THU-AM-295

Conflict-Based Cross-View Consistency for Semi-Supervised Semantic Segmentation

Zicheng Wang[1,3], Zhen Zhao[1], Xiaoxia Xing[3], Dong Xu[2], Xiangyu Kong[3] and Luping Zhou[1]

[1]University of Sydney
[2]University of Hong Kong
[3]Samsung Research China-Beijing

Brief Introduction

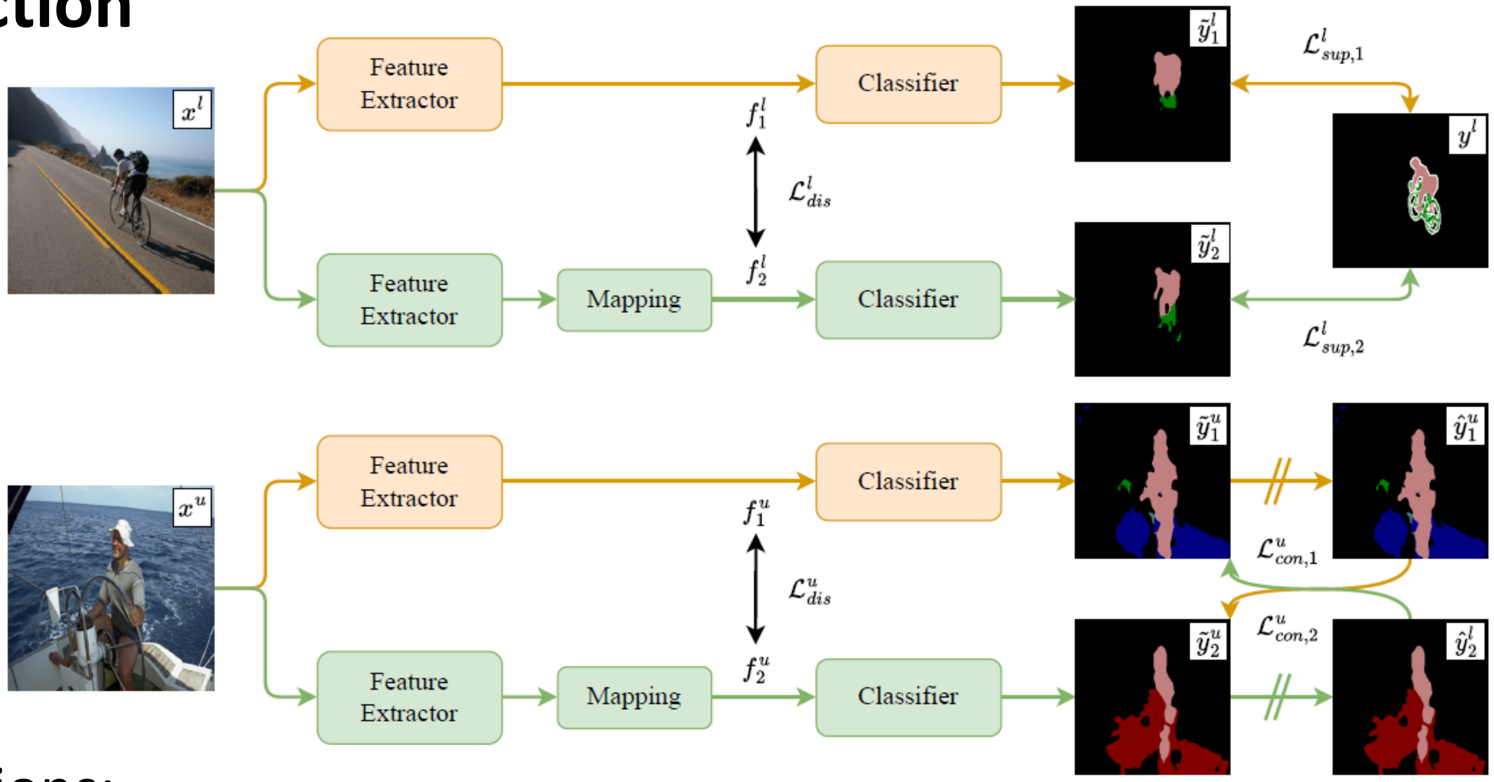
Background:

- Current semi-supervised segmentation methods may suffer from the confirmation bias problem

- The confirmation bias problem can be alleviated by the co-training framework

- The two sub-nets of the co-training framework may step into a collapse

Brief Introduction



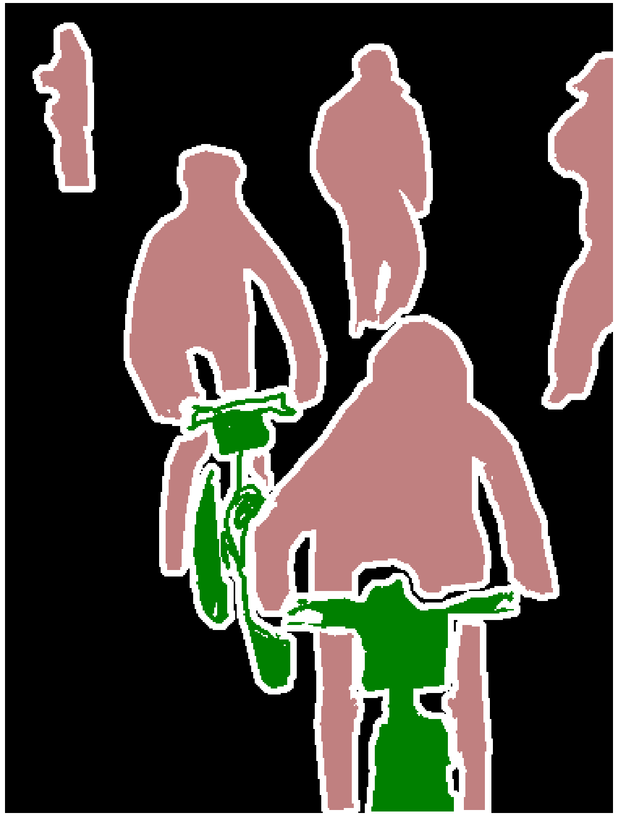
Contributions:

- We propose a feature discrepancy loss to prevent the two sub-nets from collapsing into each other

- We propose a conflict-based pseudo-labeling strategy to encourage the sub-nets learn more useful information from conflicting predictions

Problem Statement

➢ Pixel-wise annotation is extremely expensive



Problem Statement

- Semi-supervised semantic segmentation
  - How to fully take advantage of the unlabeled data ?

labeled data

unlabeled data

Background

- Self-training
  - Train a model on the labeled set
  - Generate pseudo labels for the unlabeled set
  - Re-train the model

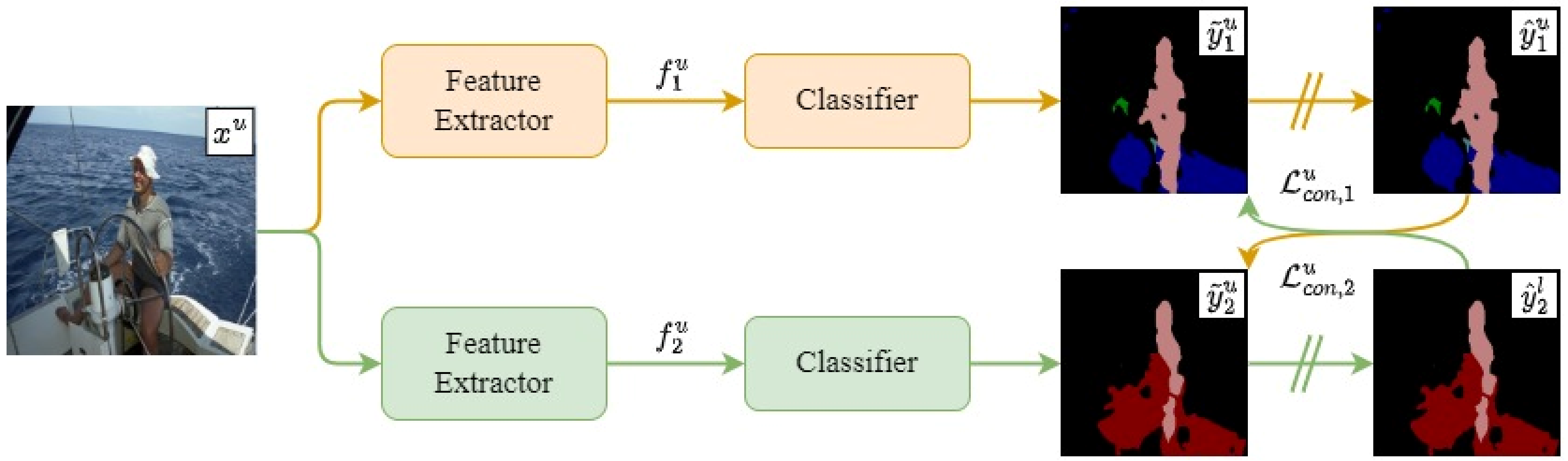
- Consistency regularization
  - Generate perturbed inputs
  - Encourage the model to generate consistent predictions for different inputs

