Box-supervised Instance Segmentation with Level Set Evolution

Wentong Li, Wenyu Liu, Jianke Zhu*, Miaomiao Cui, Xiansheng Hua, Lei Zhang

https://github.com/LiWentomng/boxlevelset

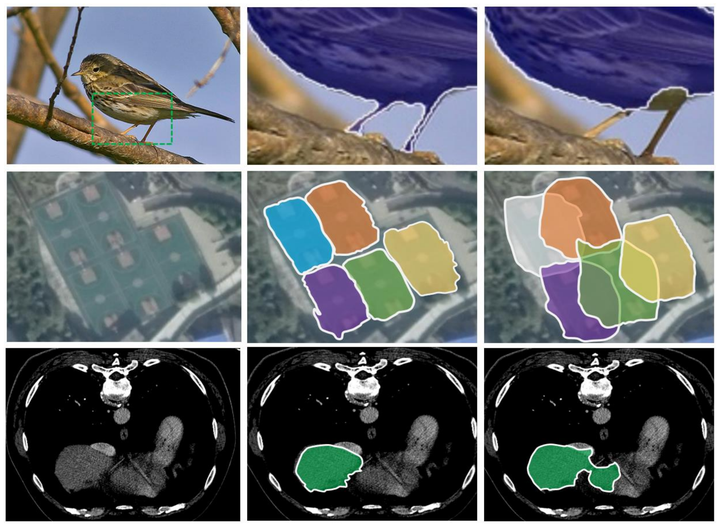
ECCV EUROPEAN CONFERENCE ON COMPUTER VISION TEL AVIV 2022

Motivation

Forms of supervisions (annotations):

- Fully pixel-wise mask ✗
- Simple bounding box ✓

Limitations of existing box-supervised method:



(a) Input (b) Ours (c) BoxInst

- Simple pairwise affinity modeling on neighboring pixel or colour pairs, which is susceptible to noisy context.

Key idea

- Instance-aware mask map as the predicted level sets.
- Level set evolution on the low-level and high-level features with robust initialization.

Contributions

- A novel single-shot level set-based approach to box-supervised instance segmentation.
- Leading performance in various scenarios, including general scene, remote sensing and medical image.

Method

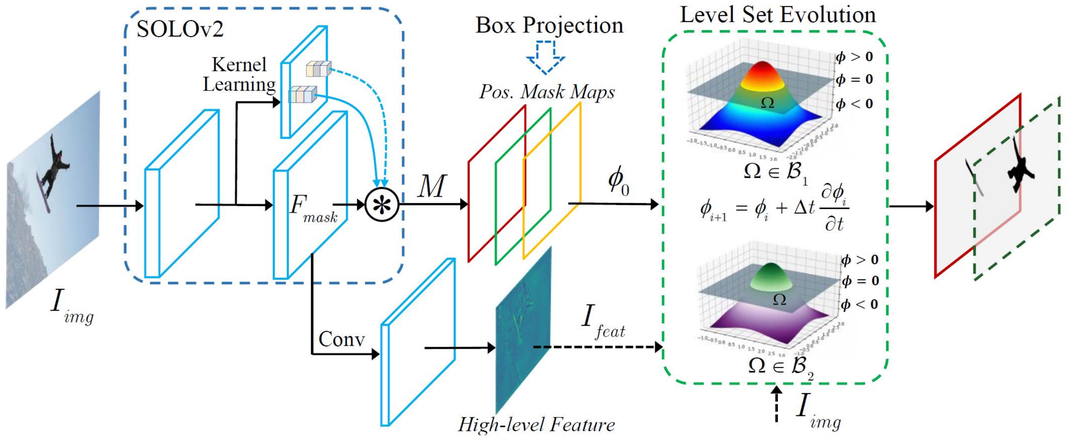


Fig. 2: Overview of our method. Our framework is designed based on SOLOv2. The positive mask maps *M* are obtained by level set evolution within the bounding box region. With the iterative energy minimization, the accurate instance segmentation can be obtained with box annotations only. The level set are only performing during training, and inference process is same as the original SOLOv2.

