Unsupervised Domain Adaptation for Semantic Segmentation with Symmetric Adaptation Consistency

Zongyao Li, Ren Togo, Takahiro Ogawa, Miki Haseyama
Hokkaido University

45th International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2020

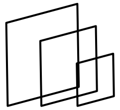
Background|Unsupervised domain adaptation (UDA)

Feature space

Domain A
(labeled)

Share the same
label space

Domain B
(unlabeled)



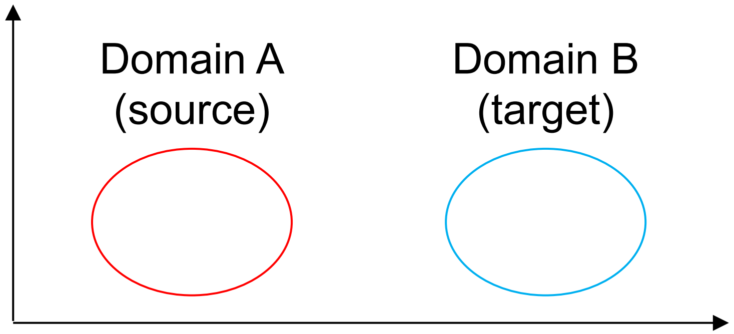
Model learned from domain A

Apply to domain B

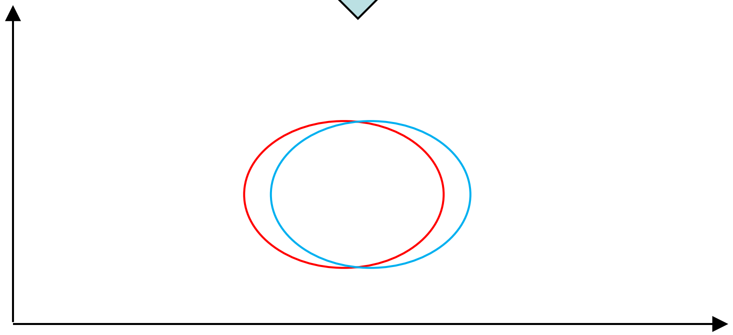
Performance drop
(cannot perform as well as in domain A)

Unsupervised domain adaptation problem

To solve this



adapt



Domains are aligned

Background|UDA for semantic segmentation

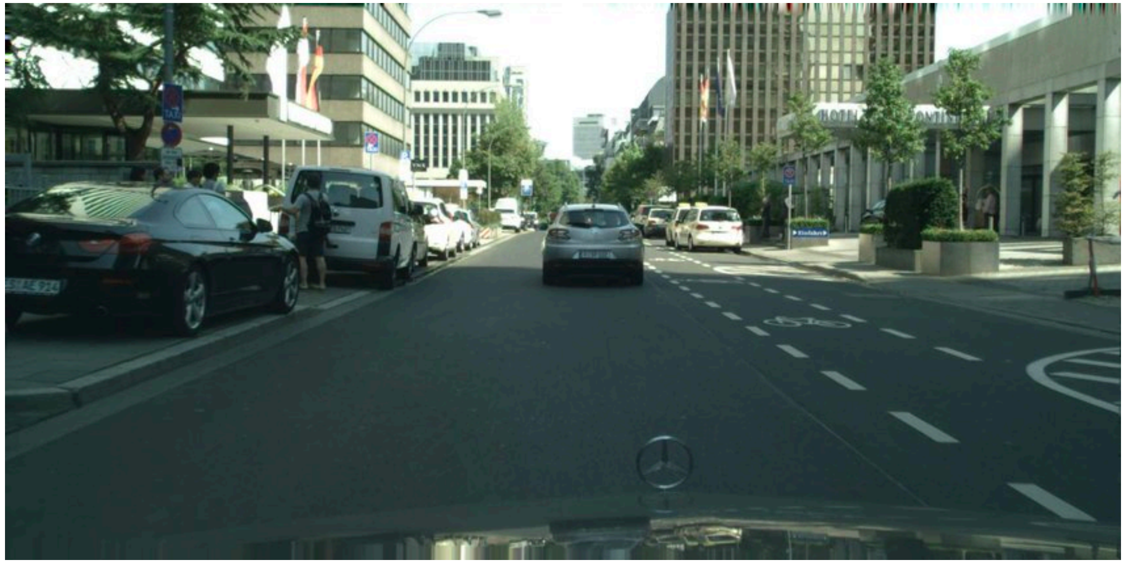
When applying semantic segmentation,
much more complex scenes
↓
much more complex feature space
↓
much more difficult

