WEEK 7

Open-World Entity Segmentation

Lu Qi[*1] **Jason Kuen**[*2] **Yi Wang**[1] **Jiuxiang Gu**[2]
Hengshuang Zhao[3] **Zhe Lin**[2] **Philip Torr**[3] **Jiaya Jia**[1,4]
[1]The Chinese University of Hong Kong [2]Adobe Research
[3] University of Oxford [4]SmartMore

01

- **Background**

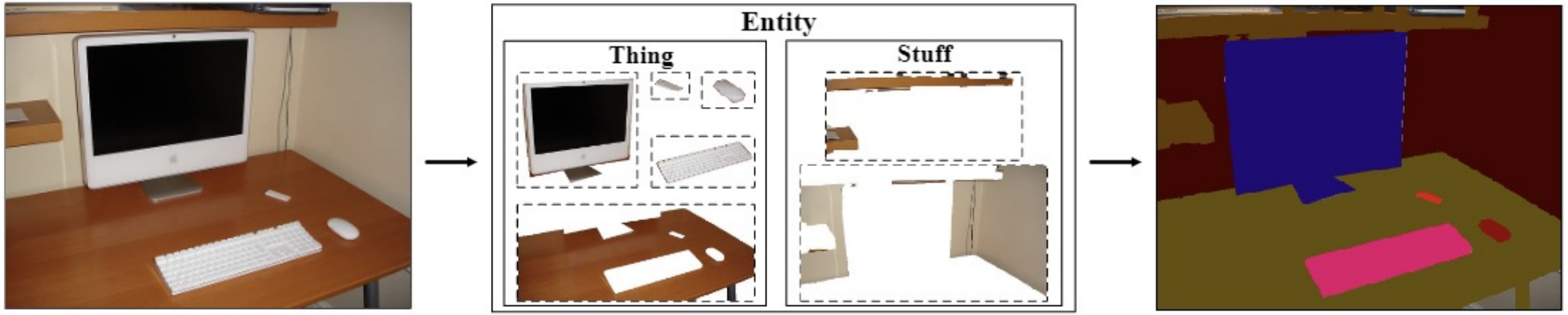
(a) image (b) semantic segmentation

(c) instance segmentation (d) panoptic segmentation

Figure 1: For a given (a) image, we show *ground truth* for: (b) semantic segmentation (per-pixel class labels), (c) instance segmentation (per-object mask and class label), and (d) the proposed *panoptic segmentation* task (per-pixel class+instance labels). The PS task: (1) encompasses both stuff and thing classes, (2) uses a simple but general format, and (3) introduces a uniform evaluation metric for all classes. Panoptic segmentation generalizes both semantic and instance segmentation and we expect the unified task will present novel challenges and enable innovative new methods.

