PseudoSeg: Designing Pseudo Labels for Semantic Segmentation

Yuliang Zou [1] Zizhao Zhang [2] Han Zhang [3] Chun-Liang Li [2]
Xiao Bian [2] Jia-Bin Huang [1] Tomas Pfister [2]

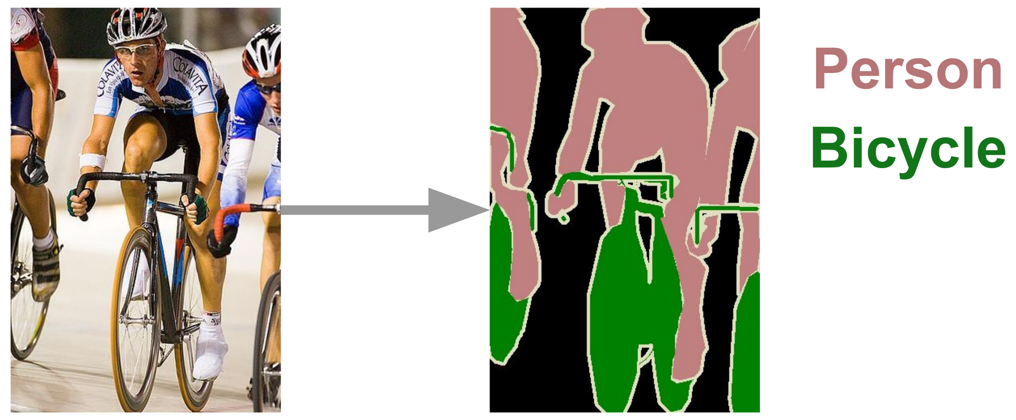
[1] Virginia Tech [2] Google Cloud AI [3] Google Brain

Code available at: https://yuliang.vision/pseudo_seg

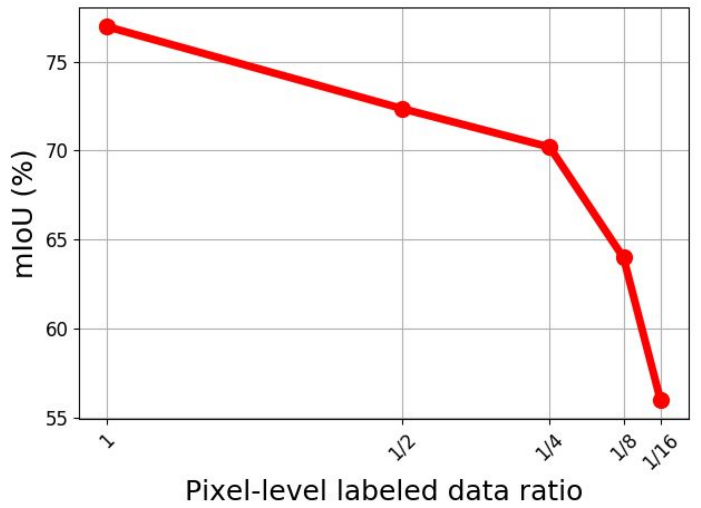
Fork me on GitHub

Motivation

Semantic Segmentation



Input Output



- Requiring a large amount of dense labels -> expensive
- Performance decreases quickly as we reduce the number of labels
- We want to improve the data efficiency of semantic segmentation

Unlabeled/Weakly-labeled Data to the Rescue

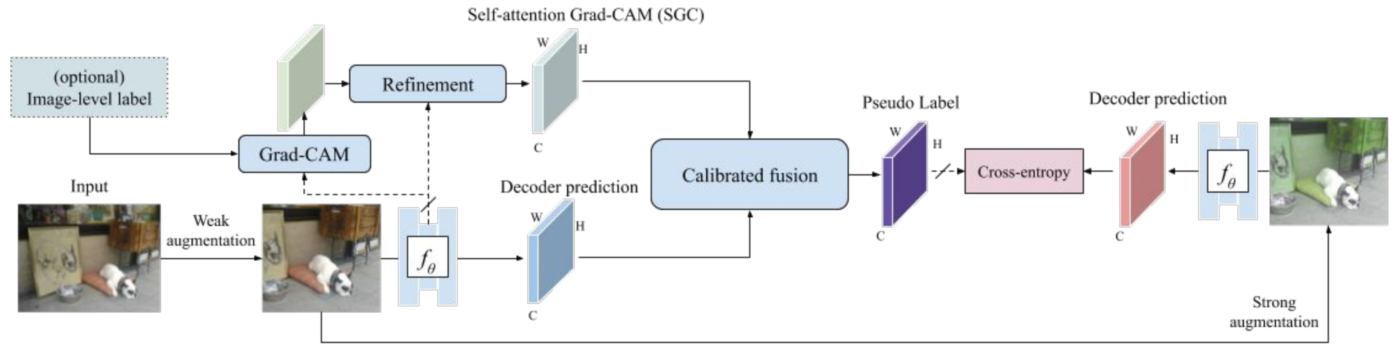
Fully-labeled data (limited)

