北京大学
PEKING UNIVERSITY

# Multi-source Domain Adaptation for Semantic Segmentation

Hexin Dong

Center For Data Science

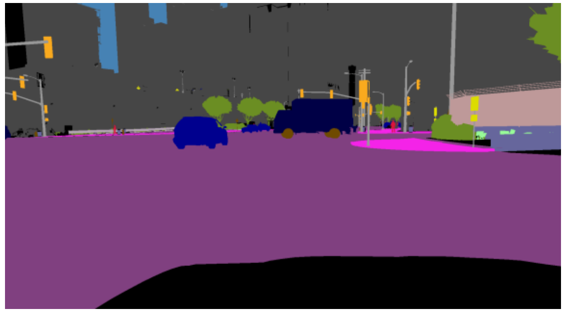
[1].Sicheng Zhao et al,Multi-source Domain Adaptation for Semantic Segmentation ,NIPS 2019

# 01 Introduction

## Semantic Segmentation:

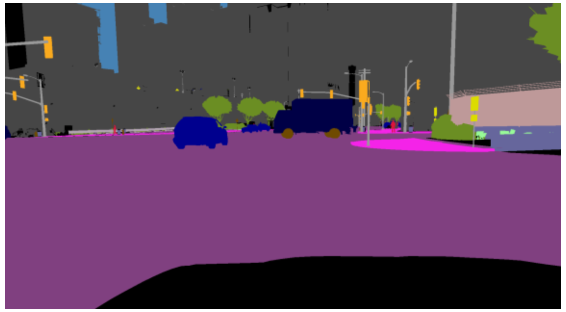
image

Segmentation model



label

## Unsupervised Domain Adaptation:



Source Domain(with label)

Segmentation model


Target Domain(no label)

**Common Approach:**

1. find a mapping from representations of the source domain to those of the target (Pixel level method: CycleGan etc)

2. find **domain-invariant** representations that are shared between the two domains(Feature level method/distengle method)

