Supplemental material: Improving Semi-supervised Semantic Segmentation via Strong-weak Dual-branch Network

Wenfeng Luo[1] and Meng Yang*[1,2]

[1] School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China
luowf5@mail2.sysu.edu.cn
[2] Key Laboratory of Machine Intelligence and Advanced Computing (SYSU), Ministry of Education yangm6@mail.sysu.edu.cn
*Corresponding author

1 Oversampling Doesn't Help with Single-branch Network

One might argue oversampling or weighted loss could be a solution. To this, we conduct experiments via oversampling the strong annotations. As shown in table 1, oversampling improves the final segmentation accuracy steadily as more strong annotations are duplicated, but it still fails to outperform the result using only the strong annotations. We argue that single-branch network in its nature is incapable of handling the inconsistency due to the competing signals sent by the strong and weak annotations.

Table 1. Segmentation accuracy concerning different oversampling rates on PASCAL VOC *val* set.

Training data	mIoU (%)
1.4k strong + 9k weak	62.8
1.4k*2 strong + 9k weak	63.5
1.4k*3 strong + 9k weak	64.2
1.4k*6 strong + 9k weak	65.9
1.4k strong	**68.9**

2 Training Strategy

Another simple strategy would be to first train the segmentation network on the weak data and then finetune it on the strong ones. However, the final result is better but only converges to the result (68.9%) achieved by using the strong ones, which means the weak annotations make no contribution.

3 Regularization between Branches

As in many semi-supervised approaches, a regularization constraint may be introduced upon the strong and weak predictions. Similarly, we also try to introduce a affinity constraint via the Kullback–Leibler divergence between the strong and weak probabilities. Mathematically, the constraint loss is defined as:

$$\mathcal{L}_{kl}(s^s, s^w) = -\frac{1}{|s^s|} \sum_{u,c} s^s_{u,c} \log \frac{s^w_{u,c}}{s^s_{u,c}} \tag{1}$$

where u denotes pixel location and c denotes class. So the overall training loss is:

$$\mathcal{L}_{total} = \mathcal{L}_{data} + \mathcal{L}_{kl} \tag{2}$$

Table 2. Ablation experiments concerning network architectures and training data. Rows marked with "*" are results from the proposed dual-branch network and others are from the single-branch network.

Backbone	**Strong branch**	**Weak branch**	**mIoU**
VGG16	10k weak	-	57.0
VGG16	10k weak (retrain)	-	60.1
VGG16	1.4k strong + 9k weak	-	62.8
VGG16	1.4k strong	-	68.9
VGG16	10k strong	-	71.4
VGG16*	1.4k strong	1.4k strong + 9k weak	72.2
VGG16+KL*	1.4k strong	1.4k strong + 9k weak	71.6
VGG16*	1.4k strong	10k strong	73.9

Interestingly, the affinity constraint via the KL divergence indeed downgrades the performance slightly by 0.6%, which implicitly demonstrates the importance of imposing separate treatment upon the strong and weak data.

Table 3. Segmentation accuracy (%) using different n's and network branches to generate predictions. Results under "strong branch" column means we use the strong branch for prediction and likewise for the "weak branch".

n	mIoU	
	Strong branch	Weak branch
0	70.9	62.3
1	71.5	63.1
2	72.2	63.8
3	72.2	64.0

4 Ablation Studies concerning Hyper-parameter n

The hyper-parameter n controls the number of shared layers besides the underlying backbone. In this experiment, we report the performance on both the strong and weak branches. As shown in table 3, the segmentation accuracy of the strong branch is robust to different choice of n's and is consistently better than combining the result from two separate networks. The performance on the weak branch increases steadily as n becomes bigger. In short, both the strong and weak branches gain performance boost over simply training them separately. We use $n = 3$ in the main experiments.

5 Extra Ablation Study concerning the Data Portion

We miss a closely related paper of the decoupled network in [1], which addressed the same problem setup based on a similar idea of decoupling the losses for fully- and weakly-labeled data. We want to point out the main difference between our approach and the related work [1]: Our instantiation of the dual-branch framework allows the network to learn directly from the labels (though noisy in its nature), which provide stronger gradient signal than the back-propagated gradient maps in [1]. It utilized back-propagation to retrieve class-specific information for the downstream segmentation network, which was only supervised by the training data with strong annotations and thus gained no benefit from the weak ones. So it actually underused the available weak annotations in the training of segmentation network.

Nevertheless, we provide extra experiments to see the potential of our model under different portion of strong and weak data. As shown in table 4, our approach achieves much better result than [1] under the same setting. We believe it's because: 1) DSRG already achieves excellent performance so the proxy ground truth provides superior supervision than the image-level labels; 2) Our dual-branch model helps mitigate the sample imbalance and annotation inconsistency between the strong and weak labels.

Table 4. Segmentation accuracy (%) using different portion of strong and weak annotations.

#strong per class	[1]	Ours
5	53.1	60.8
10	57.4	64.8
25	62.1	68.1

References

1. Hong, S., Noh, H., Han, B.: Decoupled deep neural network for semi-supervised semantic segmentation. In: NIPS (2015)