Unlabeled data (a large amount)

Weakly-labeled data (a lot)

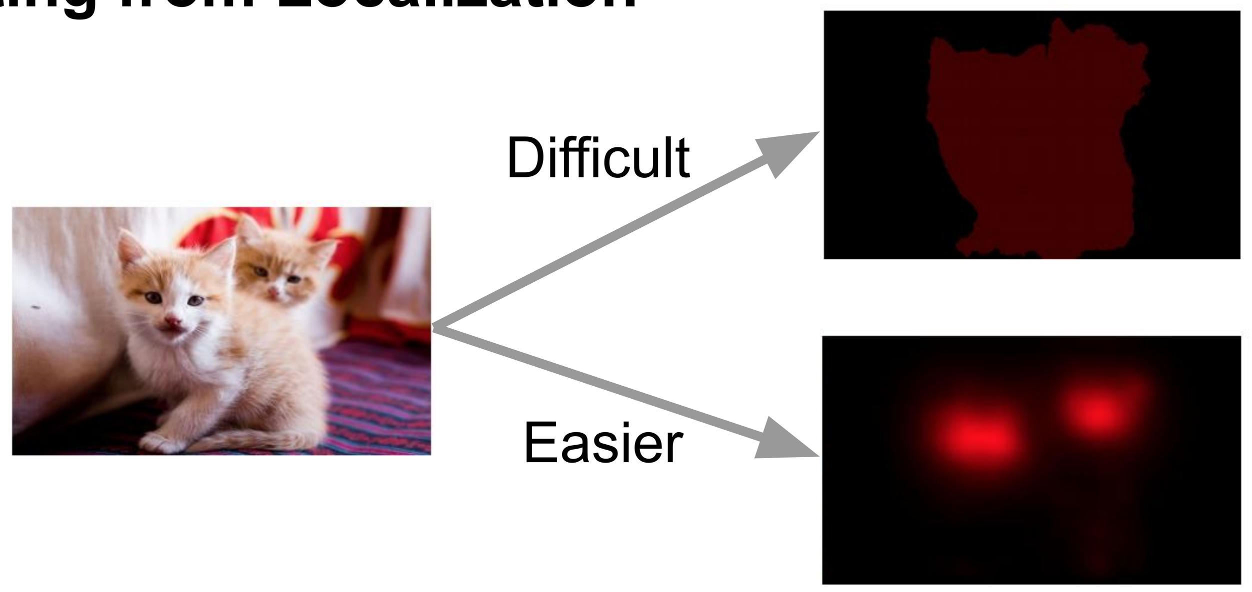
Designing Pseudo Labels for Semantic Segmentation

Overview



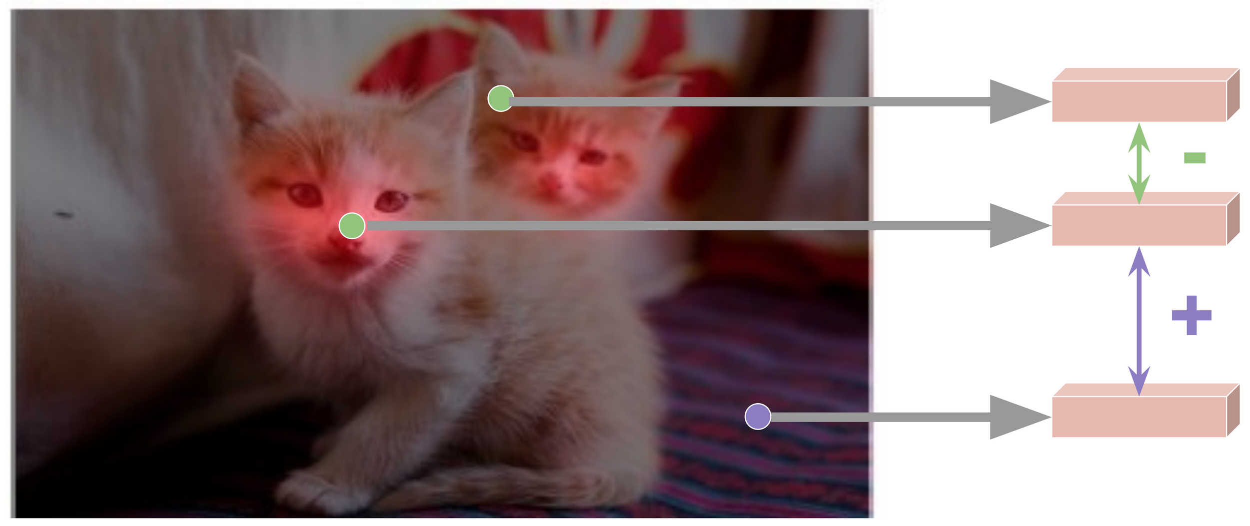
- Strong-weak dual branch consistency-based training
- Different decision mechanisms for distinct predictions
- Wisely fusing the predictions to construct well-calibrated pseudo labels

