Diverse Cotraining Makes Strong Semi-Supervised Segmentor

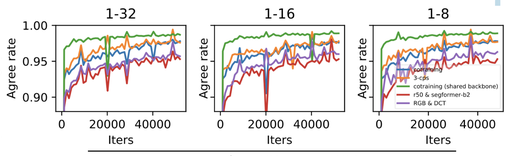
Yijiang Li[1] Xinjiang Wang[2] Lihe Yang[3] Litong Feng[2] Wayne Zhang[2] Ying Gao[4]
[1]Johns Hopkins University [2]SenseTime Research [3]The University of Hong Kong
[4]South China University of Technology

ICCV23
PARIS

Code:
https://github.com/williamium3000/diverse-cotraining
Email:
yli556@jhu.edu
wangxinjiang@sensetime.com

Problem and Motivation

Problem: Current deep co-training violates the core assumption of co-training: multiple compatible and conditionally independent views. Models are tightly coupled together leading to the homogenization of networks and confirmation bias which consequently limits the performance (see left figure).



Method	1/32 (331)	1/16 (662)	1/8 (1323)
Sup Baseline	61.2	67.3	70.8
w/o SA Co-training	**65.66**	**71.28**	**73.77**
w/o SA shared backbone	58.97	65.94	71.25
w/o SA 3-cps	65.41	70.81	72.84
w/ SA Co-training	**70.28**(+4.62)	**73.36**(+2.08)	**74.82**(+1.05)
w/ SA shared backbone	69.48	70.16	73.47
w/ SA 3-cps	69.68	71.83	74.36

Homogenization problem (measured by agree rate) and the corresponding performance.

Theoretical Framework:

Definition 1 *Definition of homogenization:*

$$H = Pr_{x\in D}[f_1(x) = f_2(x)] = \frac{1}{HW}\sum_{i=1}^{HW} 1(p_{1i} = p_{2i})$$

Theorem 1 *Given hypothesis class $\mathcal{H}$ and labeled data set D_l of size l that are sufficient to learn an initial segmentor f_i^0 with an upper bound of the generalization error of b_i^0 with probability δ ($l \geq \max\{\frac{1}{b_i^0}\ln\frac{|\mathcal{H}|}{\delta}\}$), we use empirical risk minimization to train f_i^0 on the combination of labeled and unlabeled set σ^i where pseudo label are provided by the other model f_{3-i}^0. Then we have*

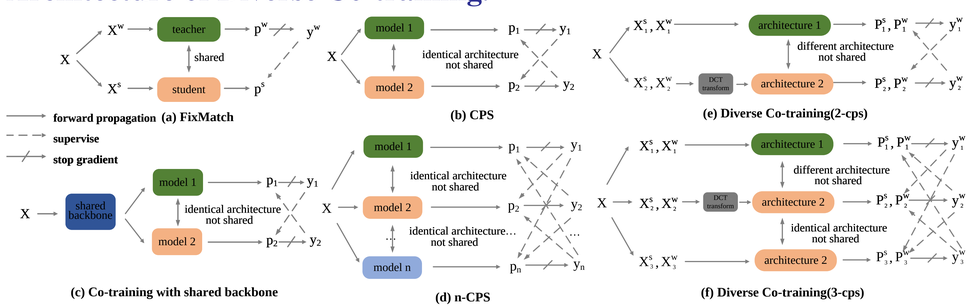
$$Pr\left[d(f_i^k, f*) \geq b_i^k\right] \leq \delta$$

if $lb_i^0 \leq e\sqrt[M]{M!} - M$, where $M = ub_{3-i}^0$ and $b_i^k = \max\{\frac{lb_i^0 + ub_{3-i}^0 - ud(f_{3-i}^{k-1}, f_i^k)}{l}, 0\}$.

Diverse Co-training

- Diverse strong augmentations provide different views.
- Different input domains as pseudo views. We propose to use DCT space as pseudo views of RGB space.
- Diverse architecture provides different inductive biases

Architecture of Diverse Co-training.



Illustrating the architectures for (a) FixMatch, (b) CPS, (c) cross heads with shared backbone, (d) n-CPS, (e) *Diverse Co-training (2-cps)* and (f) *Diverse Co-training (3-cps)*.

Experiments & Results

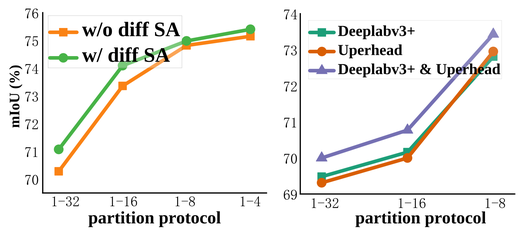
Analysis on How to Promote Diversity

	Backbone	1/32 (331)	1/16 (662)
w/o SA	R50	65.66	71.28
	mit-b2	71.01	74.53
	R50 & mit-b2	**71.58** / 71.03	**74.94** / 74.84
w/ SA	R50	70.28	73.36
	mit-b2	74.51	75.29
	R50 & mit-b2	74.85 / **74.87**	75.12 / **75.85**
	ResNeSt50	70.92	75.58
	ResNeXt50	71.18	72.77
	R50 & ResNeST50	72.70 / **73.56**	73.41 / **75.65**
	R50 & ResNeXT50	72.15 / 72.39	74.41 / 74.56

