SIM: Semantic-aware Instance Mask Generation for Box-Supervised Instance Segmentation

Ruihuang Li *, *Chenhang He**, *Yabin Zhang, Shuai Li, Liyi Chen, Lei Zhang†*

[1]The Hong Kong Polytechnic University

Code: https://github.com/lslrh/SIM

Background

Task:

Achieving high-performance instance segmentation with only **box annotations**.

Issues:

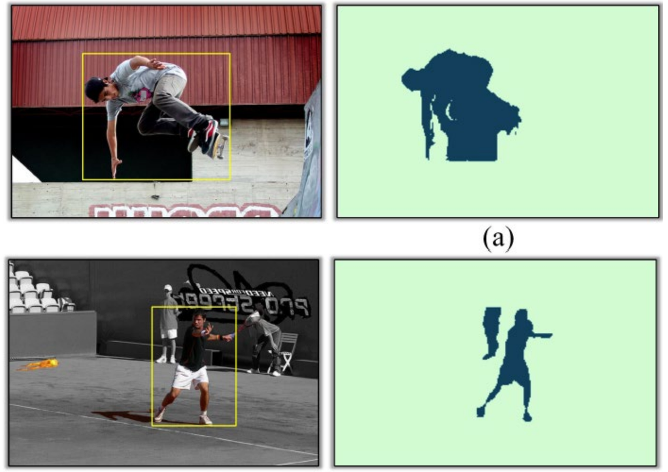
Existing methods fails in the following two cases:

- **Proximal pixels with similar color but different semantics**

  Assumption of pairwise loss: Proximal pixels with similar colors are from the same category

- **Different instances with the same semantics**

(occluded objects)



(a)

(c)

Method

- Pseudo Semantic Map

Low-level image features, such as colors, intensity, edges, blobs, could provide useful guidance to identify the object boundaries in an image. However, these features vary significantly with illuminations, motion blurs, and noises.

We first construct a group of representative prototypes to model the structural information of objects. Then we compute the semantic probability map corresponding to the *c*-th category, using the following formula:

$$M_{\mathrm{S},i}^{c} = \sigma(\max\{\frac{\langle z_i, p_l^c \rangle}{\tau}\}_{l=1}^{L}),$$

- Multi-prototype update

We update the prototypes on-the-fly with the moving average of cluster centroids computed in previous mini-batches. The cluster assignments can be obtained by solving the optimal transport problem:

$$\max_{Q \in \mathbb{Q}} \mathrm{Tr}(Q^T P_c^T Z) + \varepsilon H(Q), \quad s.t. Q \in \mathbb{Q},$$

$$\text{with} \quad \mathbb{Q} := \{Q \in \mathbb{R}_+^{L \times N_c} | Q\mathbf{1}_{N_c} = r,\ Q^T \mathbf{1}_L = h\}.$$