- Confirmation bias problem

Co-training

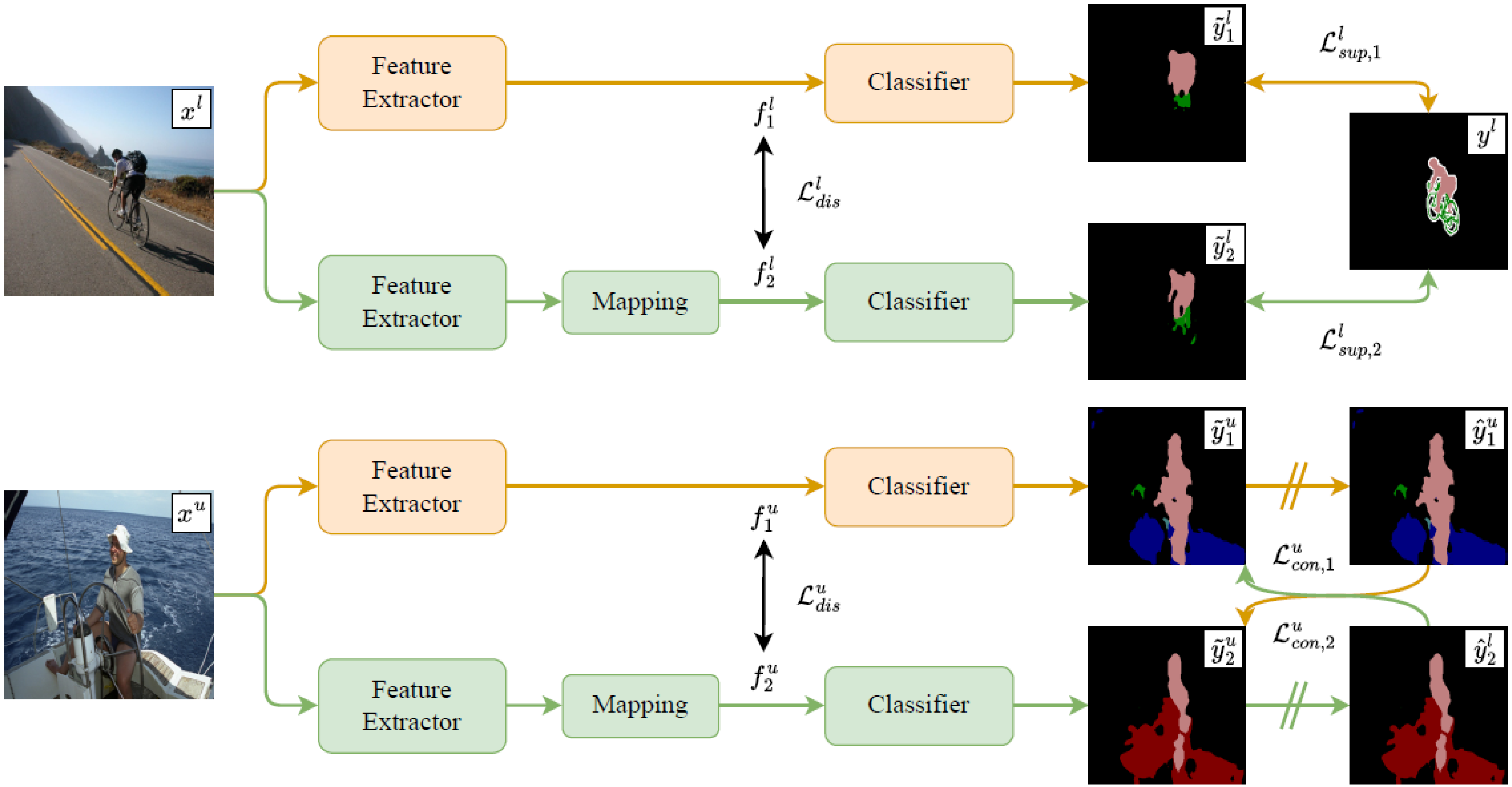
- ➢ Encourage two sub-nets to reason the input from different views
- ➢ enhancing the reliability of the generated pseudo-labels