3. Semi-Supervised Learning method (Self training)

# Method outline:

1. Image Generating Loss

   1).pixel level GAN loss 2).Reconstruction loss

   3).Sub-domain aggregation loss 4).cross-domain cycle loss

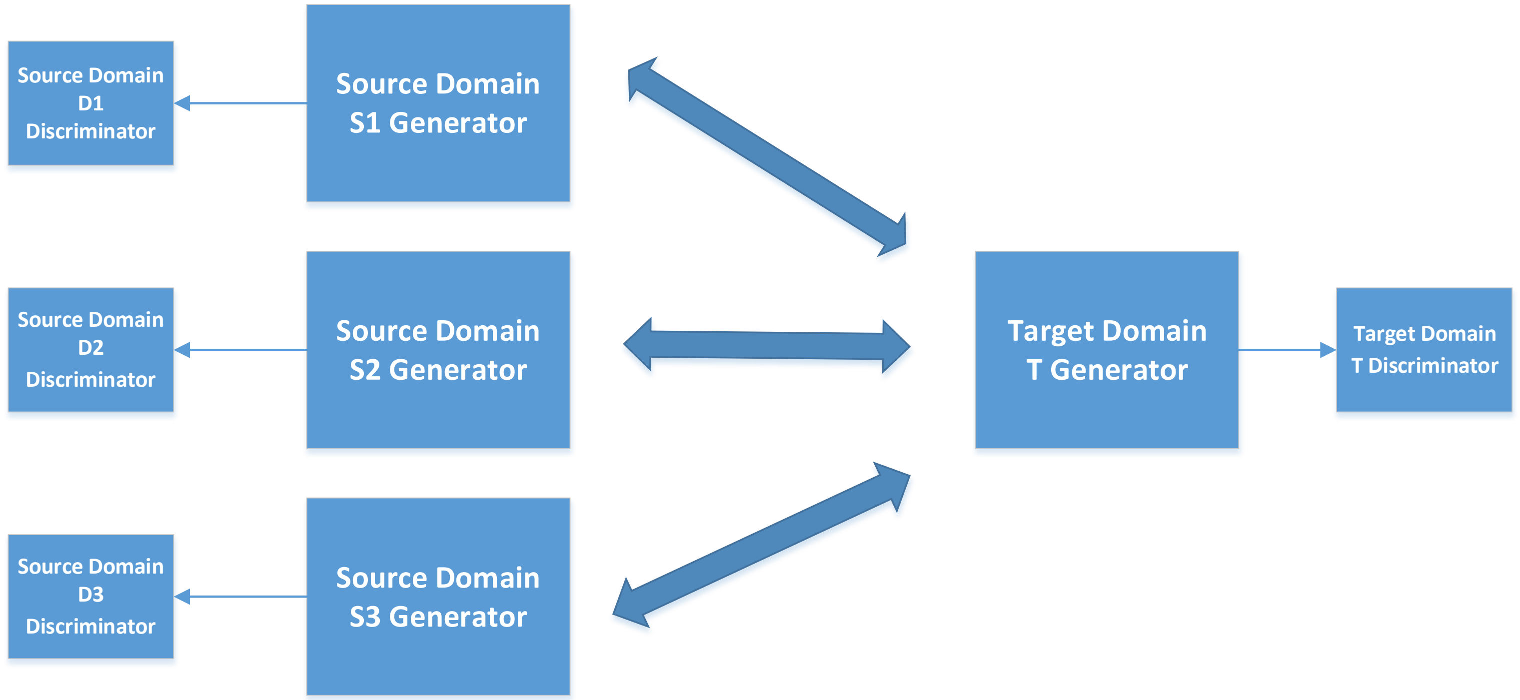
2. Semantic Based Loss

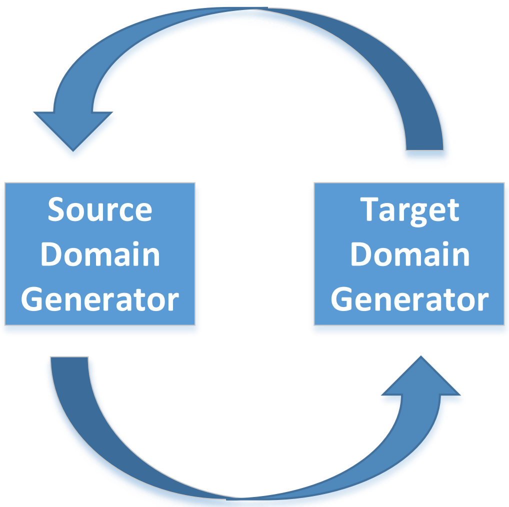
   1).task loss 2).Semantic consistency loss

3. Feature Level Loss

   -Feature level GAN loss

# 02 method




Pixel level GAN loss:

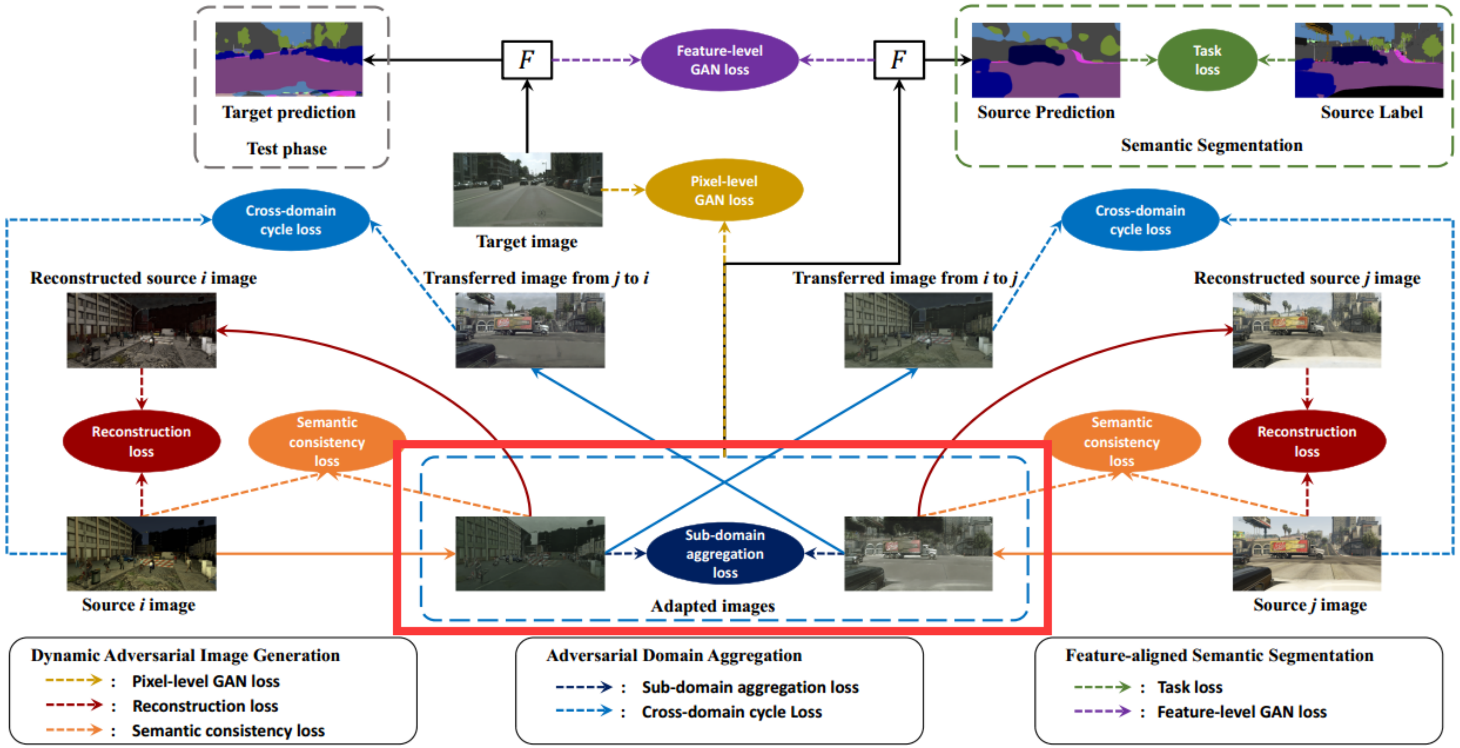
$$\mathcal{L}_{GAN}^{S_i\to T}(G_{S_i\to T}, D_T, X_i, X_T) = \mathbb{E}_{\mathbf{x}_i\sim X_i}\log D_T(G_{S_i\to T}(\mathbf{x}_i)) + \mathbb{E}_{\mathbf{x}_T\sim X_T}\log[1-D_T(\mathbf{x}_T)].$$

$$\mathcal{L}_{GAN}^{T\to S_i}(G_{T\to S_i}, D_i, X_T, X_i) = \mathbb{E}_{\mathbf{x}_i\sim X_i}\log[1-D_i(\mathbf{x}_i)] + \mathbb{E}_{\mathbf{x}_t\sim X_T}\log D_i(G_{T\to S_i}(\mathbf{x}_t)).$$

Reconstuction loss:

$$\begin{aligned}\mathcal{L}_{cyc}^{S_i\leftrightarrow T}(G_{S_i\to T}, G_{T\to S_i}, X_i, X_T) =& \mathbb{E}_{\mathbf{x}_i\sim X_i} \parallel G_{T\to S_i}(G_{S_i\to T}(\mathbf{x}_i)) - \mathbf{x}_i \parallel_1 + \\ & \mathbb{E}_{\mathbf{x}_T\sim X_T} \parallel G_{S_i\to T}(G_{T\to S_i}(\mathbf{x}_t)) - \mathbf{x}_t \parallel_1 .\end{aligned}$$

# 02 method




Sub-domain aggregation loss:

$$\mathcal{L}_{SAD}^{S_i}(G_{S_1\to T},\ldots G_{S_i\to T},\ldots,G_{S_M\to T},D_A^i) = \mathbb{E}_{\mathbf{x}_i\sim X_i}\log D_A^i(G_{S_i\to T}(\mathbf{x}_i))+ \frac{1}{M-1}\sum_{j\neq i}\mathbb{E}_{\mathbf{x}_j\sim X_j}\log[1-D_A^i(G_{S_j\to T}(\mathbf{x}_j))].$$