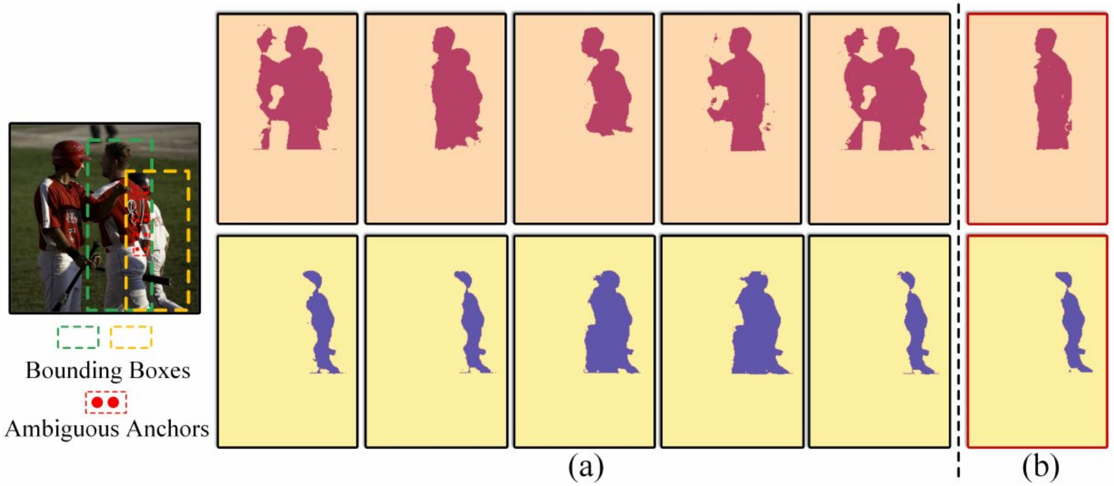
Method

- Self-Correction

Though the pseudo semantic masks could provide more reliable supervision from a global perspective, they could not distinguish different objects of the same semantics, especially when there exist overlaps or occlusions among objects.

Positive Mask Weighting

The quality of masks produced by different positive samples varies significantly. Based on this observation, we propose a positive mask weighting strategy to integrate different masks according to their quality, resulting in a high-quality instance-aware mask.



Method

- Online Weakly-Supervised Copy-Paste

It is natural to employ pseudo masks as the guidance to cut object instances from an image. To achieve online Copy-Paste, we set up a first-in-first-out memory bank to store training samples and their corresponding pseudo masks from preceding mini-batches, which ensures that the pseudo masks in memory bank could be updated on-the-fly.