- different sub-nets may step into a collapse



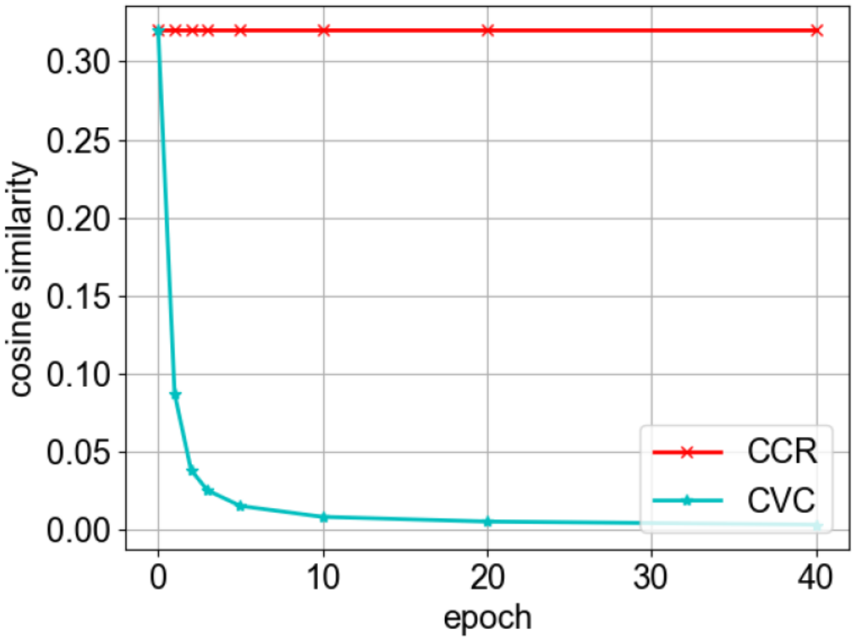
Cross-View Consistency (CVC)

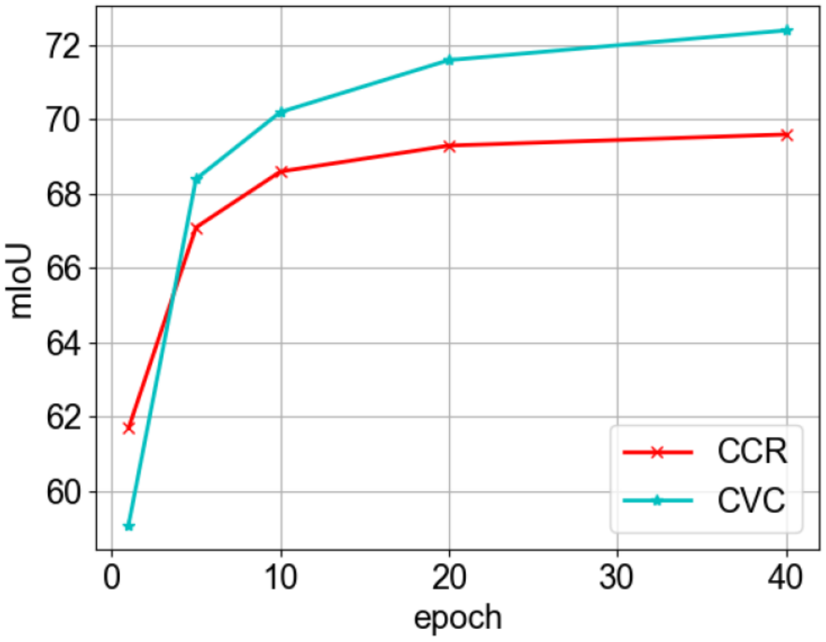
➢ We propose a new feature discrepancy loss to prevent the two sub-nets from collapsing into each other, thus learning distinct features from the same input