Cross domain cycle loss:

$$\mathcal{L}_{CCD}^{S_i}(G_{T\to S_1},\ldots G_{T\to S_{i-1}},G_{T\to S_{i+1}},\ldots,G_{T\to S_M},G_{S_i\to T},D_i) = \mathbb{E}_{\mathbf{x}_i\sim X_i}\log D_i(\mathbf{x}_i)+ \frac{1}{M-1}\sum_{j\neq i}\mathbb{E}_{\mathbf{x}_j\sim X_j}\log[1-D_i(G_{T\to S_i}((G_{S_j\to T}(\mathbf{x}_j)))].$$

# 02 method

Task loss:

$$\mathcal{L}_{task}(F, X', Y) = -\mathbb{E}_{(\mathbf{x}',\mathbf{y})\sim(X',Y)} \sum_{l=1}^{L} \sum_{h=1}^{H} \sum_{w=1}^{W} \mathbb{1}_{[l=\mathbf{y}_{h,w}]} \log(\sigma(F_{l,h,w}(\mathbf{x}'))),$$

Semantic **consistency** loss:

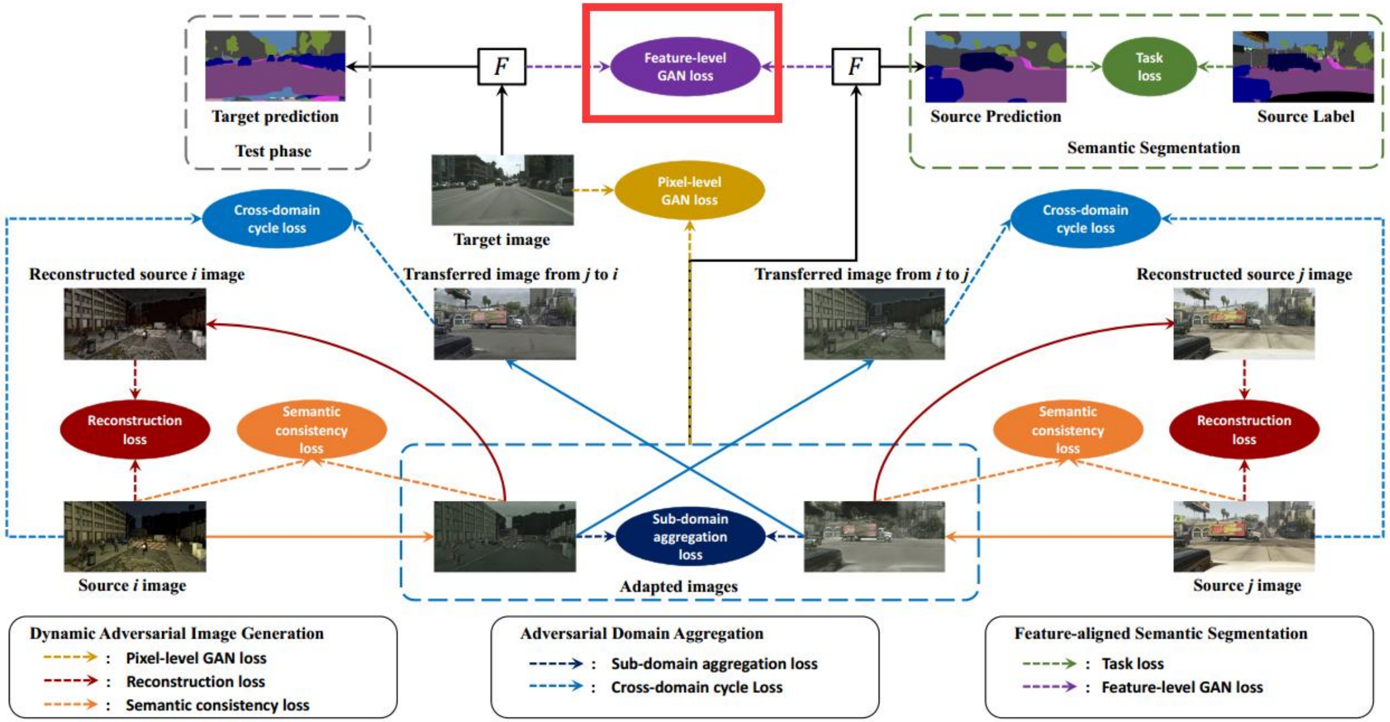
$$\mathcal{L}_{sem}^{S_i}(G_{S_i\to T}, X_i, F_i, F_A) = \mathbb{E}_{\mathbf{x}_i\sim X_i} KL(F_A(G_{S_i\to T}(\mathbf{x}_i))||F_i(\mathbf{x}_i)),$$