Example

GTA5 dataset[1] (source)

Adapt to →

CITYSCAPES[2] dataset (target)



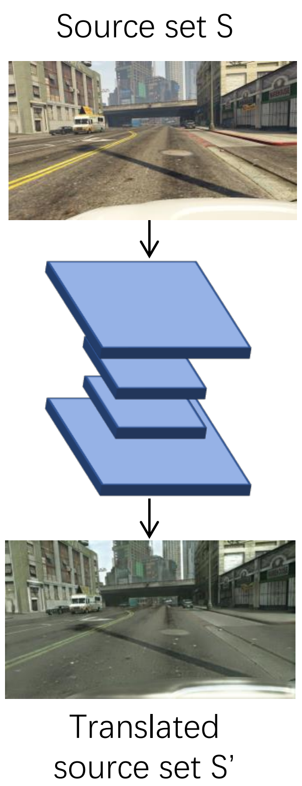
[1] Richter, Stephan R., et al. "Playing for data: Ground truth from computer games." *European conference on computer vision*. Springer, Cham, 2016.
[2] Cordts, Marius, et al. "The cityscapes dataset for semantic urban scene understanding." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Method|Overview

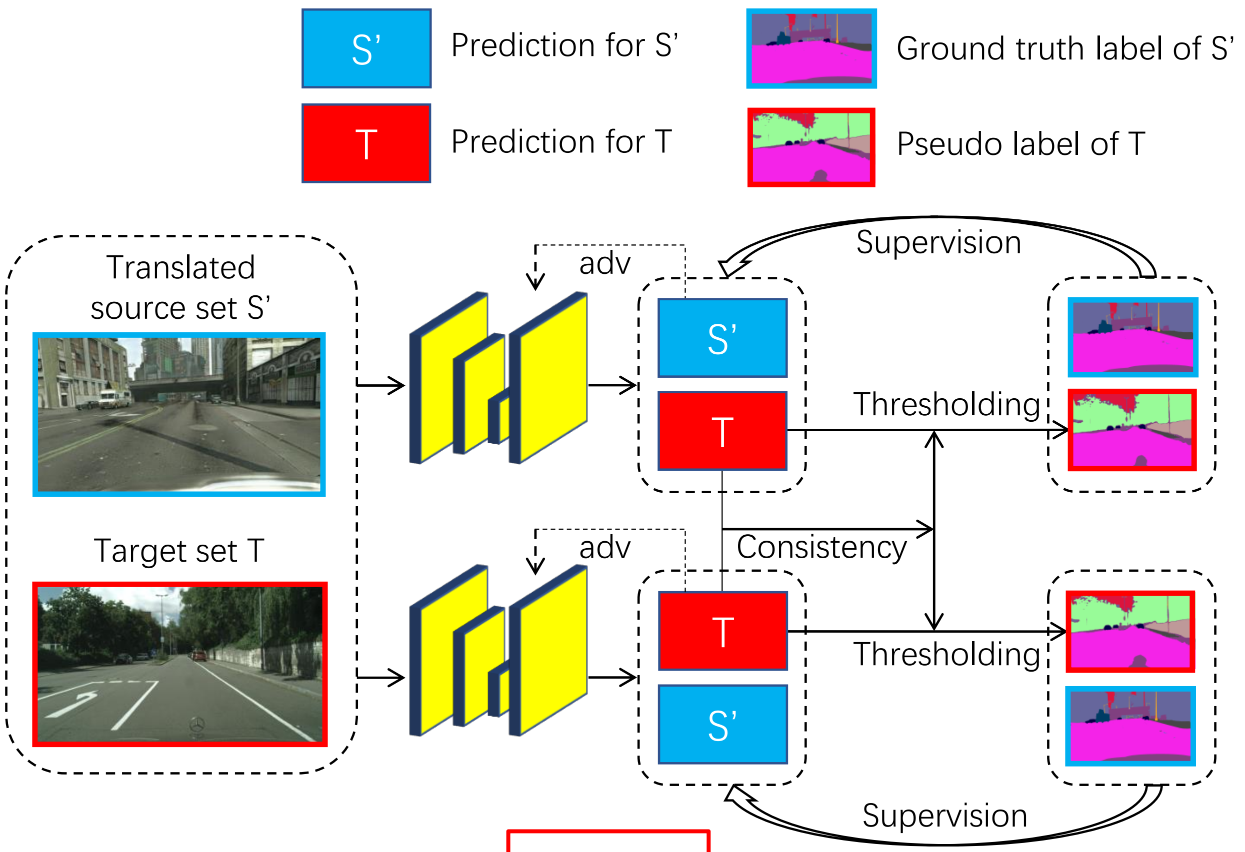
Stage 1 → Stage 2

Two stages:

Image-to-image translation



Feature-level domain adaptation



Novelty

Stage 1

Image-to-image translation:
(based on StarGAN[3])

- Translate source set S

to target domain (referred to as S') for reducing visual differences

Stage 2

Feature-level domain adaptation:

- Adversarial learning

for aligning features distributions

- Pseudo labels

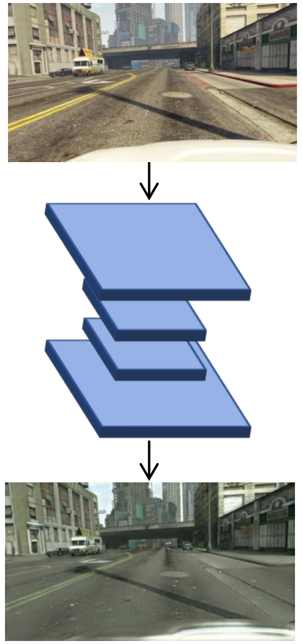
for further improvements

[3] Choi, Yunjey, et al. "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

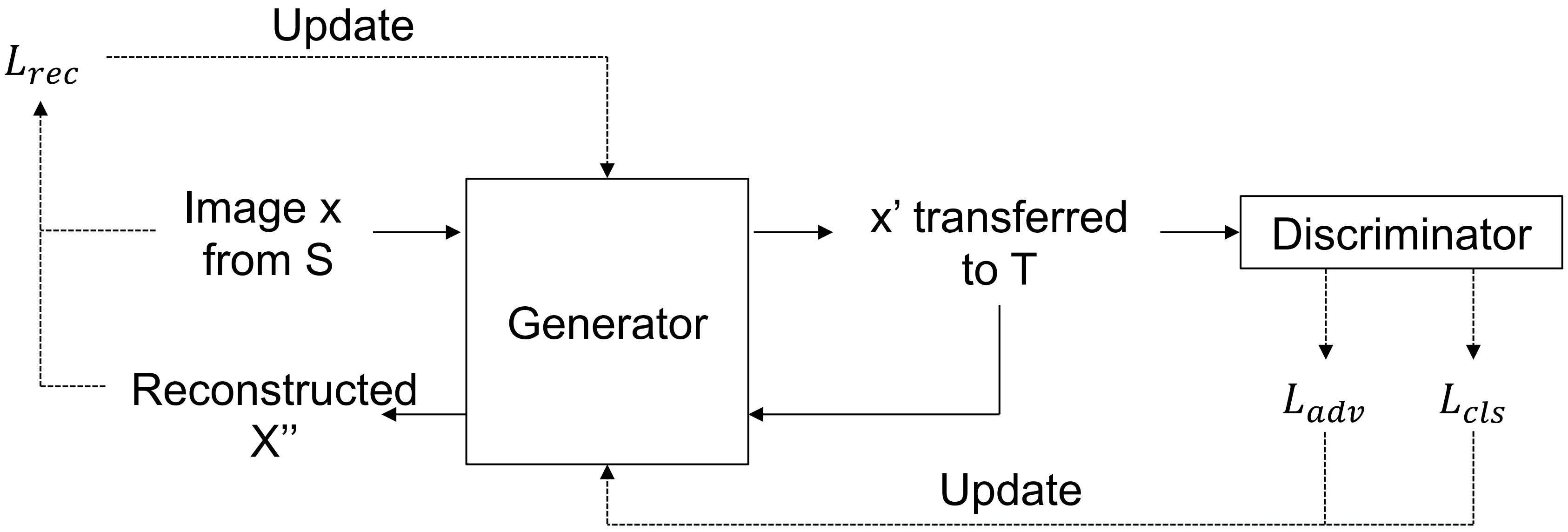
Method|Image-to-image translation

Image-to-image translation

Source set S

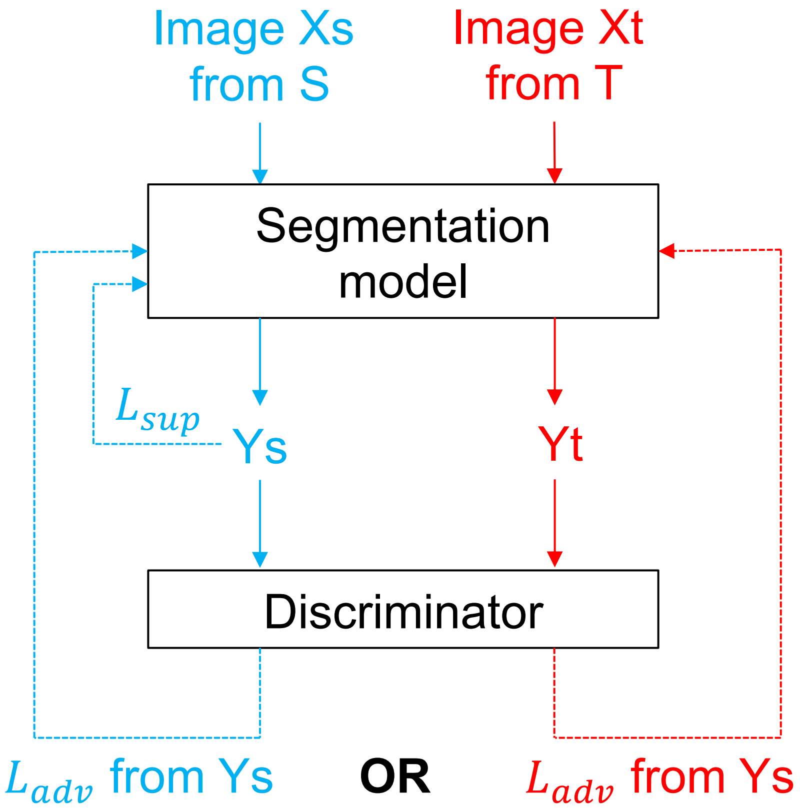