Performance

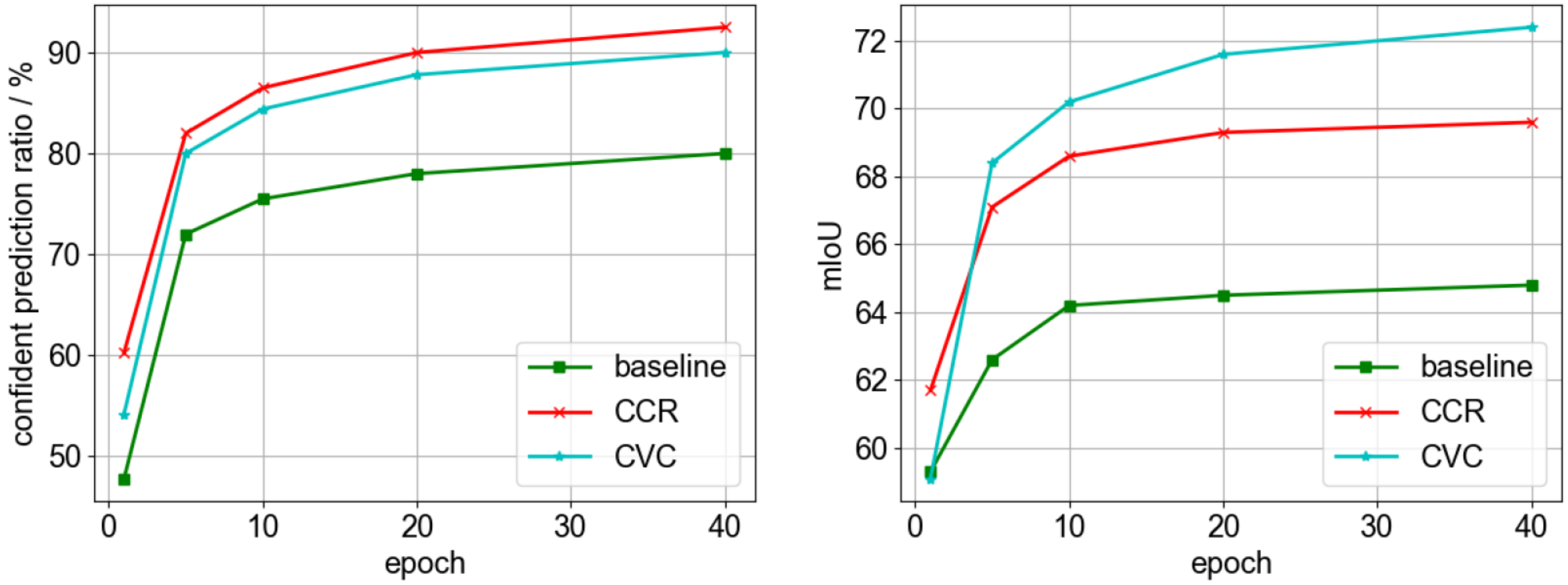
- The cosine similarity between features extracted by the two sub-nets of the traditional cross-consistency regularization (CCR) method keeps a high level
- The cosine similarity between features extracted by the two sub-nets of our cross-view consistency (CVC) keeps a low level





Performance

- CCR will generate more confident predictions, but many predictions are incorrect (confirmation bias problem)
- Our CVC method will generate more correct predictions and can reduce the influence of the confirmation bias problem



Conflict-based Pseudo-Labeling (CPL)

