SOGNet: Scene Overlap Graph Network for Panoptic Segmentation

Yibo Yang, Hongyang Li, Xia Li, Qijie Zhao, Jianlong Wu and Zhouchen Lin

Summary

- **Problem**

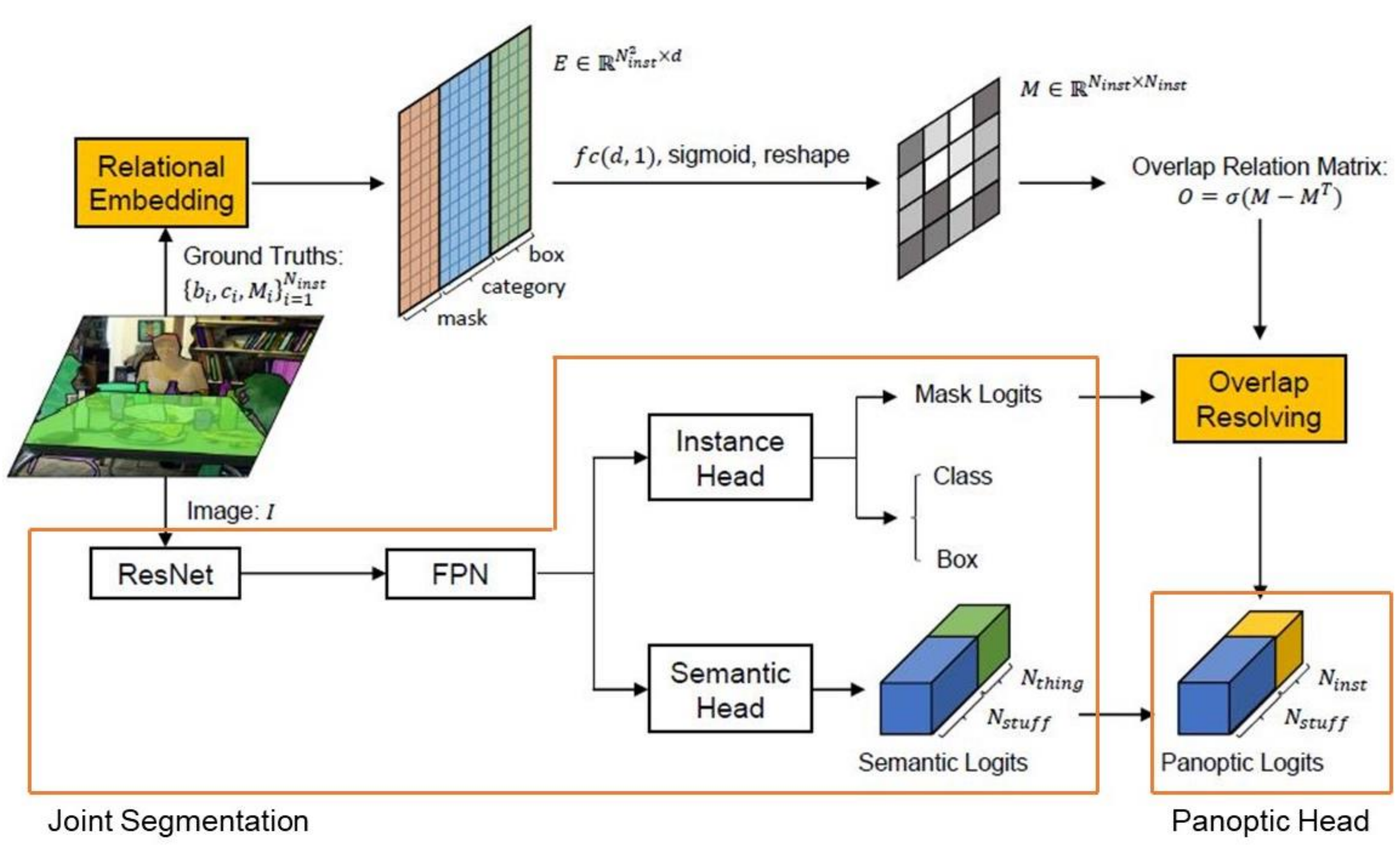
Panoptic segmentation cannot have overlapping segments. But most cutting-edge high-performance instance segmentation methods adopt the region-based strategy, and output overlapping segments.