01

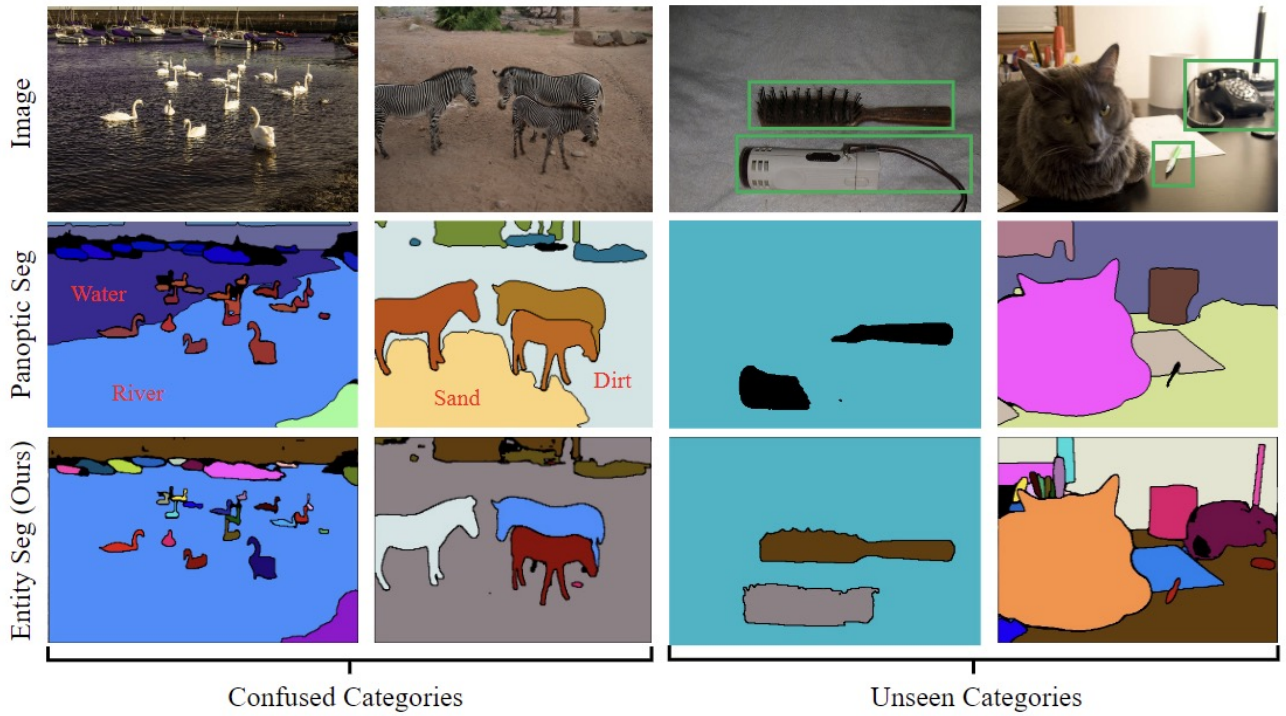
- **Background: thing and stuff**
- **thing:可数的目标，如：人、动物、车**
- **stuff:具有相似纹理或者材料的不规则区域，如草地、天空、马路**



01

- **ICCV (2021)**
- **Motivation**

In image manipulation and editing applications, where category labels are typically not required, the conventional category-oriented image segmentation may be sub-optimal and could introduce unnecessary category-related issues.

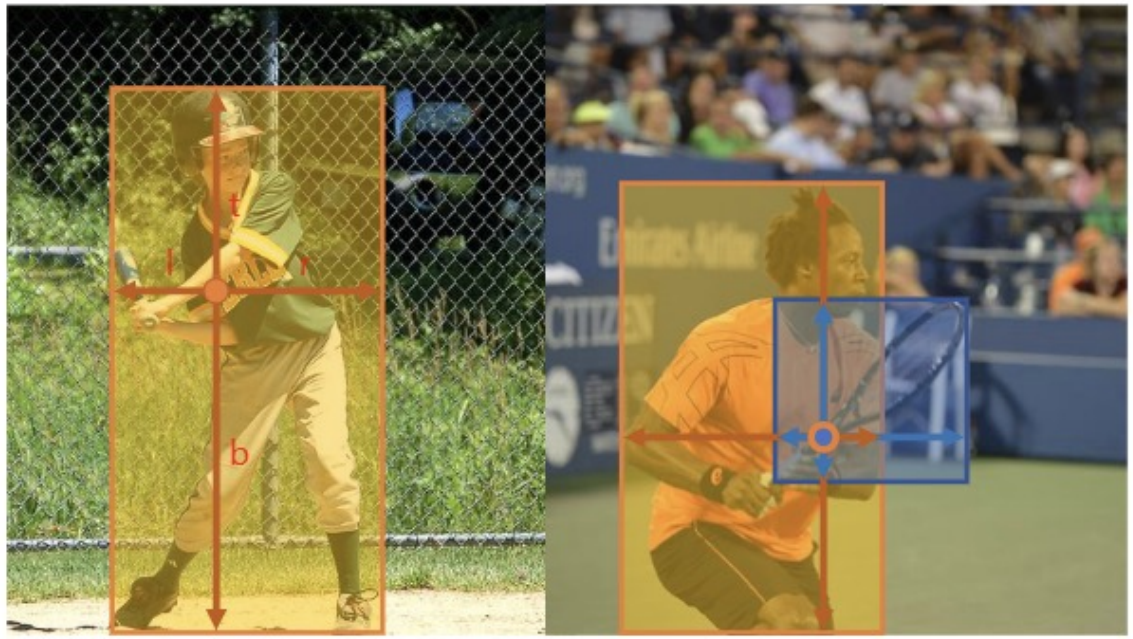


01

- **Contribution**
  - Propose a new task called entity segmentation, which aims to segment every entity without predicting its semantic category
  - Find that any entity (thing or stuff) can be very effectively and uniformly represented by center points in the network.
  - The extensive experiments show the remarkable effectiveness and generalization of our proposed method for entity segmentation