Similar semantic consistency loss in other paper[3]:

$$\begin{aligned}
&\mathcal{L}_{\text{consis}}(X_T; G_{S\to T}, G_{T\to S}, F_S, F_T) \\
&= -\mathbb{E}_{I_T\sim X_T} \sum_{h,w,c} f_{T\to S}(h,w,c) \log\left(f_T(h,w,c)\right) \\
&\quad -\mathbb{E}_{I_T\sim X_T} \sum_{h,w,c} f_T(h,w,c) \log\left(f_{T\to S}(h,w,c)\right),
\end{aligned}$$

[2].Yunchun Chen et al,CrDoCo: Pixel-level Domain Transfer with Cross-Domain Consistency CVPR2019

# 02 method




**Feature level gan loss**:

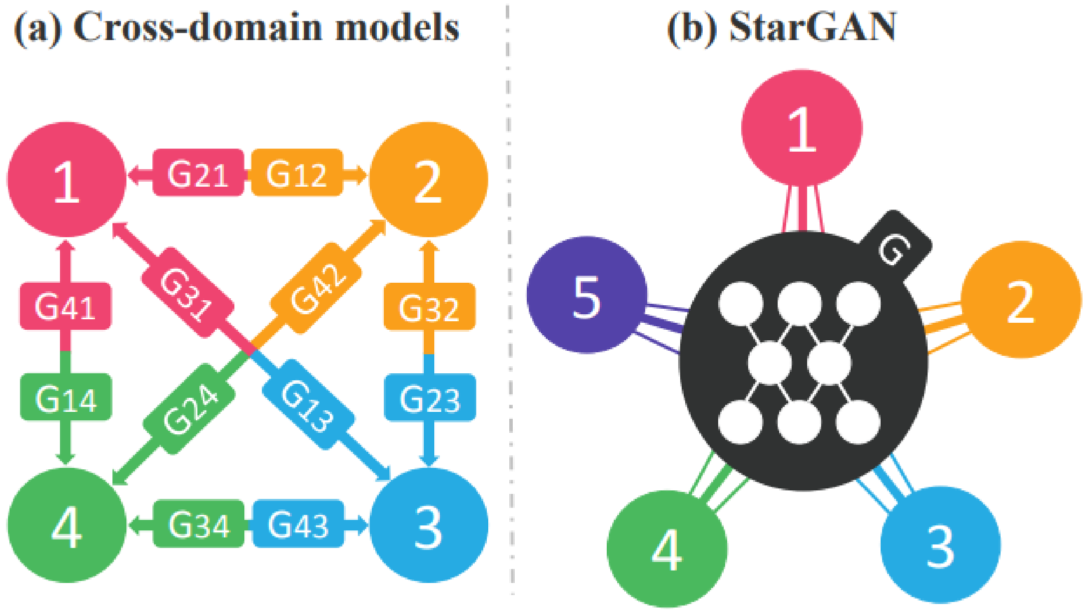
$$\mathcal{L}_{feat}(F_f, D_{F_f}, X', X_T) = \mathbb{E}_{\mathbf{x}' \sim X'} \log D_{F_f}(F_f(\mathbf{x}')) + \mathbb{E}_{\mathbf{x}_T \sim X_T} \log[1 - D_{F_f}(F_f(\mathbf{x}_T))],$$