Level Set Energy

$$\mathcal{F}(\phi, I, c_1, c_2, \mathcal{B}) = \int_{\Omega \in \mathcal{B}} |I^*(x,y) - c_1|^2 \sigma(\phi(x,y)) dxdy + \int_{\Omega \in \mathcal{B}} |I^*(x,y) - c_2|^2 (1 - \sigma(\phi(x,y))) dxdy + \gamma \int_{\Omega \in \mathcal{B}} |\nabla \sigma(\phi(x,y))| \, dxdy$$

Input Data Terms

$$\mathcal{F}(\phi) = \lambda_1 * \mathcal{F}(\phi, I_u, c_{u_1}, c_{u_2}, \mathcal{B}) + \lambda_2 * \mathcal{F}(\phi, I_f, c_{f_1}, c_{f_2}, \mathcal{B})$$

Level Set Initialization

$$\mathcal{F}(\phi_0)_{box} = \mathcal{P}_{dice}(m_x^p, m_x^b) + \mathcal{P}_{dice}(m_y^p, m_y^b)$$

Network Learning

$$L = L_{cate} + L_{inst} \qquad L_{inst} = \frac{1}{N_{pos}} \sum_k \mathbb{1}_{\{p_{i,j}^* > 0\}} \{\mathcal{F}(\phi) + \alpha \mathcal{F}(\phi_0)_{box}\}$$

Ablation of Level Set Energy

$\mathcal{F}_{\phi_0}$	$\mathcal{F}_\phi(I_u)$	$\mathcal{F}_\phi(I_f)$	$\Omega \in \mathcal{B}$	$\Omega \in I$	AP	AP_{50}	AP_{75}
✓			✓		19.7	47.4	13.9
✓	✓		✓		22.2	49.5	17.4
✓	✓	✓	✓		**24.7**	**53.3**	**20.8**
✓	✓	✓		✓	21.7	48.4	17.4

Performance

methods	backbone	AP	AP_{25}	AP_{50}	AP_{70}	AP_{75}
BBTP w/ CRF	ResNet-101	27.5	-	59.1	-	21.9
BoxInst	ResNet-50	34.3	-	59.1	-	34.2
BoxInst	ResNet-101	36.5	-	61.4	-	37.0
DiscoBox	ResNet-50	-	71.4	59.8	41.7	35.5
DiscoBox	ResNet-101	-	72.8	62.2	45.5	37.5
Ours	ResNet-50	36.3	76.3	64.2	43.9	35.9
Ours	ResNet-101	**38.3**	**77.9**	**66.3**	**46.4**	**38.7**

Table 1. Results on Pascal VOC.

method	backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
mask-supervised:							
Mask R-CNN	ResNet-101	35.7	58.0	37.8	15.5	38.1	52.4
YOLACT-700	ResNet-101	31.2	50.6	32.8	12.1	33.3	47.1
PolarMask	ResNet-101	32.1	53.7	33.1	14.7	33.8	45.3
CondInst	ResNet-101	39.1	60.9	42.0	21.5	41.7	50.9
SOLOv2	ResNet-101	39.7	60.7	42.9	17.3	42.9	57.4
box-supervised:							
BBTP†	ResNet-101	21.1	45.5	17.2	11.2	22.0	29.8
BBAM	ResNet-101	25.7	50.0	23.3	-	-	-
BoxCaseg*	ResNet-101	30.9	54.3	30.8	12.1	32.8	46.3
BoxInst	ResNet-101	33.2	56.5	33.6	16.2	35.3	45.1
BoxInst	ResNet-101-DCN	35.0	**59.3**	35.6	**17.1**	37.2	48.9
Ours	ResNet-101	33.4	56.8	34.1	15.2	36.8	46.8
Ours	ResNet-101-DCN	**35.4**	59.1	**36.7**	16.8	**38.5**	**51.3**

Table 2. Performance on COCO test 2017.

method	backbone	Sup.	AP
DeepSnake	DLA-34	*Mask*	30.5
Levelset R-CNN	ResNet-50	*Mask*	34.3
DVIS-700	ResNet-50	*Mask*	32.6
DVIS-700	ResNet-101	*Mask*	**35.7**
Ours	ResNet-101	***Box***	33.0
Ours	ResNet-101-DCN	***Box***	35.0

Table 3. Deep variational instance segmentation.

method	backbone	Sup.	AP
Mask R-CNN	R-50-C4	*Mask*	28.8
PolarMask	R-50-FPN	*Mask*	27.2
CondInst	R-50-FPN	*Mask*	29.5
BoxInst	R-50-FPN	*Box*	17.8
Ours	R-50-FPN	*Box*	**20.1**

method	backbone	Sup.	AP
Mask R-CNN	R-50-FPN	*Mask*	64.2
BoxInst	R-50-FPN	*Box*	40.7
Ours	R-50-FPN	*Box*	**44.5**

Table 4. Results on remote sensing and medical scene.

Fig 3. Visual results on COCO and iSAID.

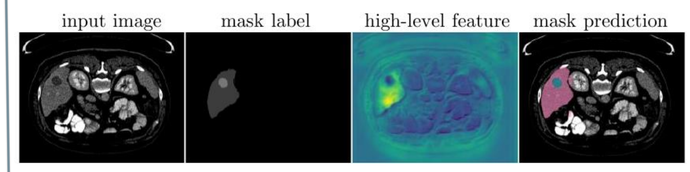


Fig 4. Visual results on LiTS.