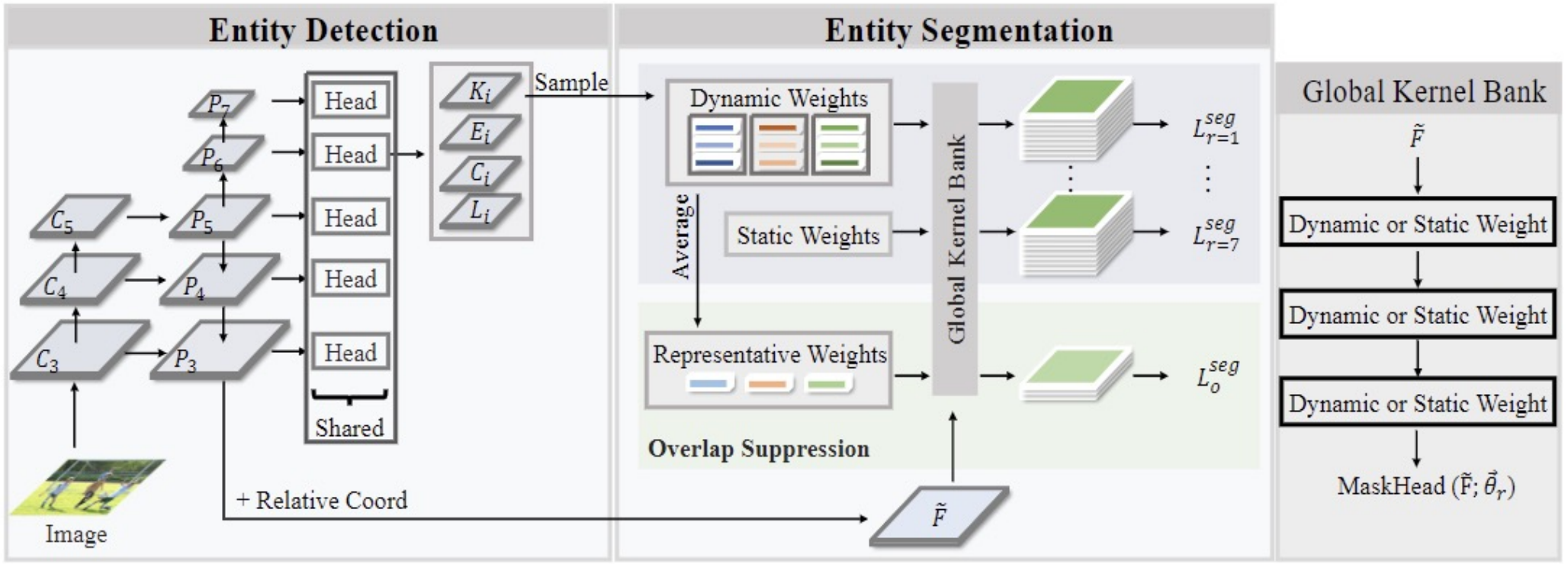
01

- **How to represent entity**
  - Here, we assume that each entity can be effectively represented by its center point