Starting from Localization



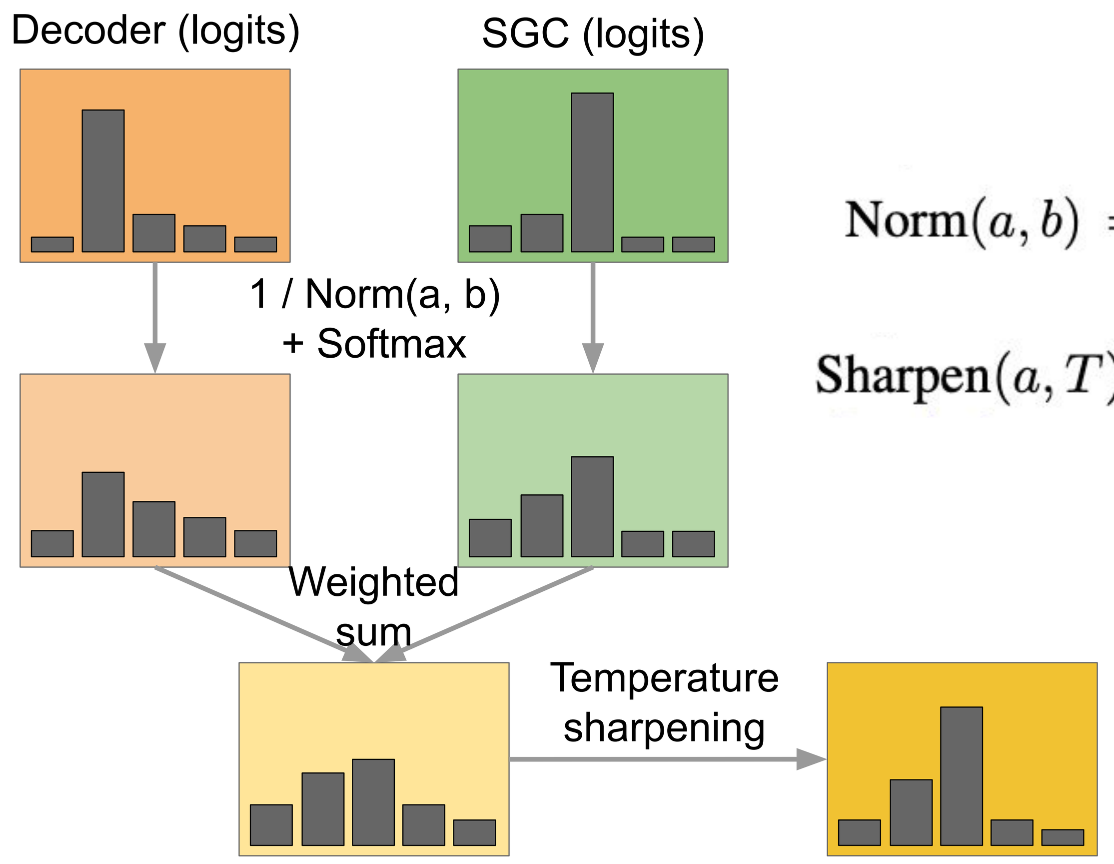
- Hard to get precise segmentation prediction in low-data regime
- Easier to get coarse localization of objects using Grad-CAM

From Localization to Segmentation



- Grad-CAM only localizes partial region of interest
- Propagating Grad-CAM scores with learned feature similarity
- Implemented with dot-product self-attention operation -> Self-attention Grad-CAM (SGC)

