Budget-Aware Semi-Supervised Semantic and Instance Segmentation

Miriam Bellver, Amaia Salvador, Jordi Torres, Xavier Giro-i-Nieto

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

WiCV Women in Computer Vision
DeepVision Workshop

Check out our paper!

Introduction

Methods that move towards less supervised scenarios are key for image segmentation, as dense labels demand significant human intervention. Generally, the annotation burden is mitigated by labeling datasets with weaker forms of supervision, e.g. image-level labels or bounding boxes. Another option are semi-supervised settings, that commonly leverage a few strong annotations and a huge number of unlabeled/weakly-labeled data. In this work, we **revisit semi-supervised segmentation schemes and narrow down significantly the annotation budget** (in terms of total labeling time of the training set) compared to previous approaches. Our contributions are the following:

- We **unify the segmentation benchmarks** regardless of the training setting and the supervision signals.
- We **outperform previous semi-supervised semantic segmentation** methods at low annotation budgets for Pascal VOC benchmark and **present first results or semi-supervised instance segmentation** for the same benchmark.
- At fixed low annotation budgets, **having fewer but stronger-labeled data** leads to better performance.

Benchmark for budget-aware segmentation

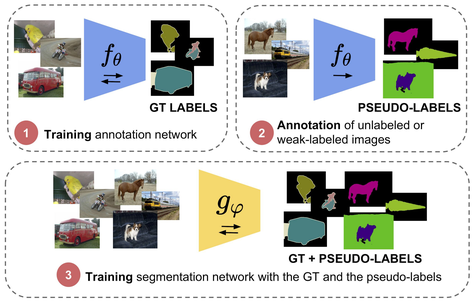
Following the analysis that **Bearman et al. 2016** did for semantic segmentation, we will compare the different semi-supervised and weakly-supervised methods for both semantic and instance segmentation in terms of annotation cost. The costs for each kind of annotation for the **Pascal VOC** dataset are obtained from the literature and can be seen in the following Table:

	IL	IL+C	Full	BB
Cost (s/image)	20	22.22	239.7	38.1

IL: Image-level labels
IL+C: Image-level + counts labels
Full: Dense labels
BB: Bounding box labels

Semi-supervised pipeline

We define a semi-supervised pipeline that consists of two networks. A first **annotation network** trained with a few strongly-annotated samples, that will be used to predict a pseudo-annotation for the unlabeled/weakly-labeled samples. We finally train a second **segmentation network** with the strong labels and the pseudo-annotations.



Semantic segmentation

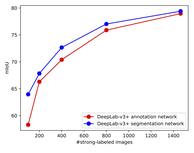
We first test our pipeline for semantic segmentation. We choose **DeepLab v3+ (Chen et al., 2018)** model for both the annotation and segmentation network. We perform our experiments on Pascal VOC dataset. The standard semi-supervised benchmark of Pascal VOC consists on 1.4k strongly-labeled images and 9k unlabeled samples.

Performance on Pascal VOC

	#Strong	#Unlabeled	val mIoU	test mIoU
DeepLab-v3+ Ours	∼1.4k		78.96	77.26
DeepLab-v3+ Ours	∼1.4k	∼9k	79.41	78.71
DeepLab-v3+ Ours	∼10k		80.42	80.29
DeepLab-v3+ [4]	∼10k		81.21	-

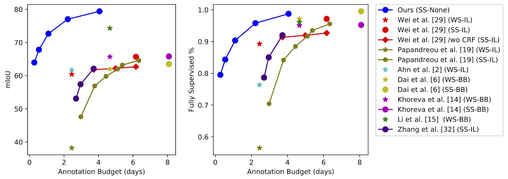
When we use the pseudo-annotations obtained by the annotation network, performance increases for all annotation budgets.

mIOU vs. #strongly-labeled samples



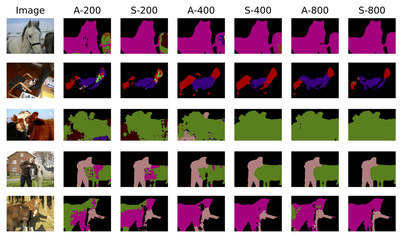
Comparison to state-of-the-art

In the following Figure we compare to other works doing semi-supervised or weakly-supervised semantic segmentation in relation to annotation budget. Our method **outperforms all previous approaches**:



SS: Semi-supervised
WS: Weakly-supervised
IL: Image-label supervision
BB: Bounding box labels supervision

Qualitative results



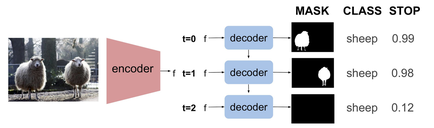
In this figure we can observe the difference between the annotation (A-) and the segmentation (S-) network performance for different annotation budgets (200, 400 or 800 strongly-annotated samples).