# 03 Result

| Standards | Method | road | sidewalk | building | wall | fence | pole | t-light | t-sign | vegettion | sky | person | rider | car | bus | m-bike | bicycle | mIoU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Source-only | GTA | 54.1 | 19.6 | 47.4 | 3.3 | 5.2 | 3.3 | 0.5 | 3.0 | 69.2 | 43.0 | 31.3 | 0.1 | 59.3 | 8.3 | 0.2 | 0.0 | 21.7 |
| | SYNTHIA | 3.9 | 14.5 | 45.0 | 0.7 | 0.0 | 14.6 | 0.7 | 2.6 | 68.2 | 68.4 | 31.5 | 4.6 | 31.5 | 7.4 | 0.3 | 1.4 | 18.5 |
| | GTA+SYNTHIA | 44.0 | 19.0 | 60.1 | 11.1 | 13.7 | 10.1 | 5.0 | 4.7 | 74.7 | 65.3 | 40.8 | 2.3 | 43.0 | 15.9 | 1.3 | 1.4 | 25.8 |
| GTA-only DA | FCN Wld [47] | 70.4 | 32.4 | 62.1 | 14.9 | 5.4 | 10.9 | 14.2 | 2.7 | 79.2 | 64.6 | 44.1 | 4.2 | 70.4 | 7.3 | 3.5 | 0.0 | 27.1 |
| | CDA [48] | 74.8 | 22.0 | 71.7 | 6.0 | 11.9 | 8.4 | 16.3 | 11.1 | 75.7 | 66.5 | 38.0 | 9.3 | 55.2 | 18.9 | 16.8 | 14.6 | 28.9 |
| | ROAD [50] | 85.4 | 31.2 | 78.6 | **27.9** | **22.2** | 21.9 | 23.7 | 11.4 | 80.7 | 68.9 | 48.5 | 14.1 | 78.0 | 23.8 | 8.3 | 0.0 | 39.0 |
| | AdaptSeg [71] | **87.3** | 29.8 | 78.6 | 21.1 | 18.2 | 22.5 | 21.5 | 11.0 | 79.7 | 71.3 | 46.8 | 6.5 | **80.1** | 26.9 | 10.6 | 0.3 | 38.3 |
| | CyCADA [32] | 85.2 | 37.2 | 76.5 | 21.8 | 15.0 | 23.8 | 22.9 | 21.5 | 80.5 | 60.7 | 50.5 | 9.0 | 76.9 | 28.2 | 4.5 | 0.0 | 38.7 |
| | DCAN [55] | 82.3 | 26.7 | 77.4 | 23.7 | 20.5 | 20.4 | **30.3** | 15.9 | **80.9** | 69.5 | **52.6** | 11.1 | 79.6 | 21.2 | 17.0 | 6.7 | 39.8 |
| SYNTHIA-only DA | FCN Wld [47] | 11.5 | 19.6 | 30.8 | 4.4 | 0.0 | 20.3 | 0.1 | 11.7 | 42.3 | 68.7 | 51.2 | 3.8 | 54.0 | 3.2 | 0.2 | 0.6 | 20.2 |
| | CDA [48] | 65.2 | 26.1 | 74.9 | 0.1 | 0.5 | 10.7 | 3.7 | 3.0 | 76.1 | 70.6 | 47.1 | 8.2 | 43.2 | 20.7 | 0.7 | 13.1 | 29.0 |
| | ROAD [50] | 77.7 | 30.0 | 77.5 | 9.6 | 0.3 | **25.8** | 10.3 | 15.6 | 77.6 | **79.8** | 44.5 | 16.6 | 67.8 | 14.5 | 7.0 | 23.8 | 36.2 |
| | CyCADA [32] | 66.2 | 29.6 | 65.3 | 0.5 | 0.2 | 15.1 | 4.5 | 6.9 | 67.1 | 68.2 | 42.8 | 14.1 | 51.2 | 12.6 | 2.4 | 20.7 | 29.2 |
| | DCAN [55] | 79.9 | 30.4 | 70.8 | 1.6 | 0.6 | 22.3 | 6.7 | **23.0** | 76.9 | 73.9 | 41.9 | 16.7 | 61.7 | 11.5 | 10.3 | **38.6** | 35.4 |
| Source-combined DA | CyCADA [32] | 82.8 | 35.8 | 78.2 | 17.5 | 15.1 | 10.8 | 6.1 | 19.4 | 78.6 | 77.2 | 44.5 | 15.3 | 74.9 | 17.0 | 10.3 | 12.9 | 37.3 |
| Multi-source DA | MDAN [69] | 64.2 | 19.7 | 63.8 | 13.1 | 19.4 | 5.5 | 5.2 | 6.8 | 71.6 | 61.1 | 42.0 | 12.0 | 62.7 | 2.9 | 12.3 | 8.1 | 29.4 |
| | **MADAN (Ours)** | 86.2 | **37.7** | **79.1** | 20.1 | 17.8 | 15.5 | 14.5 | 21.4 | 78.5 | 73.4 | 49.7 | **16.8** | 77.8 | **28.3** | **17.7** | 27.5 | **41.4** |
| Oracle-Train on Tgt | FCN [5] | 96.4 | 74.5 | 87.1 | 35.3 | 37.8 | 36.4 | 46.9 | 60.1 | 89.0 | 89.8 | 65.6 | 35.9 | 76.9 | 64.1 | 40.5 | 65.1 | 62.6 |

