Supplementary Material - Contrastive Learning for Label Efficient Semantic Segmentation

Xiangyun Zhao[1,2*] zhaoxiangyun915@gmail.com
Raviteja Vemulapalli[2] ravitejavemu@google.com
Philip Andrew Mansfield[2] memes@google.com
Boqing Gong[2] bgong@google.com
Bradley Green[2] brg@google.com
Lior Shapira[2] liorshap@google.com
Ying Wu[1] yingwu@northwestern.edu

[1]Northwestern University [2]Google Research

1. Performance on test splits

Table 1 shows the performance improvements on the test splits of the Cityscapes and PASCAL VOC 2012 datasets obtained by pretraining using within-image contrastive loss in the fully-supervised setting. Similar to the results on the validation splits (Sec. 4.4 of the main submission), we observe significant performance gains on the test splits.

Table 1. Performance on the test splits of Cityscapes and PASCAL VOC 2012 datasets.

Cityscapes		
Training images	(2975 images)	(596 images)
No pretraining	76.3	64.6
Contrastive pretraining	(1.8 ↑) 78.1	(3.4 ↑) 68.0
PASCAL VOC 2012		
Training images	(10528 images)	(2118 images)
No pretraining	67.2	39.4
Contrastive pretraining	(7.6 ↑) 74.8	(21.7 ↑) 61.1

2. Performance gain in semi-supervised setting

Figures 1 and 2 show the performance improvements on the validation splits of the Cityscapes and PASCAL VOC 2012 datasets, respectively, obtained by contrastive pretraining in the semi-supervised setting. Here, we use contrastive pretraining for both the initial model that is used to generate the pseudo labels, and the final model that is trained with labeled and pseudo-labeled images. Contrastive pretraining consistently improves the performance on both the datasets for different amounts of labeled and unlabeled training data. On the Cityscapes dataset, we see large gains (2.8 - 7.4 points) in terms of mean IOU, and on the PASCAL VOC 2012 dataset, we see huge gains (up to about 30 points). Similar to the fully-supervised setting, we are able to reduce the labeling requirements by 2× while improving the performance on the PASCAL VOC 2012 dataset.

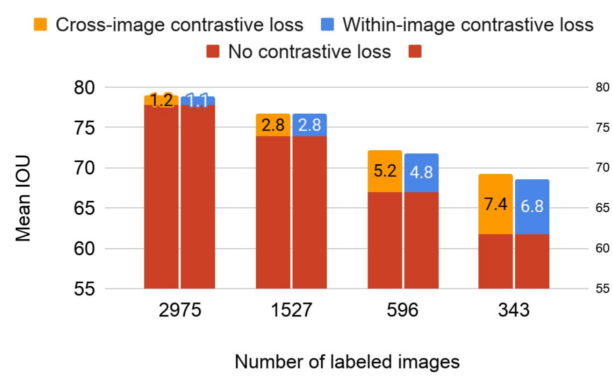

Figure 1. Improvement on Cityscapes validation dataset due to contrastive pretraining in the semi-supervised setting. Note that # unlabeled images = 2975 − # labeled images.

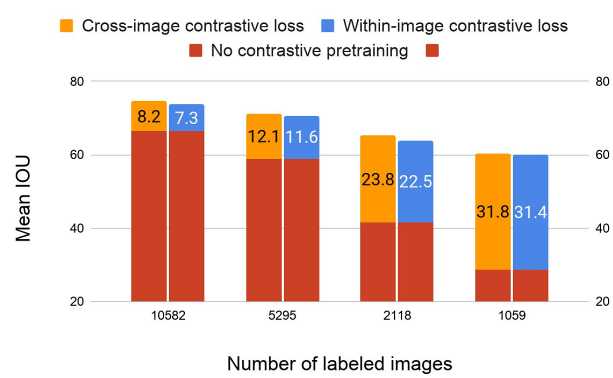

Figure 2. Improvement on PASCAL VOC 2012 validation dataset due to contrastive pretraining in the semi-supervised setting. Note that # unlabeled images = 10582 − # labeled images.

*This work was done when Xiangyun Zhao was interning at Google.