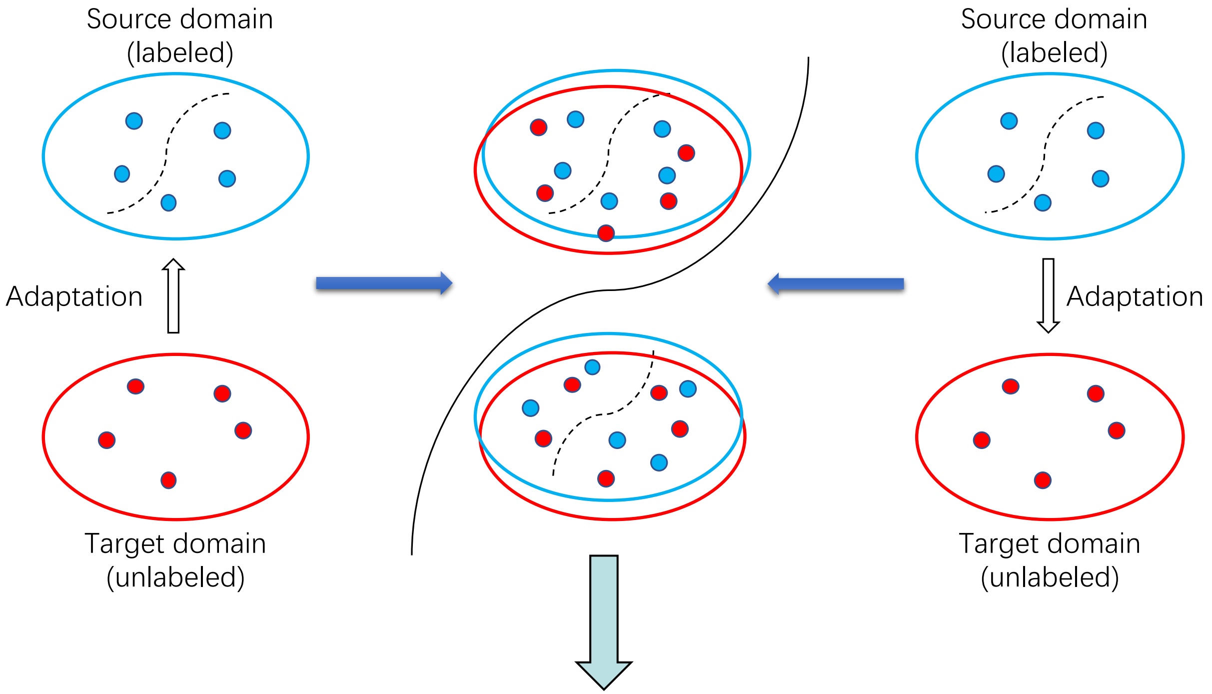
Translated source set S’



L_{rec}: reconstruction loss
L_{adv}: adversarial loss (real or fake)
L_{cls}: domain classification loss (S or T)

Method|Symmetric adaptation





More robust than single one

More accurate pseudo labels

Accuracy of pseudo labels has great impact on final performance.

Confidence map for selecting pixels:

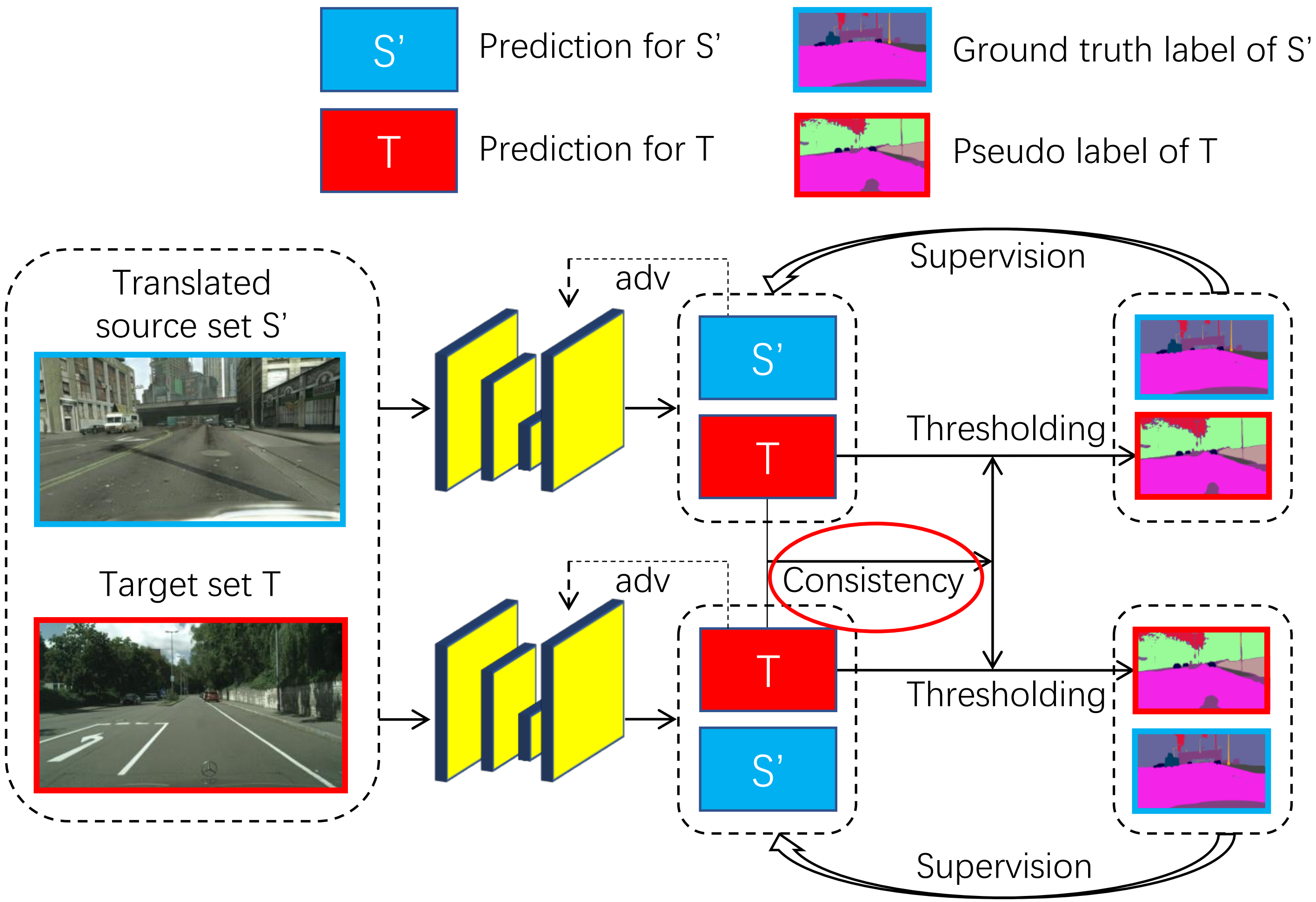
$$M_{confi} = \frac{M_{proba} + M_{consist}}{2}$$

M_{proba}: probability map

$M_{consist}$: consistency map

Consistency can help to produce more reliable pseudo labels.

Feature-level domain adaptation



Experiment|Datasets and implementation details

GTA5 dataset (used as source domain)

- Labeled synthesis data
- Including 24966 images of urban scenes
- 1914 x 1052 resolution

19 shared categories (road, sky, tree, car, building …)

CITYSCAPES dataset (used as target domain)

