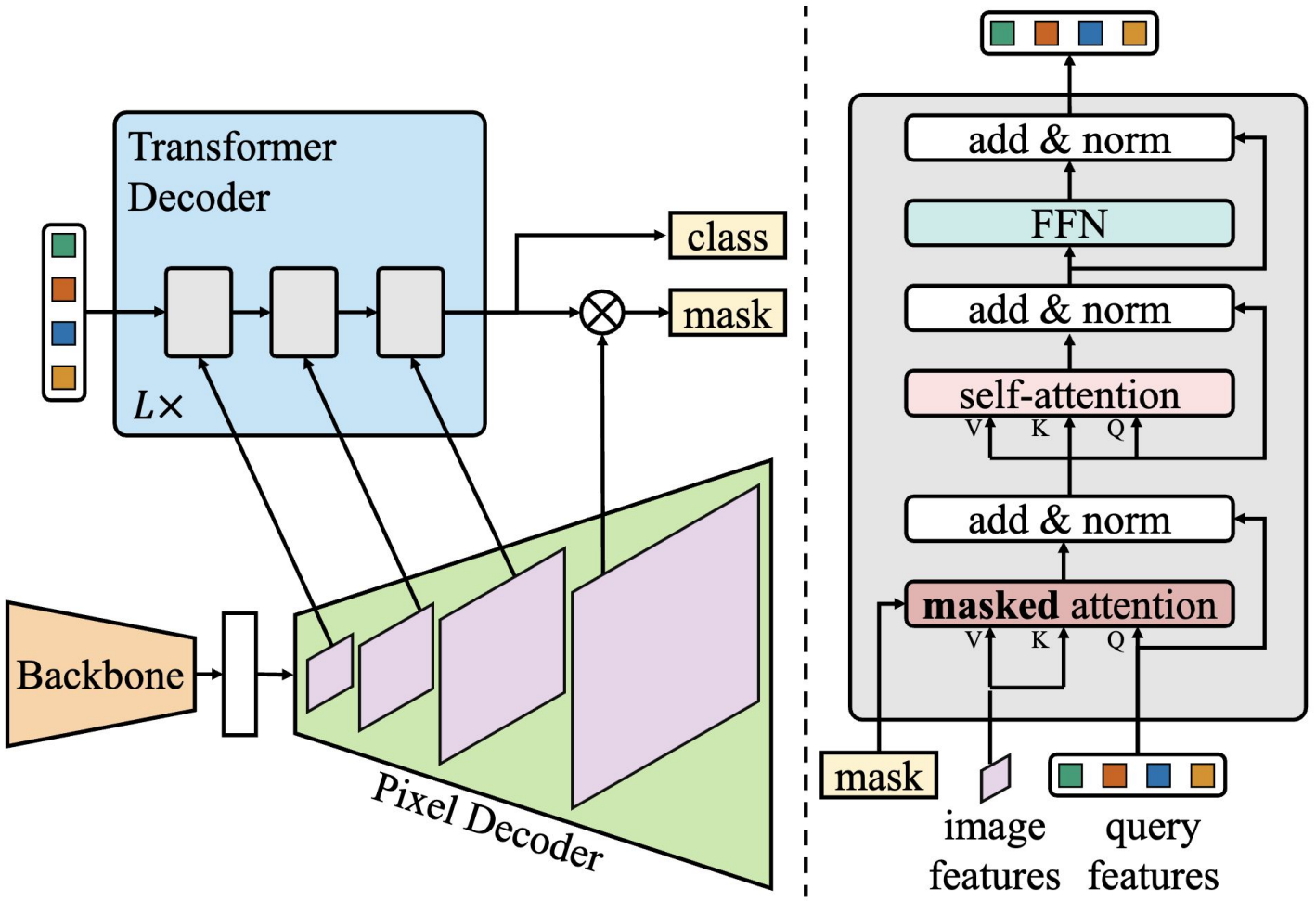


Figure 3. **RoIAlign:** The dashed gr
resents a feature map, the solid lines
(with 2×2 bins in this example), and t
the 4 sampling points in each bin. Ro
computes the value of each samplin
by bilinear interpolation from the near
points on the feature map. No quantiz
performed on any coordinates involve
RoI, its bins, or the sampling points.