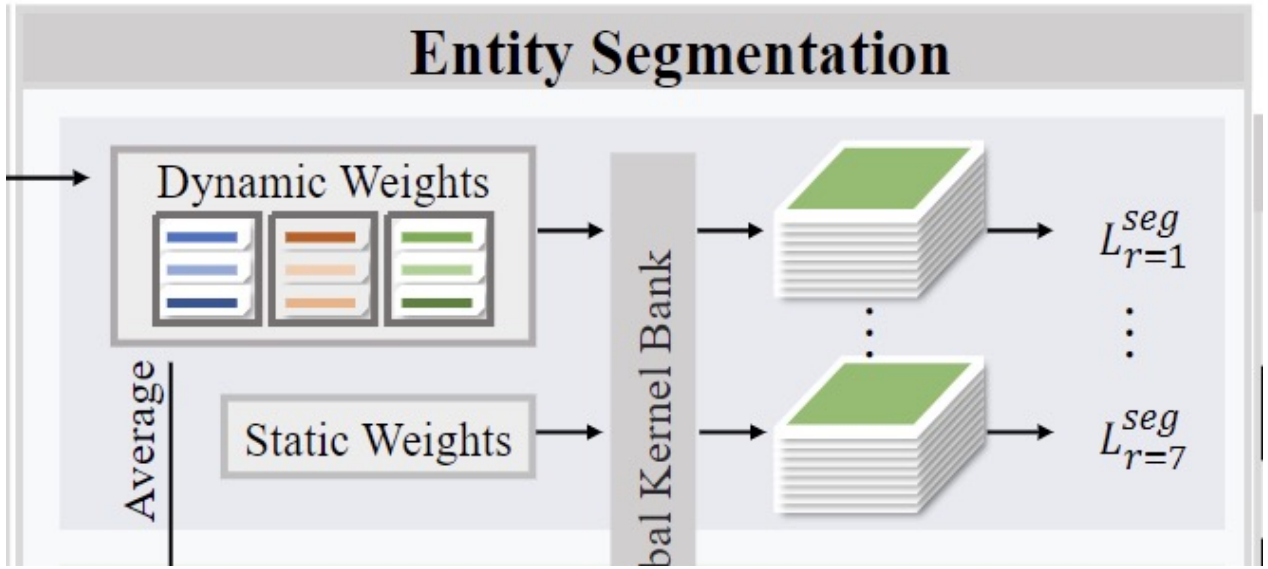
01

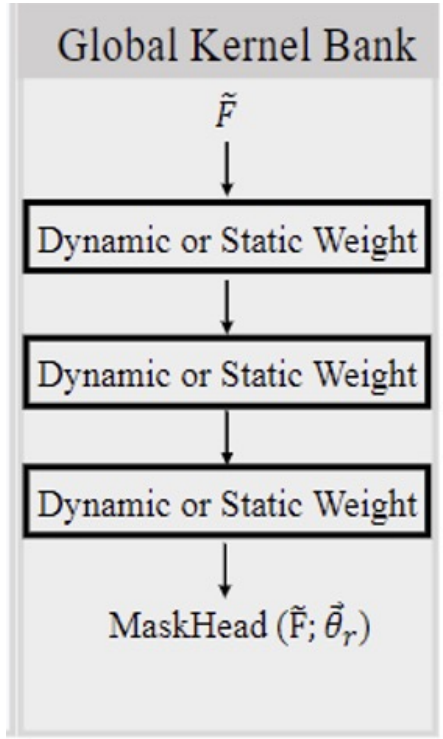
- **Overview of the Proposed Model**



01

- **Entity Segmentation**

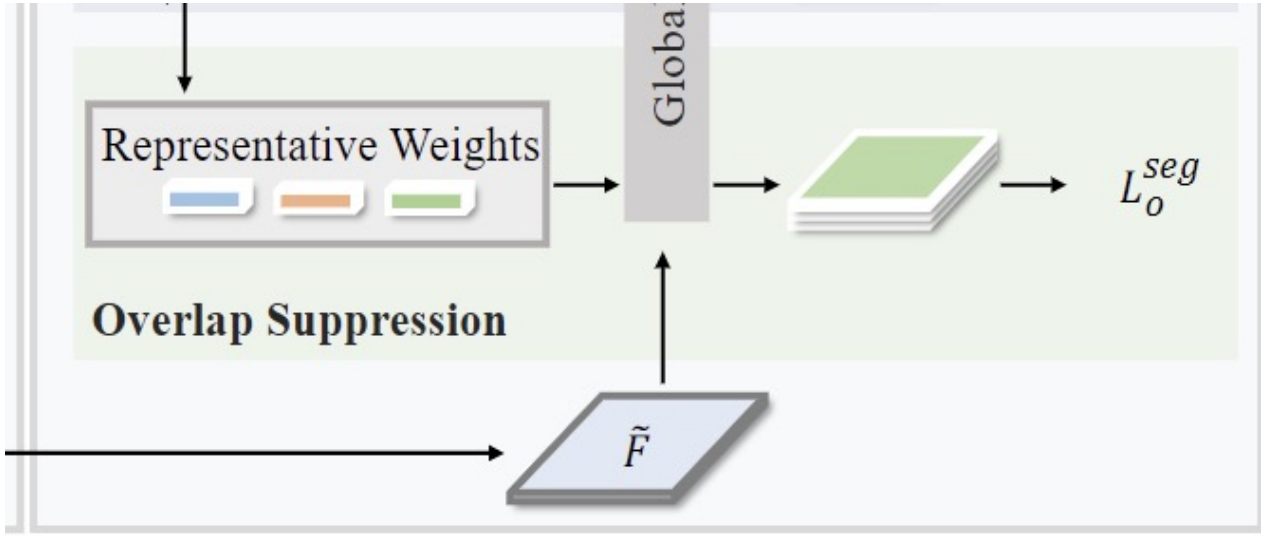




$$\mathcal{L}_r^{seg} = \lambda_r \times \text{Dice}(\text{Sigmoid}(\text{MaskHead}(\tilde{\mathbf{F}}; \vec{\theta}_r)), \mathbf{Y}),$$

01

- **Overlap Suppression**



$$\vec{\theta}_o^n = \text{Average}(K_n) = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} \mathbb{1}^{m\in n} K_n}{\sum_{n=1}^{N}\sum_{m=1}^{M} \mathbb{1}^{m\in n}}$$

$$\mathcal{L}_o^{seg} = \text{Dice}(\text{Softmax}(\text{MaskHead}(\tilde{\mathbf{F}}; \vec{\theta}_o)), \mathbf{Y}),$$

- **Experimental Results**

Model	PQ	AP^m	AP^m_e
PanopticFPN [18]	39.4	-	23.2
PanopticFCN [20]	41.1	-	24.5
DETR [21]	43.4	-	24.8
Ours	-	34.6	29.8

(a)

$\mathcal{L}_R^{seg}$	$\mathcal{L}_o^{seg}$	AP^m_e
○	○	28.3
✓	○	29.1
○	✓	29.3
✓	✓	29.8

(b)

Softmax	Sigmoid	AP^m_e
○	○	28.3
○	✓	28.6
✓	○	29.1
✓	✓	29.0

(c)

Table 2: **(a)**: **Comparison with the existing panoptic segmentation methods**. For the existing panoptic segmentation methods, we merely convert their panoptic results to the ES format and obtain the entity scores for "*stuff*" entities by averaging the scores within each *stuff* mask. PQ is the evaluation metric for conventional category-aware panoptic segmentation. AP^m is the overlap-tolerated entity segmentation evaluation metric