Calibrated Prediction Fusion



$$\mathrm{Norm}(a,b) = \sqrt{\sum_i^{|a|}(a_i^2 + b_i^2)}$$

$$\mathrm{Sharpen}(a,T)_i = a_i^{1/T} / \sum_j^C a_j^{1/T}$$

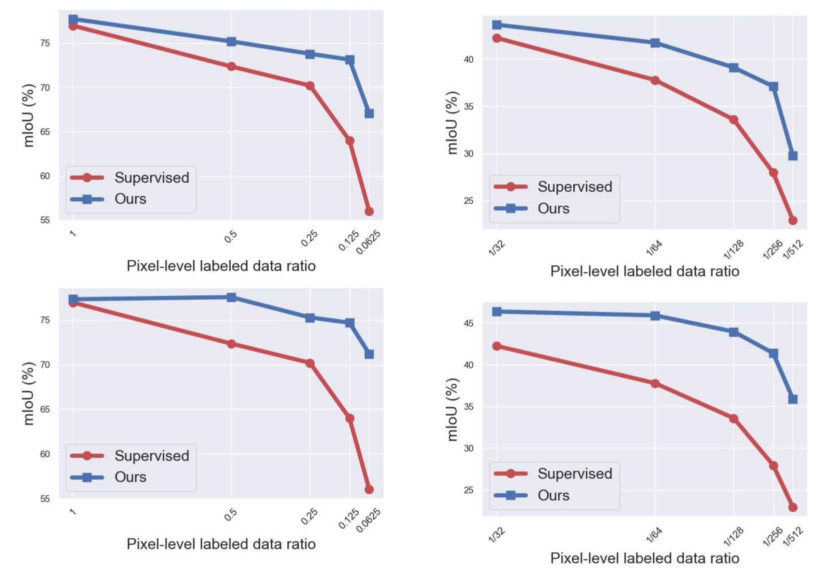
- Constructing well-calibrated pseudo labels from two predictions

Experimental Results

Improvement over Supervised Baseline

Fully-labeled and unlabeled

Fully-labeled and weakly-labeled



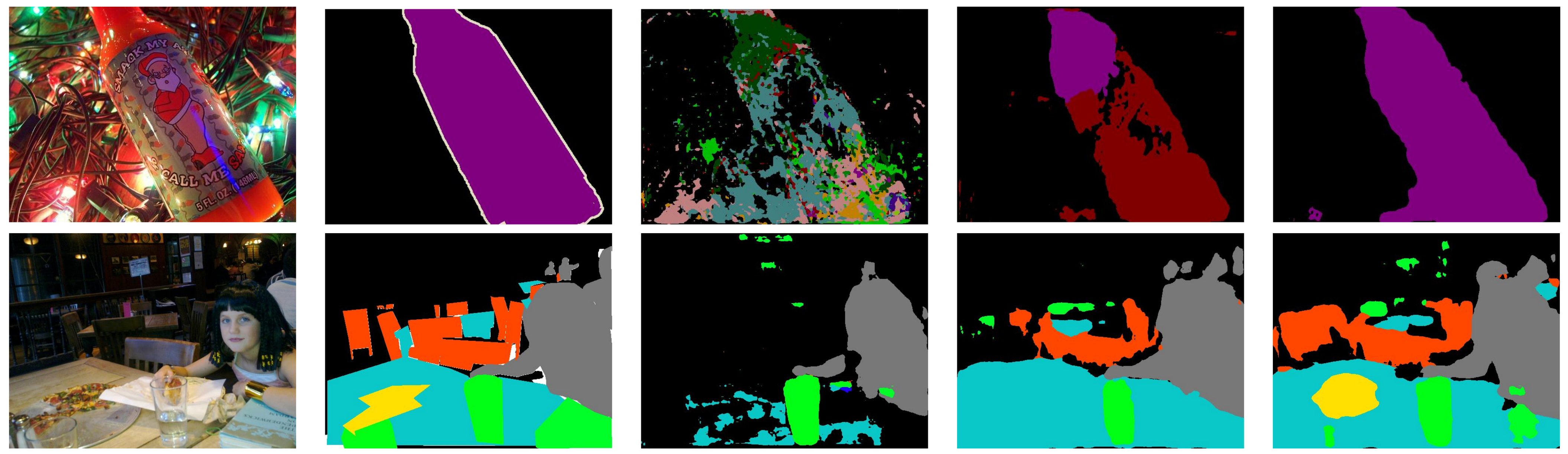
VOC2012 COCO

Improving Fully-Supervised Model with Extra Data

Method	Baseline	PseudoSeg (w/o image-level labels)		PseudoSeg (w/ image-level labels)	
Extra data	-	C_{tr}+C_u	$C_{tr} + V_{9k}$	C_{tr}	$C_{tr} + V_{9k}$
mIoU (%)	76.96	77.40 (+0.44)	78.20 (+1.24)	77.80 (+0.84)	79.28 (+2.32)

- Cross-dataset semi-supervised learning setting (VOC+COCO)
- Improving fully-supervised learning also in high-data regime

Qualitative Results



Input Ground truth Baseline Ours (w/ unlabeled) Ours (w/ weak labels)