Instance segmentation

Strongly and unlabeled samples

We test our pipeline for instance segmentation. We choose **RSIS model (Salvador et al., 2018)** for both the annotation and segmentation network, and also evaluate the performance for Pascal VOC benchmark.

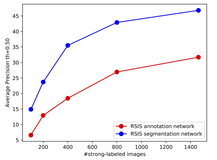
Recurrent Semantic Instance Segmentation (RSIS)



Performance on Pascal VOC

	#Strong	#Unlabeled	val AP 50
RSIS Ours	∼1.4k		31.7
RSIS Ours	∼1.4k	∼9k	46.8
RSIS Ours	∼10k		56.4
RSIS [26]	∼10k		57.0

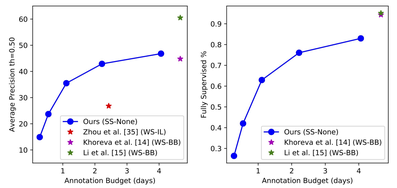
mAP vs. #strongly-labeled samples



For the instance segmentation case, using the pseudo-annotations obtained by the annotation network leads to a significant improvement of performance.

Comparison to state-of-the-art

In the following Figure we compare to other works that tackle instance segmentation with low annotation budgets. We outperform the state-of-the-art at low annotation budgets, **Zhou et al, 2018**, by a remarkable margin. Thus, the semi-supervised setup leads to better results than the weakly-supervised pipeline given a low annotation budget.



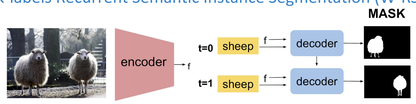
SS: Semi-supervised
WS: Weakly-supervised
IL: Image-label supervision
BB: Bounding box labels supervision

Strongly and weakly-labeled samples

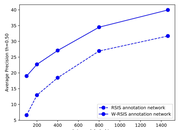
We also guide experiments with strongly-labeled data combined with weakly-labeled data. As weakly-labeled samples we choose i**mage-level labels plus counts of how many objects appear in the image (IL+C).** Now we should consider the annotation cost of these weak labels together with the strong ones.

For the annotation network, we modify the baseline RSIS model to benefit from the weak labels information. The **weak labels are an input** of the network.

Weak-labels Recurrent Semantic Instance Segmentation (W-RSIS)



W-RSIS vs. RSIS annotation networks



Feeding the class labels of the different objects in the network leads to significant better performance.

W-RSIS vs. RSIS annotation + RSIS segmentation networks

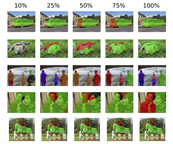
	#Strong	#Unlabeled	#Weak	Budget	AP 50
RSIS	100	10482	-	0.27	14.9
RSIS	200	10382	-	0.55	23.7
W-RSIS	100	-	912	0.51	25.2
RSIS	400	10182	-	1.1	35.5
W-RSIS	200	-	2279	1.14	30.8
RSIS	800	9782	-	2.21	42.9
W-RSIS	200	-	6838	2.31	32.7
Zhou et al. [35]	-	-	10582	2.43	26.8

The configurations that use W-RSIS reach better performance at very low budgets.

Qualitative results

W-RSIS vs. RSIS pseudo-annotations

Increasing weak labels with fixed strong labels



Examples of typical failures of RSIS

Bearman, A., Russakovsky, O., Ferrari, V., & Fei-Fei, L. (2016, October). What's the point: Semantic segmentation with point supervision. In ECCV 2016.
Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In ECCV 2018.
Salvador, A., Bellver, M., Campos, V., Baradad, M., Marques, F., Torres, J., & Giro-i-Nieto, X. (2017). Recurrent neural networks for semantic instance segmentation. In DeepVision Workshop CVPR 2018.
Zhou, Y., Zhu, Y., Ye, Q., Qiu, Q., & Jiao, J. (2018). Weakly supervised instance segmentation using class peak response. In CVPR 2018.

Acknowledgments:

EXCELENCIA SEVERO OCHOA
Obra Social Fundació "la Caixa"
NVIDIA
Generalitat de Catalunya