Different Architecture

	Input Domain	1/32 (331)	1/16 (662)
w/o SA	RGB	65.66	71.28
	DCT	65.33	67.37
	RGB & DCT	69.45 / 69.03	**72.46** / 72.03
	RGB & HSV	**69.65** / 67.05	71.74 / 69.89
w/ SA	RGB	70.28	73.36
	DCT	70.65	73.26
	RGB & DCT	71.88 / **72.00**	**74.10** / 73.94
	RGB & HSV	70.40 / 68.30	72.64 / 70.91

Different Input Space



(a) Different strong augmentation (SA) (b) Different head architecture

Different augmentation and seg head

Comparison with state-of-the-arts on Pascal VOC and Cityscapes.

Method	Resolution	92	183	366	732	1464
ResNet50						
Sup Baseline	513x513	39.1	51.3	60.3	65.9	71.0
PseudoSeg [103]	512x512	54.9	61.9	64.9	70.4	-
PC^{2}Seg [100]	512x512	56.9	64.6	67.6	70.9	-
Ours (2-cps)	513x513	71.8	74.5	**77.6**	78.6	79.8
Ours (3-cps)	513x513	**73.1**	**74.7**	77.1	**78.8**	**80.2**
ResNet101						
Sup Baseline	321x321	44.4	54.0	63.4	67.2	71.8
ReCo [49]	321x321	64.8	72.0	73.1	74.7	-
ST++ [91]	321x321	65.2	71.0	74.6	77.3	79.1
ours (2-cps)	321x321	74.8	**77.6**	79.5	80.3	**81.7**
ours (3-cps)	321x321	**75.4**	76.8	**79.6**	**80.4**	81.6
Sup Baseline	512x512	42.3	56.6	64.2	68.1	72.0
MT [72]	512x512	48.7	55.8	63.0	69.16	-
CPS[12]	512x512	64.1	67.4	71.7	75.9	-
U^{2}PL [82]	512x512	68.0	69.2	73.7	76.2	79.5
PS-MT [50]	512x512	65.8	69.6	76.6	78.4	80.0
ours (2-cps)	513x513	**76.2**	76.6	**80.2**	80.8	81.9
ours (3-cps)	513x513	75.7	**77.7**	80.1	**80.9**	**82.0**

Comparison with state-of-the-art methods on the Pascal dataset. Labeled images are from the high-quality training set.

Method	Resolution	1/32 (331)	1/16 (662)	1/8 (1323)	1/4 (2646)
Sup Baseline	321x321	55.8	60.3	66.8	71.3
CAC[43]	320x320	-	70.1	72.4	74.0
ST++[91]	321x321	-	72.6	74.4	75.4
Ours (2-cps)	321x321	**75.2**	76.0	76.2	76.5
Ours (3-cps)	321x321	74.9	**76.4**	**76.3**	**76.6**
Sup Baseline	513x513	54.1	60.7	67.7	71.9
CPS[12]	512x512	-	72.0	73.7	74.9
3-CPS [20]	512x512	-	72.0	74.2	75.9
ELN [42]	512x512	-	-	73.2	74.6
PS-MT [50]	512x512	-	72.8	75.7	76.4
U^{2}PL* [82]	513x513	-	72.0	75.1	76.2
Ours (2-cps)	513x513	**75.2**	76.2	77.0	77.5
Ours (3-cps)	513x513	74.7	**76.3**	**77.2**	**77.7**

Comparison with ResNet50 on the Pascal VOC 2012 dataset. Labeled images are sampled from the blended training set.

Method	ResNet50 1/30 (100)	1/8 (372)	1/4 (744)	Method	ResNet101 1/16 (186)	1/8 (372)	1/4 (744)
Sup Baseline	54.8	70.2	73.6	Sup Baseline	66.8	72.5	76.4
CAC [43]	-	69.7	72.7	CutMix [23]	67.9	73.5	75.4
CPS [12]	-	74.4	76.9	CPS [12]	70.5	75.7	77.4
ST++ [91]	61.4	72.7	73.8	U^{2}PL [82]	74.9	76.5	78.5
U^{2}PL* [82]	59.8	73.0	76.3	PS-MT [50]	-	76.9	77.6
Ours (2-cps)	64.5	76.3	77.1	Ours (2-cps)	75.0	77.3	**78.7**
Ours (3-cps)	**65.5**	**76.5**	**77.9**	Ours (3-cps)	**75.7**	**77.4**	78.5

Comparison with state-of-the-art methods on the Cityscapes dataset

Conclusion

- We theoretically prove that the homogenization of networks accounts for the generalization error of co-training training and discover the lack of diversity in current co-training methods that violate the assumptions.
- We comprehensively explore the different dimensions of co-training to promote diversity including input domains, augmentations and architectures and demonstrate the significance of diversity in co-training.
- We propose a holistic framework combining the above three techniques to increase diversity and discuss two variants with high empirical performances.