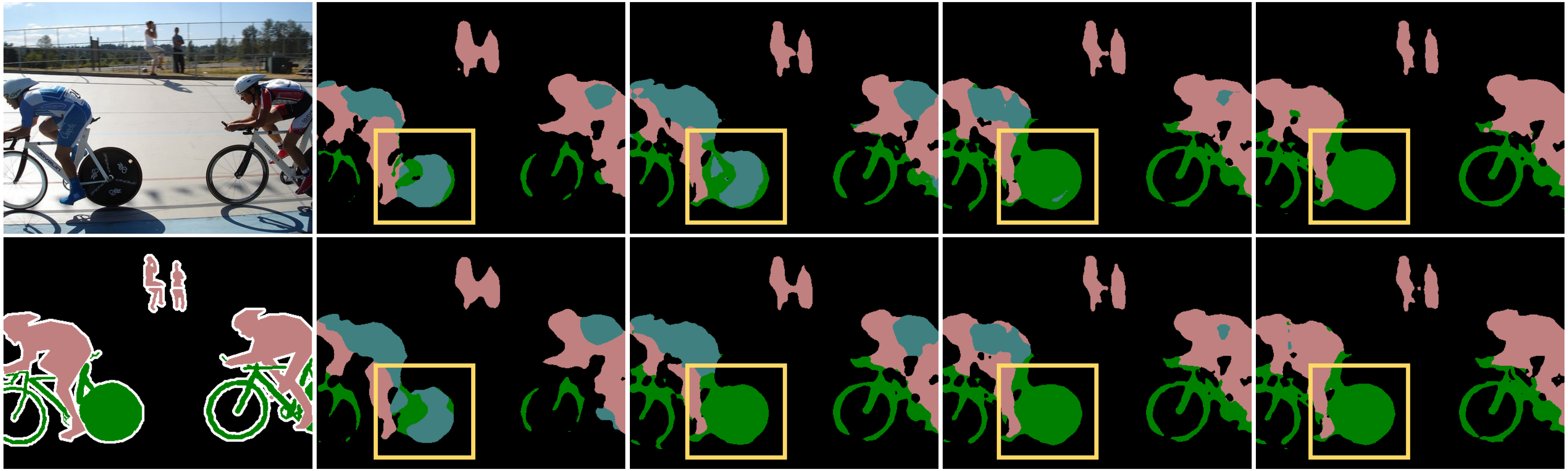
- The feature discrepancy loss might introduce a too strong perturbation
- The training might be unstable

- We enable the sub-nets to learn more useful information from conflicting but confident (cc) predictions
- The sub-nets can generate consistent predictions
- The training would be stable

$$\mathcal{L}_{con,i}^{u} = \omega_c \mathcal{L}_{con,i}^{u,cc} + \mathcal{L}_{con}^{u,e}$$

Performance

- Our CPL method can prevent the two sub-nets from making inconsistent predictions



