Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision

Xiaokang Chen[1], Yuhui Yuan[2], Gang Zeng[1], Jingdong Wang[2]
[1]Key Laboratory of Machine Perception (MOE), Peking University [2]Microsoft Research Asia

CVPR VIRTUAL JUNE 19-25

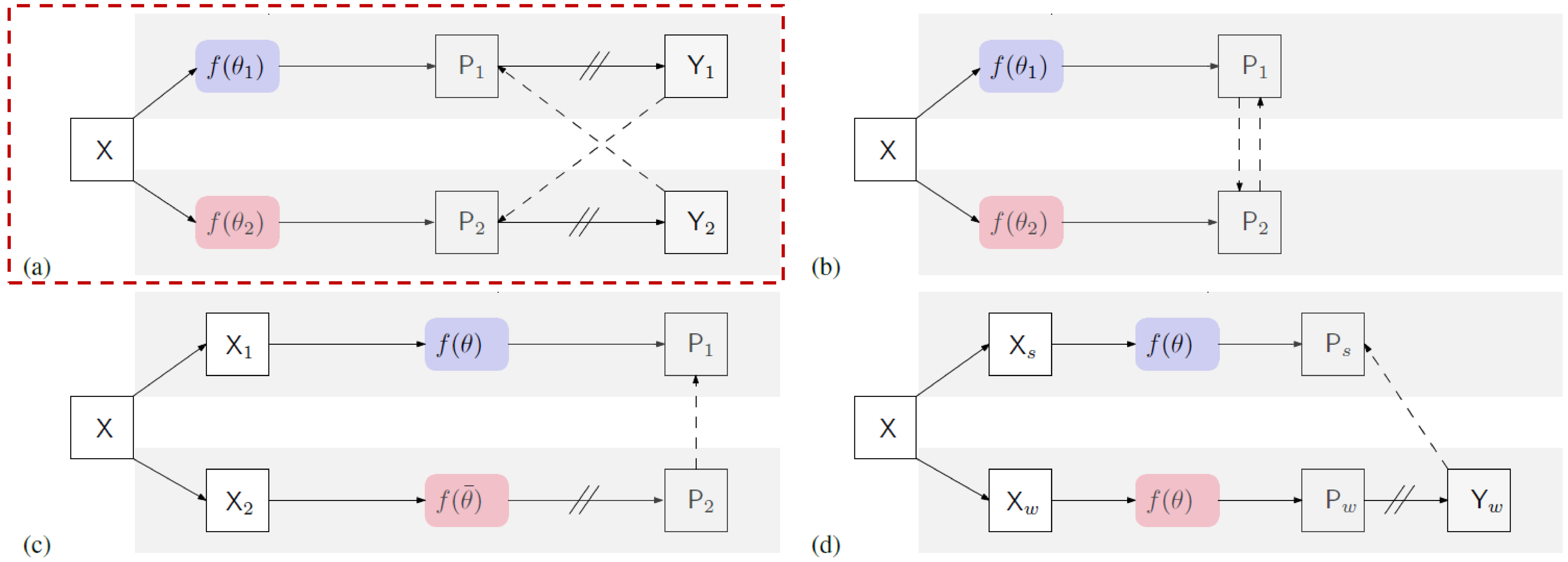
1. Motivation

- Data labelling is hard and expensive for segmentation. We could easily obtain the images of the city by taking photos, but it is hard to label each pixel carefully, such as the pole.
- Self-training expands the labeled set through pseudo labelling, but it is not end-to-end. Consistency learning encourages the network to learn a compact feature embedding. Why not combine these two types of methods?

2. Contributions

- We present a simple but effective semi-supervised semantic segmentation approach. Different from previous methods that have complicated and carefully-designed modules, our CPS is model agnostic and simply imposes the consistency between two networks.
- We propose that the cross pseudo supervision (CPS) with the one-hot label is curial for the semantic segmentation task.
- Our model outperforms state-of-the-arts on two public benchmarks Cityscapes and VOC.