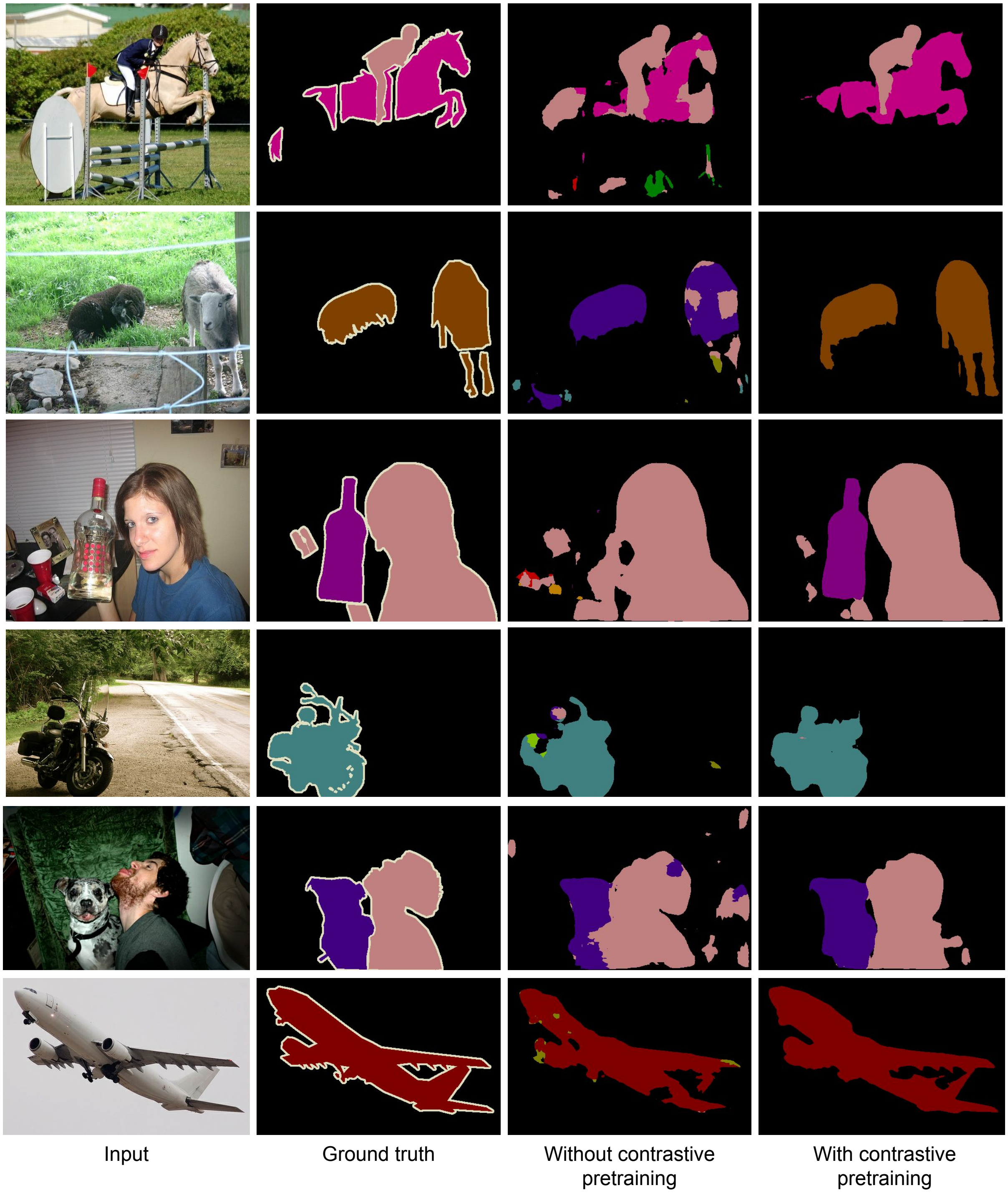

Figure 3. Comparison of models trained with and without contrastive pretraining on 2118 labeled images from PASCAL VOC 2012 dataset. Contrastive pretraining improves the results by reducing the confusion between various classes.

3. Visual results

Figure 3 shows some segmentation results of models trained with and without label-based contrastive pretraining using 2118 labeled images from the PASCAL VOC 2012 dataset. Contrastive pretraining improves the segmentation results by reducing the confusion between background and various foreground classes, and also the confusion between different foreground classes.