Think of this as an additional pass after running Faster R-CNN on the image

Mask2Former: unified approach for labeled segmentation



Promptable masks with Segment Anything (SAM)

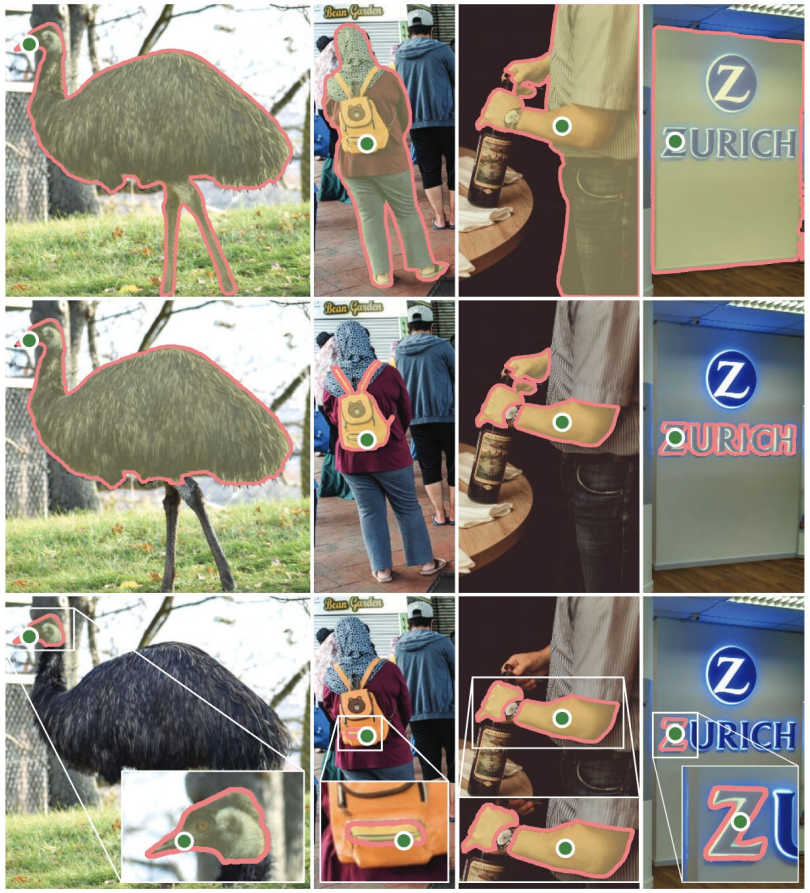


Figure 3: Each column shows 3 valid masks generated by SAM from a single ambiguous point prompt (green circle).

Image from Segment Anything

Class-agnostic segmentation with Segment Anything (SAM)

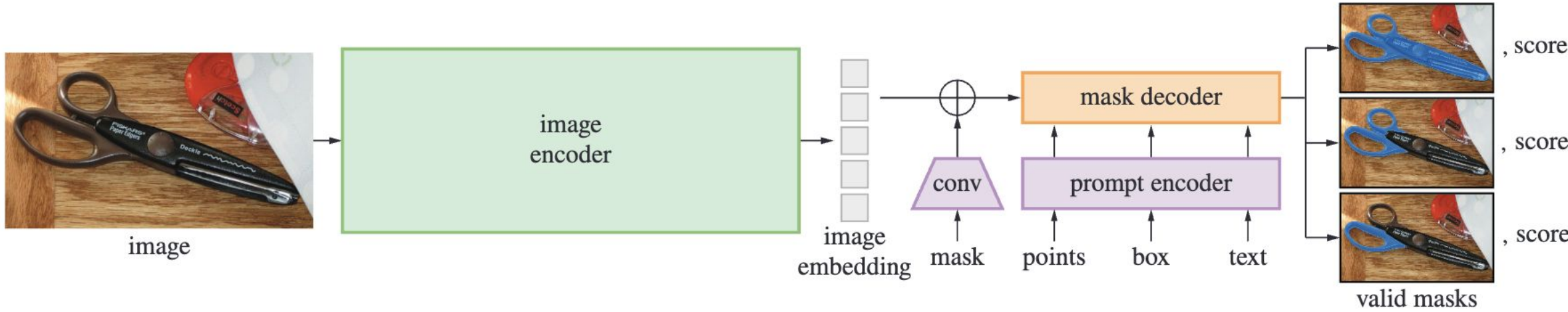


Figure 4: Segment Anything Model (SAM) overview. A heavyweight image encoder outputs an image embedding that ca then be efficiently queried by a variety of input prompts to produce object masks at amortized real-time speed. For ambiguou prompts corresponding to more than one object, SAM can output multiple valid masks and associated confidence scores.

