C^3-SemiSeg: Contrastive Semi-supervised Segmentation via Cross-set Learning and Dynamic Class-balancing

Yanning Zhou
CUHK

Hang Xu*
Huawei Noah's Ark Lab

Wei Zhang
Huawei Noah's Ark Lab

Bin Gao
Huawei Noah's Ark Lab

Pheng-Ann Heng
CUHK

1. Details for Data Augmentation Strategy

Region-level data mixing. As stated in our main text, we perform region-level data mixing among labeled and unlabeled data using CutMix [3] operation. Specifically, we fix the rectangle region's size to the half of the inputs and vary the aspect ratio and the position for each mixing pair.

Although the mixing image does not exist in the real world, it is reasonable to consider it as a kind of augmentation technique for training, which is similar to the random crop operation. Besides, experiments show that conducting cross-set region-level data mixing enriches not only the diversity of unlabeled samples but also deduces the feature misalignment between labeled and unlabeled set.

Overall augmentation strategy. In a semi-supervised learning setting, we apply the asymmetric data augmentation strategy for the mean teacher network. Specifically, weak augmentation is conducted on the labeled data, as well as the unlabeled data for the teacher's input. The strong augmentation with Auto Augmentation[1] is applied to the unlabeled data to generate the student's input. We summarize the operations in Table 1.

2. Mask Quality from Dynamic Confident Region Selection

DCRS is designed to reduce the side effect from the noisy predictions during consistency regularisation. Although DCRS cannot eliminate all noisy predictions, it can filter out most of them in low confident regions. In this section, we further analyze how large DCRS can improve the quality of the soft pseudo-label. We evaluate the predictions using the converged models on the validation set for all data ratio and summarize them in Table 2. The name with $\star$ means the low confident predictions from DCRS are assigned as ignored region during evaluation. It is observed that results from high confident region have higher mIoU for each labeled data ratio, with an increase of about 10. Furthermore, the improvement of moU for the relatively low quality categories (e.g., rider, pole and wall in 744) is larger than that form relatively accurate categories (road, sky and car). We have also tried to give a smaller s to the harder class so that it would preserve fewer ratio of samples similar to [2], but did not find improvement. This may be because different reservation ratio will arise label distribution misalignment and harm the optimization.

Name	Parameters
Student	
Brightness	$p = 0.5$
Contrast	$p = 0.5$
Saturation	$p = 0.5$
Hue	$p = 0.25$
Rotation	degree $= 10°$
Random horizontal flip	$p = 0.5$
CutMix [3]	prop_range $= 0.5$
	num_layers $= 2$
Auto Augmentation [1]	magnitude $= 10$
	magnitude_std $= 0.5$
Teacher	
Rotation	degree $= 10°$
Random horizontal flip	p = 0.5

Table 1. Augmentation Strategy.

3. Qualitative Results

More qualitative results of C^3-SemiSeg on Cityscapes validation set with different proportions of labeled data are presented in Figure 1.

*Corresponding Author: xbjxh@live.com

[1]Implementation from https://github.com/HobbitLong/PyContrast

	road	sidewalk	building	wall	fence	pole	traffic_light	traffic_sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	Avg.
Ours-100	96.3	73.7	86.9	23.0	28.9	39.7	42.3	54.5	88.1	47.3	92.1	64.6	32.3	88.9	39.3	17.2	25.8	44.3	59.5	55.0
Ours-100⋆	**98.5**	**85.7**	**93.4**	**31.6**	**37.4**	**54.5**	**60.1**	**68.8**	**94.9**	**57.4**	**98.0**	**76.0**	**37.7**	**95.7**	**49.4**	**22.7**	**29.8**	**54.2**	**70.2**	**64.0**
Ours-372	96.7	76.1	88.1	37.6	40.3	43.1	48.4	62.9	89.2	56.2	92.0	70.3	47.9	91.3	40.1	64.1	49.0	51.3	65.6	63.7
Ours-372⋆	**98.8**	**87.8**	**95.1**	**49.6**	**52.8**	**61.5**	**67.2**	**81.0**	**96.2**	**69.3**	**98.4**	**82.6**	**57.4**	**97.3**	**52.7**	**75.4**	**58.9**	**62.0**	**75.5**	**74.7**
Ours-744	97.1	78.1	88.5	42.9	46.6	43.2	48.6	62.5	89.5	57.4	92.4	70.3	46.4	91.6	57.2	77.5	58.7	48.6	65.3	66.4
Ours-744⋆	**99.1**	**90.0**	**95.3**	**54.2**	**58.2**	**61.8**	**67.3**	**80.1**	**96.5**	**70.7**	**98.5**	**82.5**	**56.7**	**97.6**	**69.5**	**88.0**	**71.2**	**60.4**	**76.3**	**77.6**
Ours-all	97.2	79.1	89.4	49.9	49.2	42.3	51.6	64.7	90.0	61.3	92.6	71.3	51.7	92.3	69.0	80.4	62.6	53.4	68.9	69.3
Ours-all⋆	**99.1**	**91.0**	**95.9**	**62.9**	**60.4**	**61.7**	**69.1**	**81.7**	**96.8**	**75.0**	**98.4**	**83.3**	**62.7**	**98.1**	**81.8**	**88.3**	**70.1**	**66.7**	**81.7**	**80.2**

Table 2. Performance analysis over different labeled data ratio on Cityscapes validation set. The name with ⋆ means the low confident predictions from DCRS are assigned as ignored region during evaluation.