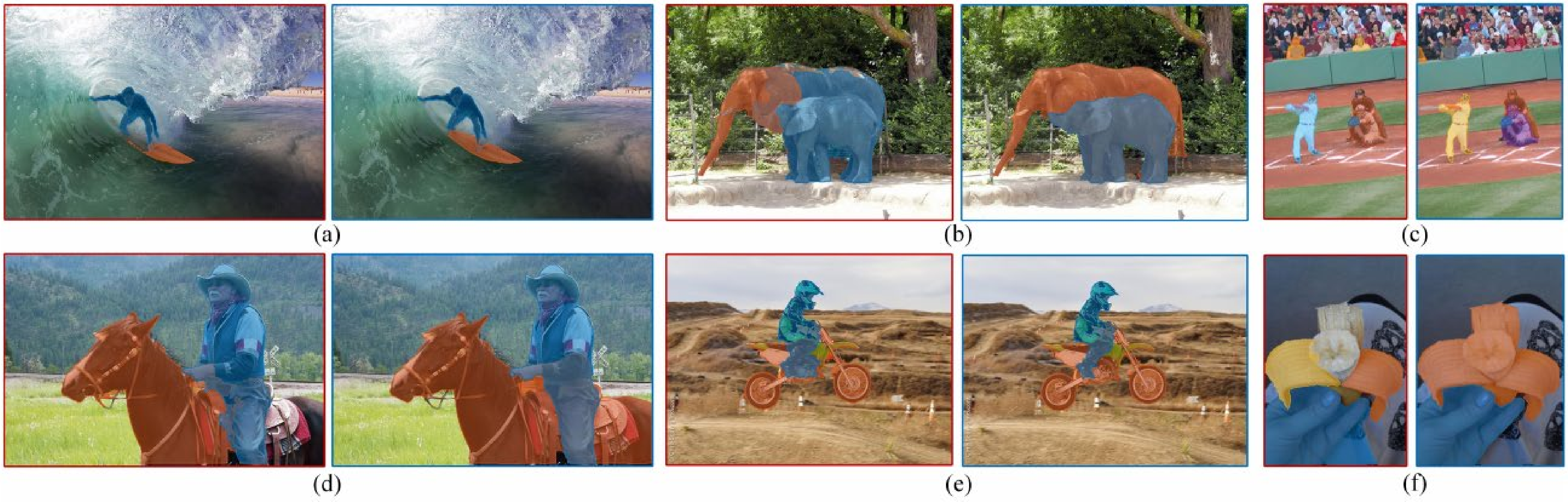
Experiments

- Comparisons with state-of-the-art methods

	method	backbone	sche.	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
fully-supervised	Mask R-CNN [12]	ResNet-101-FPN	3×	37.5	59.3	40.2	21.1	39.6	48.3
	YOLACT-700 [3]	ResNet-101-FPN	4.5×	31.2	50.6	32.8	12.1	33.3	47.1
	PolarMask [40]	ResNet-101-FPN	2×	32.1	53.7	33.1	14.7	33.8	45.3
	SOLOv2 [38]	ResNet-101-FPN	3×	39.7	60.7	42.9	17.3	42.9	57.4
	CondInst [34]	ResNet-101-FPN	3×	39.1	60.9	42.0	21.5	41.7	50.9
	Mask2Fomer† [7]	ResNet-101-MSDefomAttn	50e	44.2	-	-	23.8	47.7	66.7
box-supervised	BBTP† [14]	ResNet-101-FPN	1×	21.1	45.5	17.2	11.2	22.0	29.8
	BBAM [21]	ResNet-101-FPN	1×	25.7	50.0	23.3	-	-	-
	BoxCaseg‡ [37]	ResNet-101-FPN	1×	30.9	54.3	30.8	12.1	32.8	**46.3**
	SIM (Ours)	ResNet-101-FPN	1×	**34.0**	**56.8**	**35.0**	**17.2**	**36.8**	45.5
	BoxLevelSet [23]	ResNet-101-FPN	3×	33.4	56.8	34.1	15.2	36.8	46.8
	BoxInst [36]	ResNet-101-FPN	3×	33.2	56.5	33.6	16.2	35.3	45.1
	SIM (Ours)	ResNet-101-FPN	3×	**35.3**	**58.9**	**36.4**	**18.4**	**38.0**	**47.5**
	BoxLevelSet [23]	ResNet-DCN-101-BiFPN	3×	35.4	59.1	36.7	16.8	38.5	51.3
	BoxInst [36]	ResNet-DCN-101-BiFPN	3×	35.0	59.3	35.6	17.1	37.2	48.9
	SIM (Ours)	ResNet-DCN-101-BiFPN	3×	**37.4**	**61.8**	**38.6**	**18.6**	**40.2**	**51.6**
	BoxInst [36]	Swin-B-FPN	3×	37.9	63.2	39.0	20.0	41.2	53.1
	SIM (Ours)	Swin-B-FPN	3×	**40.2**	**66.9**	**41.3**	**21.1**	**43.5**	**56.0**
	BoxInst† [36]	Mask2Former-ResNet-101	50e	35.7	59.8	36.4	16.6	38.5	55.4
	SIM† (Ours)	Mask2Former-ResNet-101	50e	**37.4**	**62.2**	**38.7**	**17.6**	**41.3**	**56.6**