similar to common instance segmentation metrics. **(b)**: **Proposed modules**. The effect of global kernel bank and overlap suppression module. **(c)**: **Overlap suppression**. The ablation study of scoring activation function in the module.

- **Experimental Results**

111	110	101	100	011	010	001	AP^m_e	$AP^m_{e_{50}}$	$AP^m_{e_{75}}$
1.0	0.0	0.0	0.0	0.0	0.0	0.0	28.3	48.7	28.8
2.0	0.0	0.0	0.0	0.0	0.0	0.0	28.1	48.4	28.5
1.0	1.0	0.0	0.0	0.0	0.0	0.0	28.9	49.3	29.1
1.0	1.0	1.0	0.0	0.0	0.0	0.0	29.1	49.8	29.6
1.0	1.0	1.0	1.0	1.0	1.0	1.0	27.8	47.8	27.7
1.0	1.0	1.0	0.25	0.25	0.25	0.25	29.3	50.2	29.8

(a)

MODEL	AP^m_e	$AP^m_{e_{50}}$	$AP^m_{e_{75}}$
Baseline	29.8	50.3	30.9
R-50	31.8	53.5	33.8
R-50-DCNv2	33.7	56.1	35.6
R-101	33.2	55.5	34.8
R-101-DCNv2	35.5	58.2	37.1
Swin-L	38.6	62.4	40.8

(b)

Table 3: Ablation studies. **(a)**: **Global Kernel Bank**. The r in $\mathcal{L}_R^{seg}$ ranges from 1 to 7. In the first row, "xxx" is binary representation of r. For example, "100" corresponds to the 4-th path with the first layer (1) and last two (00) layers using dynamic and static weights, respectively.Each entry below path IDs indicates the loss weight λ_r. **(b)**: **High-Performance Regime**. The performance of our models enhanced by stronger backbones and a longer training duration. “Swin-L” and “DCNv2” refer to Swin Transformer [86] in large series with window size 7 and deformable Convolution v2 [53].