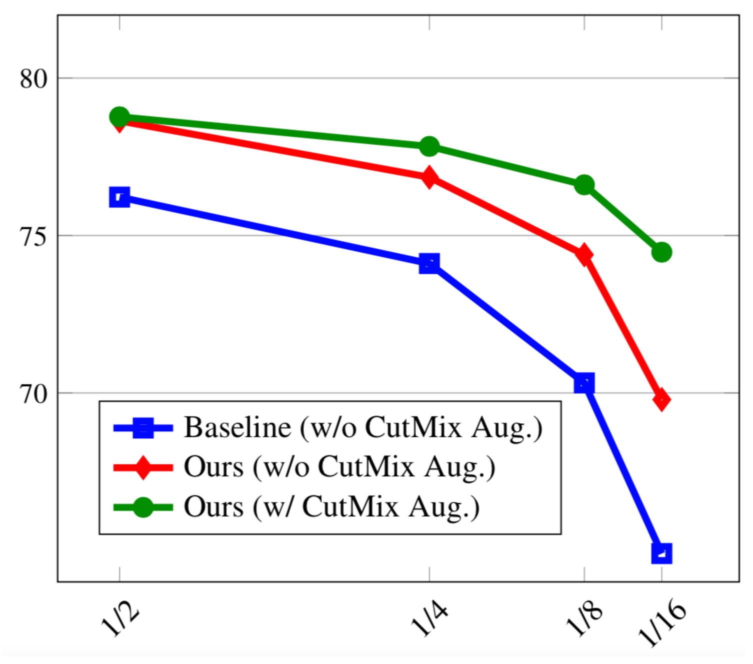
3. Methodology



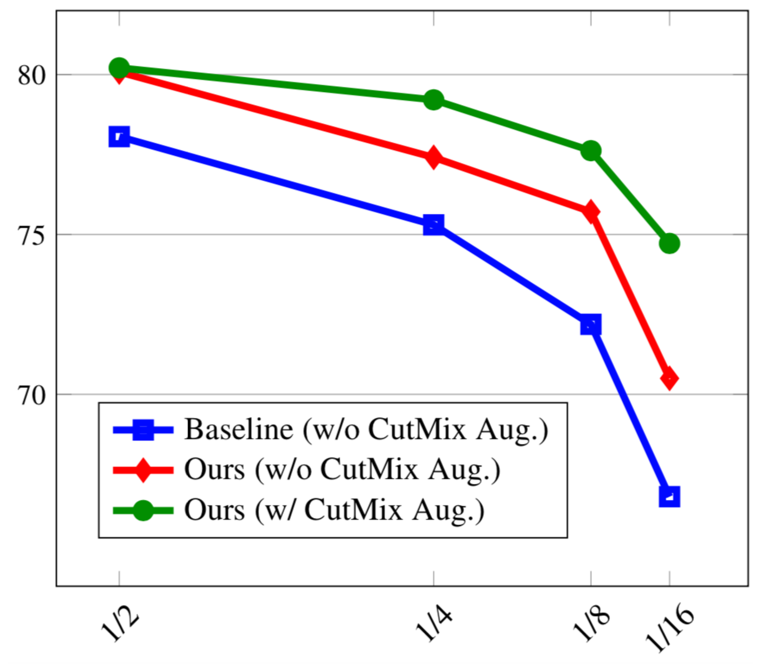
(a) our approach cross pseudo supervision (CPS), (b) cross confidence consistency (CPC)
(c) mean teacher, (d) PseudoSeg
'-->' means forward operation and '//' means stop gradients.