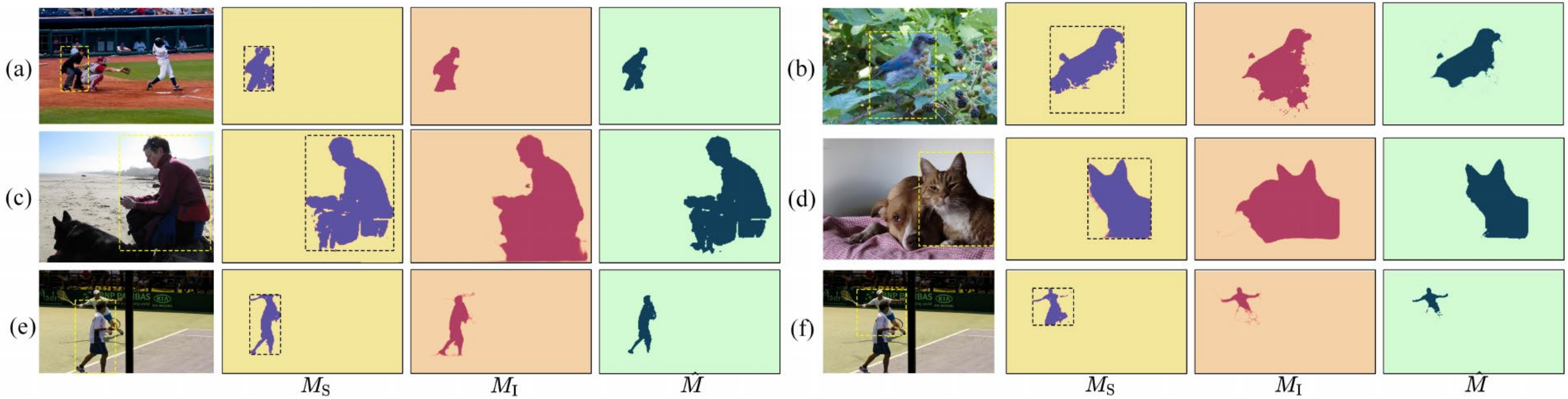
Experiments

- Qualitative Results



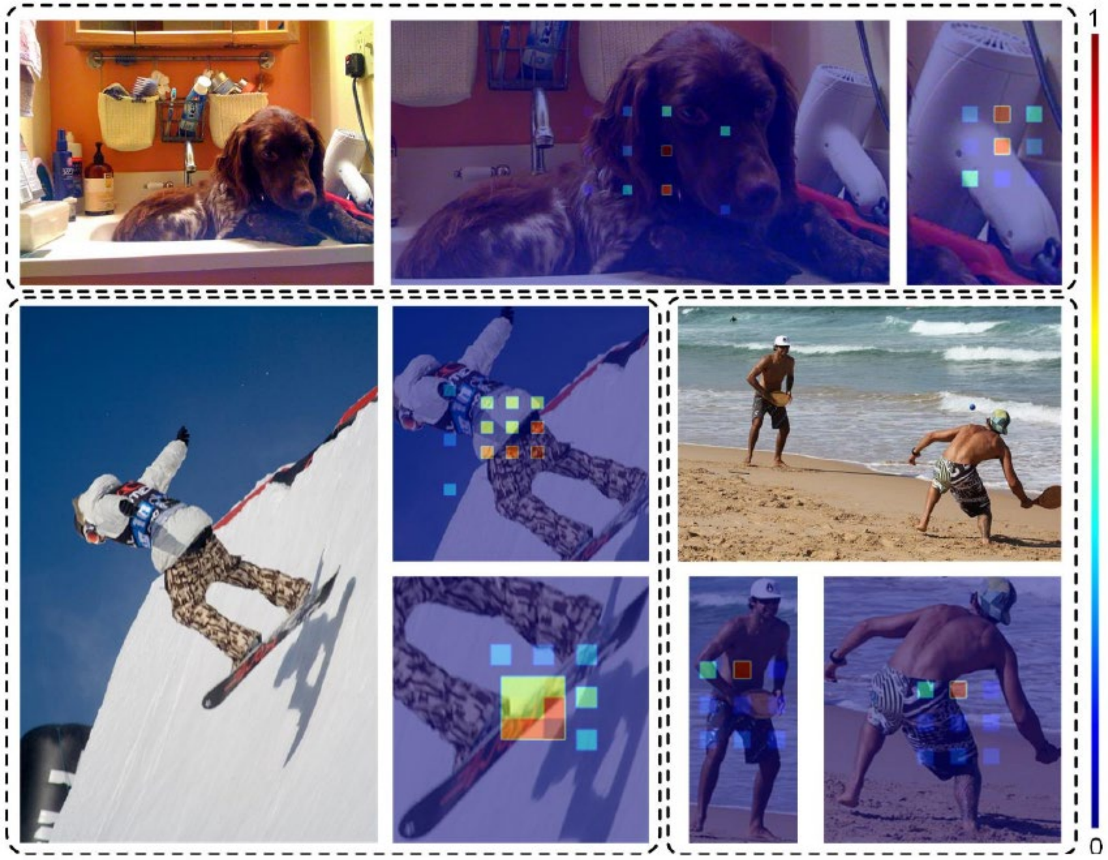
Experiments

- Comparison of different pseudo masks



Experiments

- Visualizations of weights for different positive samples



The End

Thanks!