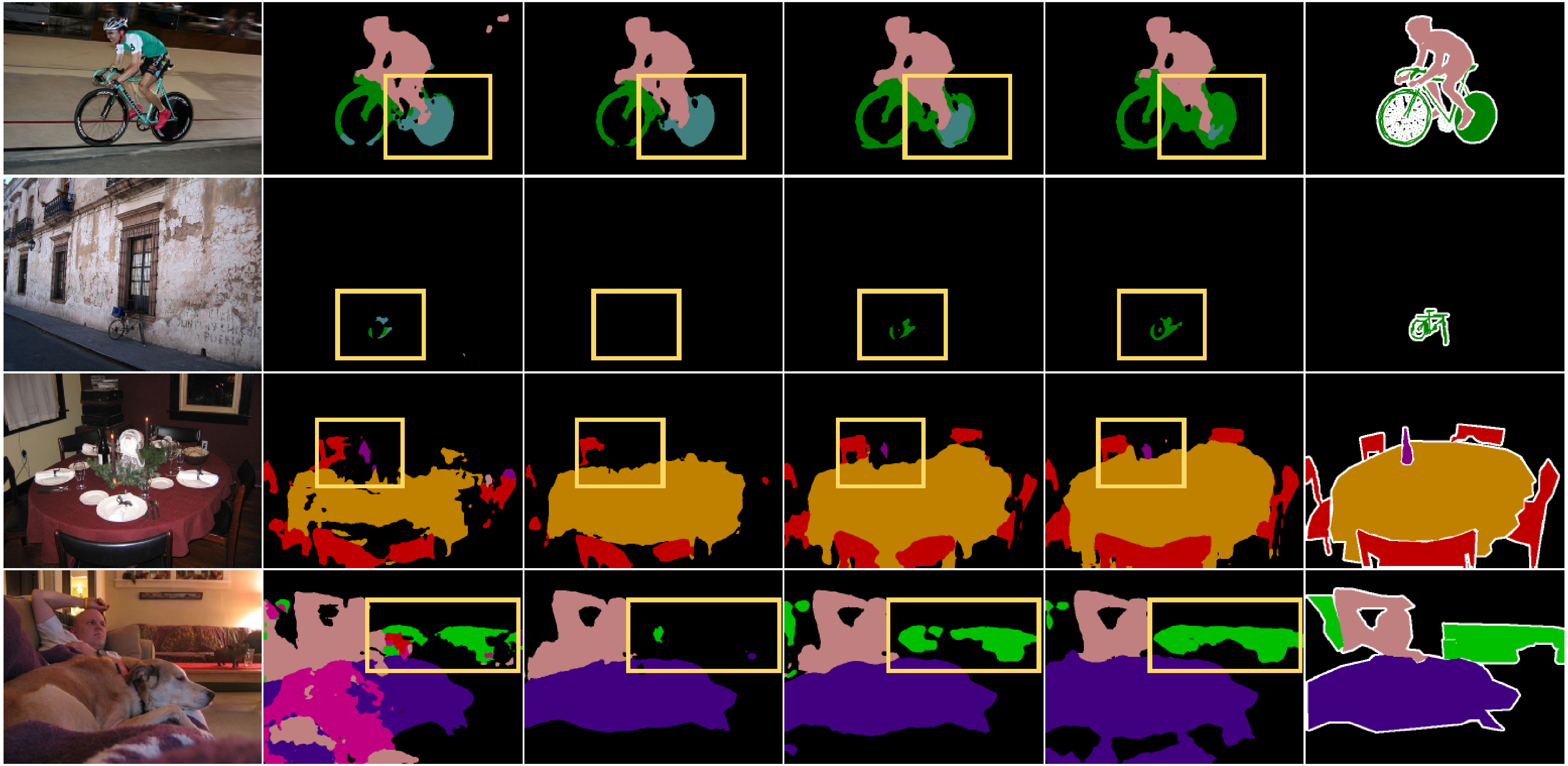
input & gt　　5 epoch　　10 epoch　　20 epoch　　40 epoch

Experiments

- Our CCVC method achieves the new SOTA performance

Dataset	Pascal VOC					CityScapes			
Methods	1/16 (92)	1/8 (183)	1/4 (366)	1/2 (732)	Full (1464)	Methods	1/16 (186)	1/8 (372)	1/4 (744)
Supervised Baseline	45.1	55.3	64.8	69.7	73.5	Supervised Baseline	63.3	65.8	68.4
CutMix-Seg	52.2	63.5	69.5	73.7	76.5	CCT	66.4	72.5	75.7
PseudoSeg	57.6	65.5	69.1	72.4	73.2	GCT	65.8	71.3	75.3
PC^2Seg	57.0	66.3	69.8	73.1	74.2	CPS	69.8	74.3	74.6
CPS	64.1	67.4	71.7	75.9	-	ELN	-	70.3	73.5
ReCo	64.8	72.0	73.1	74.7	-	ST++	-	72.7	73.8
ST++	65.2	71.0	74.6	77.3	79.1	U^2PL	69.0	73.0	76.3
U^2PL	68.0	69.2	73.7	76.2	79.5	USRN	71.2	75.0	-
PS-MT	65.8	69.6	76.6	78.4	80.0	PS-MT	-	75.8	76.9
Ours	**70.2**	**74.4**	**77.4**	**79.1**	**80.5**	Ours	**74.9**	**76.4**	**77.3**

Qualitative Results



inputs | Baseline | CCR | CVC | CCVC | gt

Thank you for listening!