Image — Instance Segmentation — Panoptic Segmentation

- **Contributions**

1. We propose an end-to-end framework, SOGNet, to explicitly encode overlap relations among objects, and resolve the overlap between any pair of objects in a differentiable way.
2. State-of-the-art performance on the COCO and Cityscapes datasets.

Architecture



Relations Predicted by SOGNet

The activation on location (i, j) represents that the object i is covered by (lies below) object j. The indices of objects are marked in the images.

Methods

- **Joint Segmentation**

We use ResNet with FPN as the shared backbone of semantic and instance branches. The Mask R-CNN structure is adopted for instance segmentation head. For semantic head, the FPN feature maps first go through three 3x3 deformable convolution layers, and then are up-sampled to the 1/4 scale.

- **Relational Embedding**

$$E^{(c)}_{i|j} = P^T\left(\sigma(V^T c_i)\circ\sigma(U^T c_j)\right), \quad E^{(c)} = \left[E^{(c)}_{1|1}, E^{(c)}_{1|2}, \dots, E^{(c)}_{N_{inst}|N_{inst}}\right]^T \in R^{N^2_{inst}\times d_c},$$

$$E^{(b)}_{i|j} = K^T\left(\frac{x_i-x_j}{w_j}, \frac{y_i-y_j}{h_j}, \log\left(\frac{w_i}{w_j}\right), \log\left(\frac{h_i}{h_j}\right)\right)^T, \quad E = \left[E^{(m)}, E^{(c)}, E^{(b)}\right] \in R^{N^2_{inst}\times d},$$

where $\circ$ denotes the Hadamard product, σ is the ReLU activation, V, U, P and $K \in R^{4\times d_b}$ are linear embeddings, and $d = d_m + d_c + d_b$.

- **Overlap Resolving**

$$E \in R^{N^2_{inst}\times d} \xrightarrow{fc(d,1),sigmoid,reshape} M \in R^{N_{inst}\times N_{inst}}, \quad O = \sigma(M - M^T) \in R^{N_{inst}\times N_{inst}},$$

$$A'_i = A_i - A_i \circ \left[s(A_i)\circ s(A_j)\right]O_{ij}, \quad A'_i = A_i - A_i \circ s(A_i) \circ \sum_{j=1}^{N_{inst}} s(A_j)O_{ij},$$

(overlap of j on i) (overlap of all the other objects on i)

The computational step of overlap resolving is formulated as:

$$\mathcal{A}' = \mathcal{A} - \mathcal{A}\circ s(\mathcal{A})\circ\left(s(\mathcal{A})\times_3 O^T\right)$$

where s denotes the sigmoid function, $\mathcal{A} = \left[A_1, A_2, \dots, A_{N_{inst}}\right] \in R^{H\times W\times N_{inst}}$, and $\times_3$ denotes the Tucker product along the 3-rd dimension (reshape $s(\mathcal{A})$ as $R^{HW\times N_{inst}}$ for inner product with O^T, and then return to $R^{H\times W\times N_{inst}}$). In this way, our method explicitly encodes overlap relations by O, and is differentiable for resolving the overlap between any pair of objects.

- **Panoptic Head**

Panoptic Head 1: $Z_i = X_i + A'_i$

Panoptic Head 2: $Z_i = k\cdot X_i \circ s(A'_i) + A'_i$

where Z_i is the combines logit, and k is a factor to balance the numerical difference between semantic output values and mask logits.

Selected Experimental Results

Models	backbone	PQ	PQ^{Th}	PQ^{St}
Cityscapes				
Q.Li *et al.*	ResNet-101	53.8	42.5	62.1
Panoptic FPN	ResNet-101	58.1	52.0	62.5
TASCNet	ResNet-50	59.3	56.3	61.5
UPSNet	ResNet-50	59.3	54.6	62.7
SOGNet	**ResNet-50**	**60.0**	**56.7**	**62.5**
COCO				
JSIS	ResNet-50	26.9	29.3	23.3
Panoptic FPN	ResNet-101	40.3	47.5	29.5
OCFusion	ResNet-50	41.2	49.0	29.0
UPSNet	ResNet-50	42.5	48.5	33.4
SOGNet	**ResNet-50**	**43.7**	**50.6**	**33.2**

Arxiv: https://arxiv.org/abs/1911.07527.pdf

Contact: ibo@pku.edu.cn