4. Experimental Results

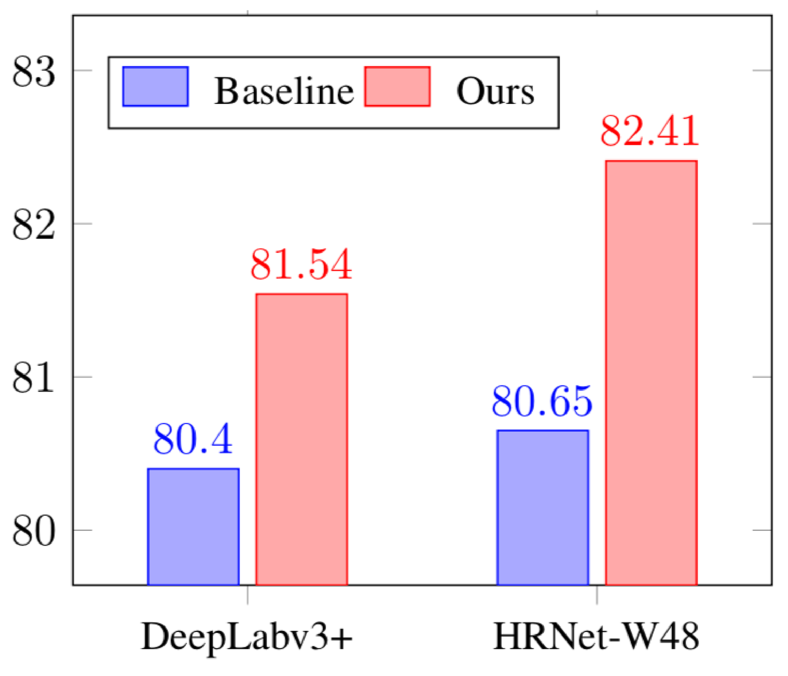
- Comparison with supervised baseline with ResNet-50



- Comparison with supervised baseline with ResNet-101



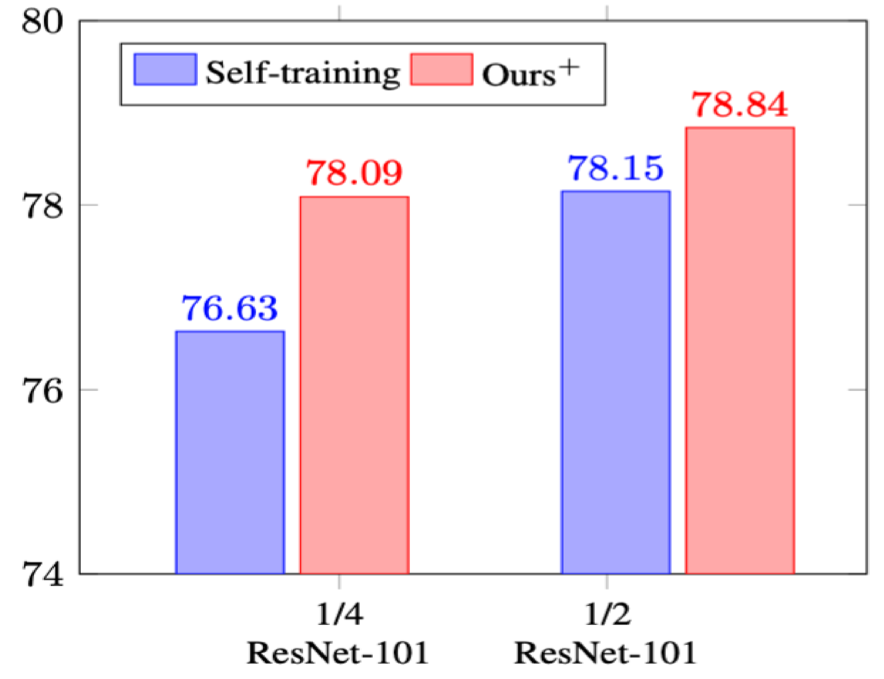
- Improvements over strong baselines on Cityscapes *val* set