Image from Segment Anything

Creating labeled masks with object detectors and SAM

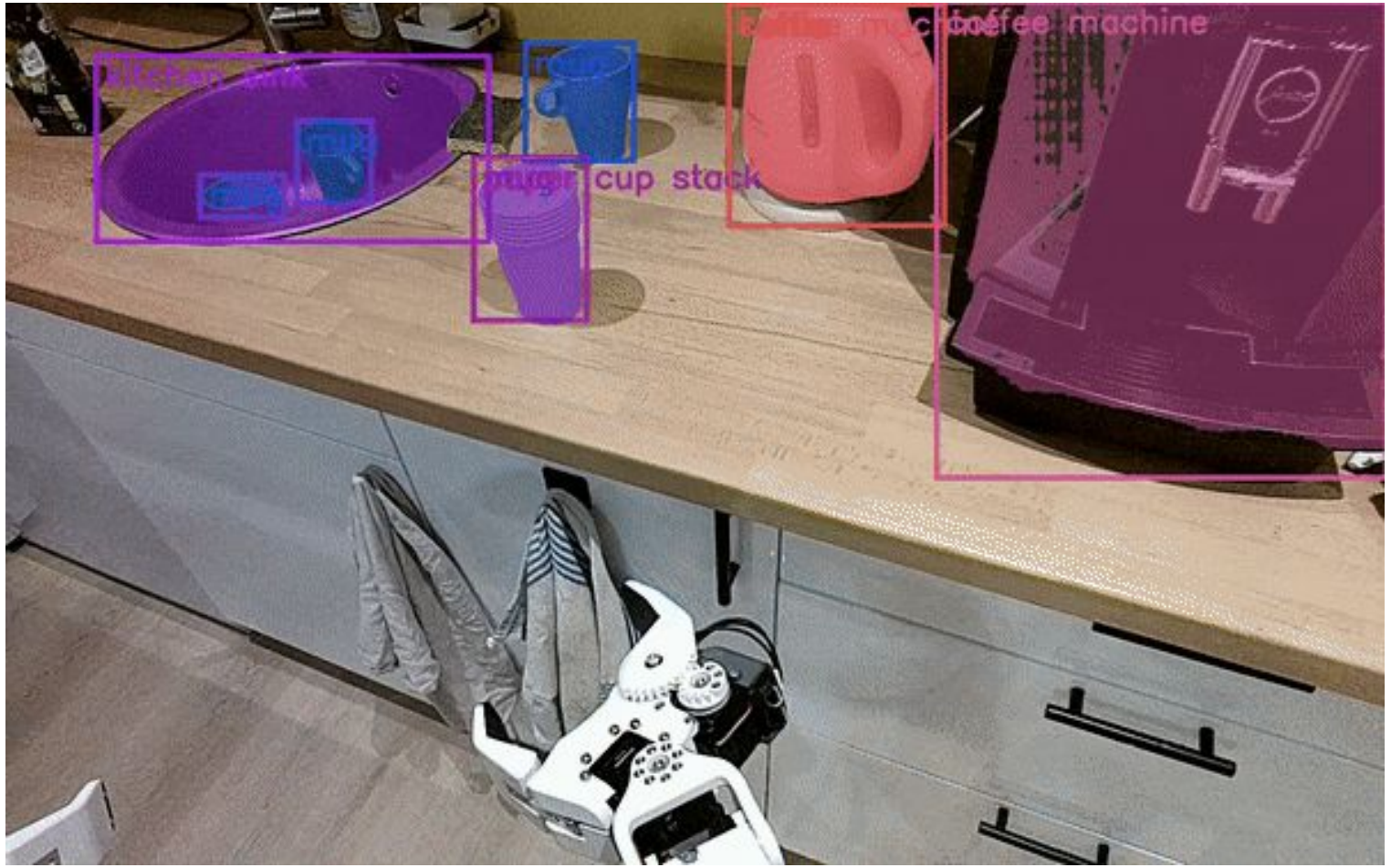


Image from https://github.com/pollen-robotics/pollen-vision

Summary

Semantic segmentation assigns class labels to individual pixels

- Loss functions compare predictions with ground truth masks at the pixel level

Instance segmentation separates objects of the same class

- Each detected object receives a unique mask identifier

Panoptic segmentation combines instance and semantic segmentation

- Labels pixels as countable (instance) or uncountable (semantic) classes

Segment anything (SAM) produces zero-shot masks

- Generates class-agnostic masks from prompts
- Integrates with object detectors for class labels

SPONSORED BY THE
Federal Ministry of Education and Research

Further reading and references

U-Net: Convolutional Networks for Biomedical Image Segmentation

- https://arxiv.org/abs/1505.04597

Mask R-CNN

- https://arxiv.org/abs/1703.06870

Mask2Former: Masked-attention Mask Transformer for Universal Image Segmentation

- https://arxiv.org/abs/2112.01527

Segment Anything

- https://arxiv.org/abs/2304.02643

SPONSORED BY THE
Federal Ministry of Education and Research