References

[1] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical automated data augmentation with a reduced search space. In *CVPR workshop*, 2020.

[2] Ke Mei, Chuang Zhu, Jiaqi Zou, and Shanghang Zhang. Instance adaptive self-training for unsupervised domain adaptation. In *ECCV*, 2020.

[3] Sangdoo Yun, Dongyoon Han, Sanghyuk Chun, Seong Joon Oh, Youngjoon Yoo, and Junsuk Choe. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *ICCV*, 2019.

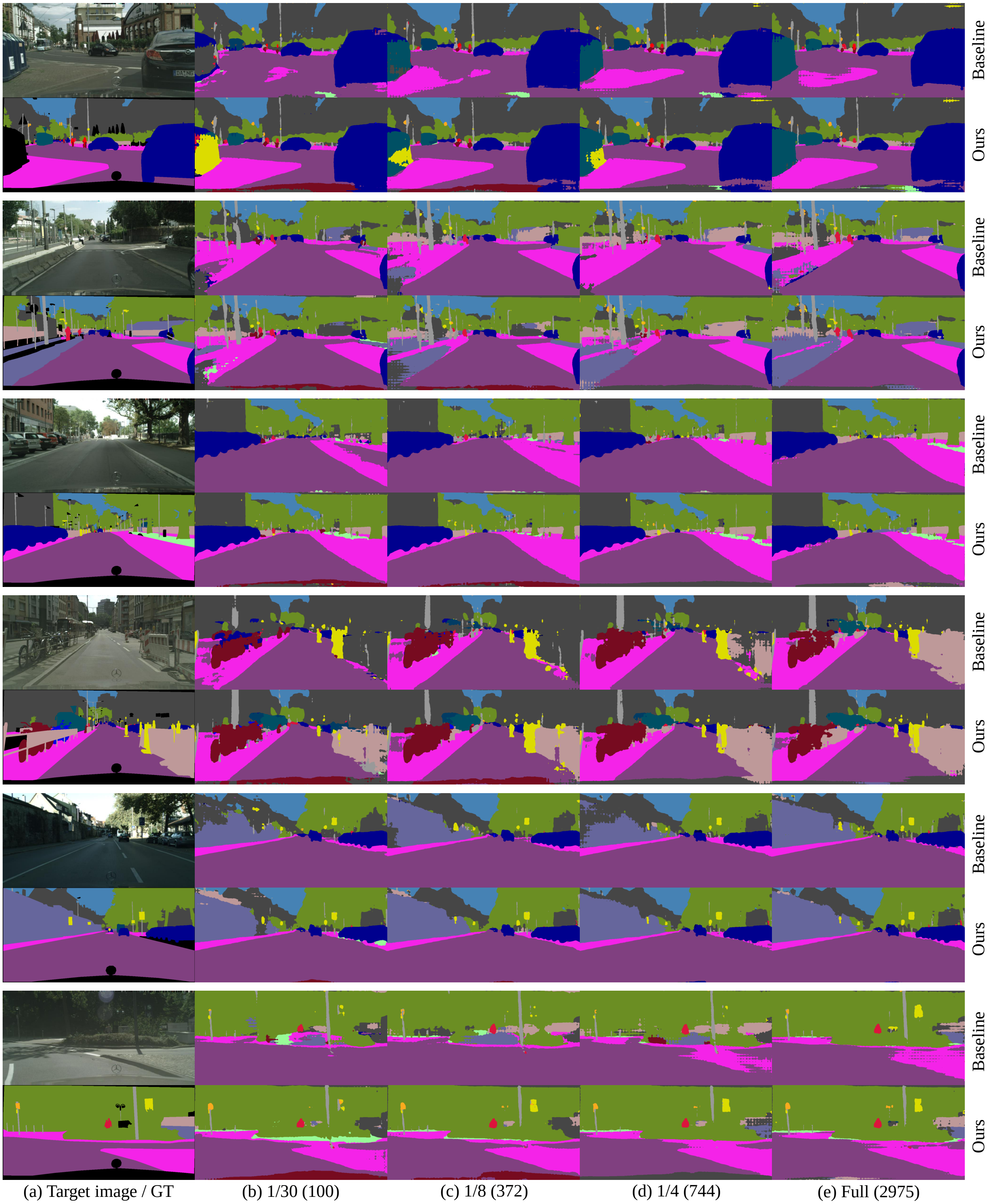

Figure 1. Qualitative results of our method and baseline method on different proportions of labeled images on Cityscapes validation dataset. (a) target images and corresponding ground truth (GT), (b)-(e) segmentation results of different proportions of labeled images.