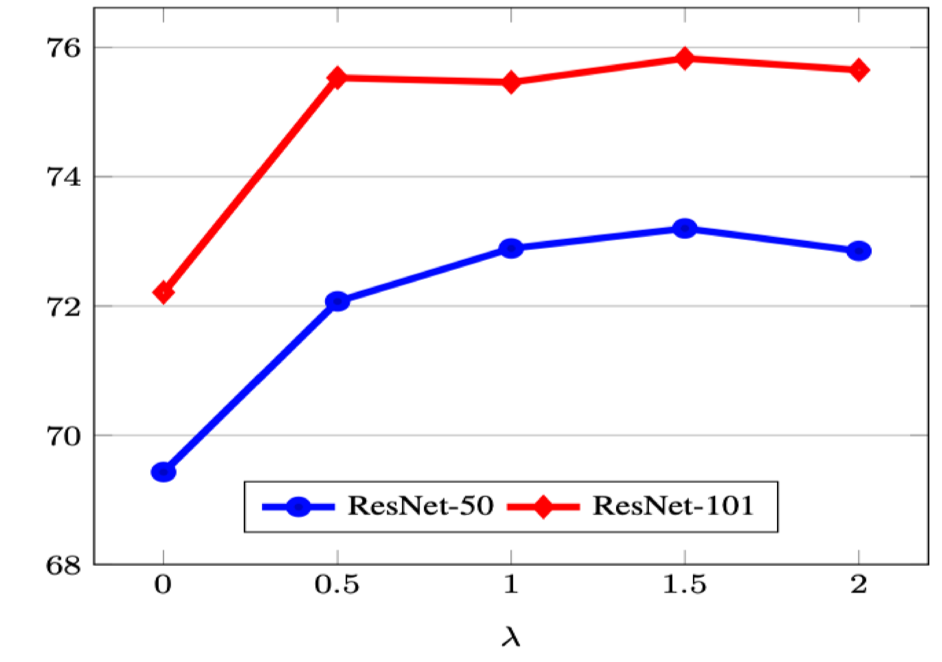
Compare different loss functions: CPS & CPC

Losses					PASCAL VOC 2012		Cityscapes	
$\mathcal{L}_s$	$\mathcal{L}^l_{cps}$	$\mathcal{L}^u_{cps}$	$\mathcal{L}^l_{cpc}$	$\mathcal{L}^u_{cpc}$	ResNet-50	ResNet-101	ResNet-50	ResNet-101
✓					69.43	72.21	70.32	72.19
✓	✓				69.99	72.98	71.73	73.08
✓		✓			73.00	75.83	73.97	75.28
✓	✓	✓			**73.20**	**75.85**	**74.39**	**75.71**
✓			✓	✓	71.23	74.01	72.03	73.77

Comparison with self-training on City



Different weights for CPS on VOC



Comparison with SOTA on PASCAL VOC under different partition protocols

Method	ResNet-50				ResNet-101			
	1/16 (662)	1/8 (1323)	1/4 (2646)	1/2 (5291)	1/16 (662)	1/8 (1323)	1/4 (2646)	1/2 (5291)
MT [33]	66.77	70.78	73.22	75.41	70.59	73.20	76.62	77.61
CCT [27]	65.22	70.87	73.43	74.75	67.94	73.00	76.17	77.56
CutMix-Seg [11]	68.90	70.70	72.46	74.49	72.56	72.69	74.25	75.89
GCT [17]	64.05	70.47	73.45	75.20	69.77	73.30	75.25	77.14
Ours (w/o CutMix Aug.)	68.21	73.20	74.24	75.91	72.18	75.83	77.55	78.64
Ours (w/ CutMix Aug.)	**71.98**	**73.67**	**74.90**	**76.15**	**74.48**	**76.44**	**77.68**	**78.64**

Comparison with SOTA on Cityscapes under different partition protocols

Method	ResNet-50				ResNet-101			
	1/16 (186)	1/8 (372)	1/4 (744)	1/2 (1488)	1/16 (186)	1/8 (372)	1/4 (744)	1/2 (1488)
MT [33]	66.14	72.03	74.47	77.43	68.08	73.71	76.53	78.59
CCT [27]	66.35	72.46	75.68	76.78	69.64	74.48	76.35	78.29
GCT [17]	65.81	71.33	75.30	77.09	66.90	72.96	76.45	78.58
Ours (w/o CutMix Aug.)	69.79	74.39	76.85	78.64	70.50	75.71	77.41	80.08
Ours (w/ CutMix Aug.)	**74.47**	**76.61**	**77.83**	**78.77**	**74.72**	**77.62**	**79.21**	**80.21**