Bidirectional Learning for Domain Adaptation of Semantic Segmentation

Yunsheng Li
UC San Diego
yul554@eng.ucsd.edu

Lu Yuan
Microsoft
luyuan@microsoft.com

Nuno Vasconcelos
UC San Diego
nvasconcelos@ucsd.edu

Supplemental Material

In the supplemental material, we further analyze our algorithm by showing more results. We first give results to explain the role played by image translation model and adversarial learning in *self-supervised learning* (SSL). Secondly, we show the influence given by the perceptual loss on the image translation model.

1. Self-Supervised Learning

In this section, we further show how adversarial learning (AL) and image-to-image translation model $\mathbf{F}$ can improve the performance of self-supervised learning(SSL). In Table 1, we show the segmentation results in the first iteration ($k = 1$) of our bidirectional learning system with the threshold being 0.9 for SSL. We list 4 results in Table 1

1. $\mathbf{M}^{(0)}$: the baseline model trained only with source dataset $\mathcal{S}$.
2. 'SSL only': a segmentation model is first trained with the source dataset $\mathcal{S}$ and then it is used to predict the pseudo labels for the target dataset $\mathcal{T}$. With the source data and the target data with pseudo labels, a new segmentation is trained. While the segmentation model is learned, there is no adversarial learning process.
3. 'SSL+$\mathbf{F}$': the same as 'SSL only' but with translated source dataset $\mathcal{S}'$ given by $\mathbf{F}$.
4. 'SSL+$\mathbf{F}$+AL': adversarial learning is added to 'SSL+F' when the segmentation model is learned.

Table 1. Influence of translation model and adversarial learning

GTA5 → Cityscapes	
model	mIoU
$\mathbf{M}^{(0)}$	33.6
SSL only	40.8
SSL+$\mathbf{F}$	43.6
SSL+$\mathbf{F}$+AL	46.8

We can get some interesting findings from Table 1:

1. SSL is effective in domain alignment, it can improve mIoU from 33.6 ot 40.8 directly.
2. image translation model can benefit SSL which promotes the mIoU to 43.6.
3. Adversarial learning can further reduce the domain discrepancy for SSL to improve the mIoU to 46.8.

In summary, while SSL is only used which is similar to CBST [1], SSL is not very effective. Image translation model can reduce the visual domain gap between $\mathcal{S}$ and $\mathcal{T}$ which is very useful for SSL process. The adversarial learning which is able to further reduce the domain gap by aligning the data that is not well aligned by SSL can promote the SSL process to a better performance.

2. Perceptual Loss for Image Translation

In this section, we will show directly about how a better perceptual loss given by a better segmentation adaptation model can promote the image translation model.

As the quality of translated images is hard to be evaluated and our task is mainly about segmentation, we propose to test the performance of image translation model through segmentation results. We train a segmentation model with the translated source dataset $\mathcal{S}'$ without adversarial learning and test it directly with the target dataset. The results are shown in Table 2.

Table 2. Influence of perceptual loss on image translation

GTA5 → Cityscapes	
model	mIoU
$\mathbf{F}^{(1)}$ + $\mathbf{M}^{(0)}$	40.9
$\mathbf{F}^{(2)}$ + $\mathbf{M}_0^{(1)}$	42.9
$\mathbf{F}^{(2)}$ + $\mathbf{M}_2^{(1)}$	43.8

1. $\mathbf{F}^{(1)}$ + $\mathbf{M}^{(0)}$: the image translation model $\mathbf{F}^{(1)}$ in the first iteration ($k = 1$) of bidirectional learning is trained with perceptual loss given by $\mathbf{M}^{(0)}$.
2. $\mathbf{F}^{(2)}$ + $\mathbf{M}_0^{(1)}$: the translation model $\mathbf{F}^{(2)}$ in the second iteration ($k = 2$) is trained with perceptual loss given by $\mathbf{M}_0^{(1)}$ which is not trained with SSL and can achieve mIoU 42.7.

3. $\mathbf{F}^{(2)}$ is trained with a better perceptual loss given by $\mathbf{M}_2^{(1)}$ which is trained with SSL for two iterations and achieve mIoU 47.2..

We can find from the results that *the perceptual loss with a better segmentation model can give a better image translation model*. When the perceptual loss is replaced with a better segmentation model (mIoU increases from 33.6 to 47.2), the quality of image translation for the segmentation task can keep increasing. Thus we can find the perceptual loss is very crucial for the image translation part in our bidirectional learning system, which can in return influence the whole bidirectional learning system.

References

[1] Y. Zou, Z. Yu, B. V. Kumar, and J. Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 289–305, 2018. 1