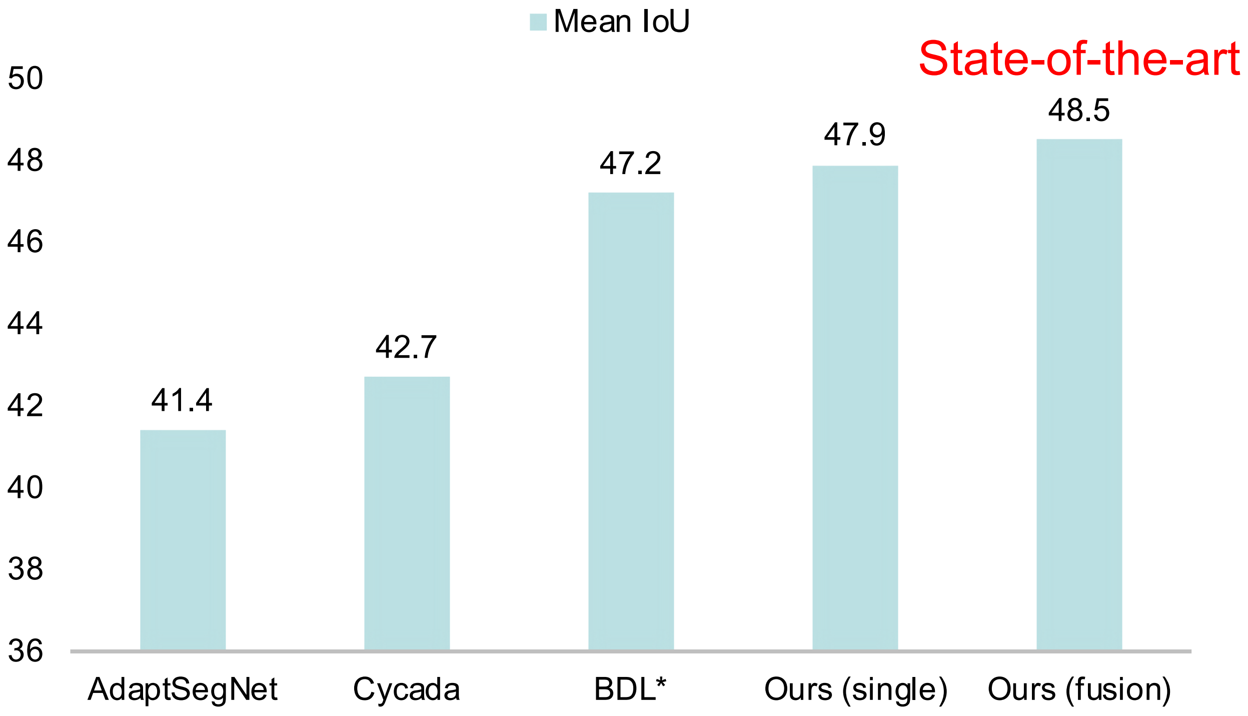
- Unlabeled real-world data
- Including 2975 images as training set and 500 images as test set
- 2048 x 1024 resolution

Implementation details

Implementation environment: Python3 + Pytorch1.1

Network architecture: DeepLab V2 with Resnet 101

Experiment|Experiment results



	Using consistency	Not using consistency
Accuracy of pseudo labels	72.6	70.6

AdaptSegNet[4]: adversarial learning

Cycada[5]: Image-to-image translation
+ adversarial learning

BDL[6]: Image-to-image translation
+ adversarial learning
+ pseudo label

Ours: Image-to-image translation
+ symmetric adversarial learning
+ pseudo label (using symmetric consistency)

Ours(single) – performance of single model

Ours(fusion) – performance fusing two models

*results when training image-to-image translation model once

[4] Tsai, Yi-Hsuan, et al. "Learning to adapt structured output space for semantic segmentation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
[5] Hoffman, Judy, et al. "Cycada: Cycle-consistent adversarial domain adaptation." *arXiv preprint arXiv:1711.03213* (2017).
[6] Li, Yunsheng, Lu Yuan, and Nuno Vasconcelos. "Bidirectional learning for domain adaptation of semantic segmentation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.

Conclusion

We have proposed an unsupervised domain adaptation method for semantic segmentation.

Architecture of our method:

- Consisting of two stages, image-to-image translation and feature-level adaptation.
- In feature-level adaptation employing adversarial learning and pseudo labels.

Advantages of our method:

- Symmetric adaptation with adversarial learning is more robust.
- Pseudo labels produced using symmetric consistency are more reliable.

Achievement:

- Our method achieved state-of-the-art performance on GTA5-to-CITYSCAPES scenario.

Future work:

- The image-to-image translation method can still be improved for domain adaptation task.