# 04 Improvement

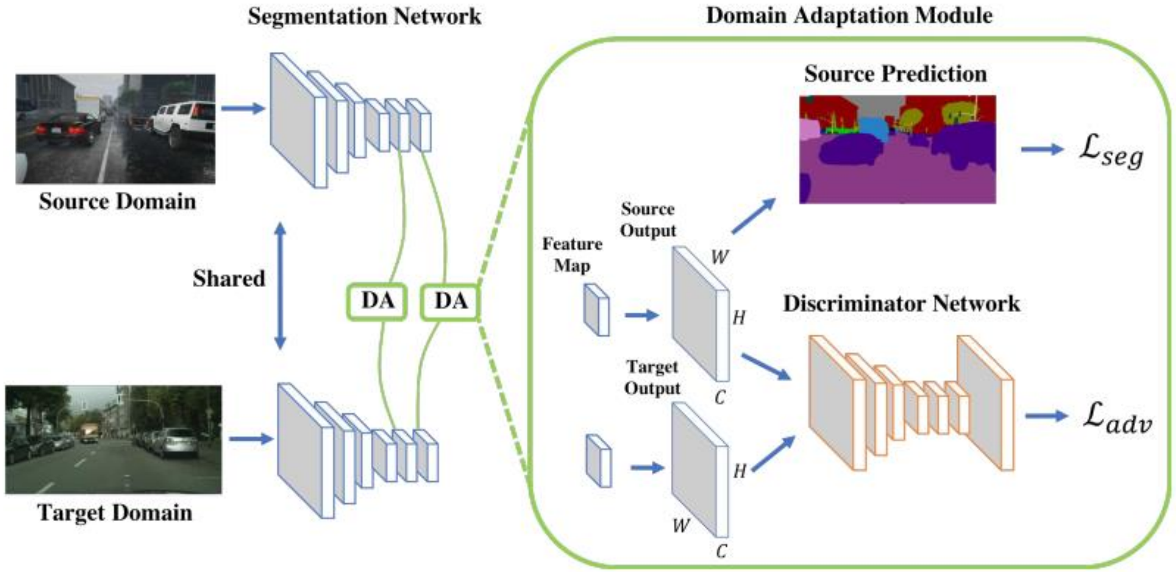
Model的可扩展性，文章中的Generator和Discriminator的数量是O(n)的量级，不利于扩展。简化为一个模型：




[3]. Yunjey Choi,et al.StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation CVPR2018

# 04 Improvement

Feature Adversarial Training vs Output space Adversarial Training：

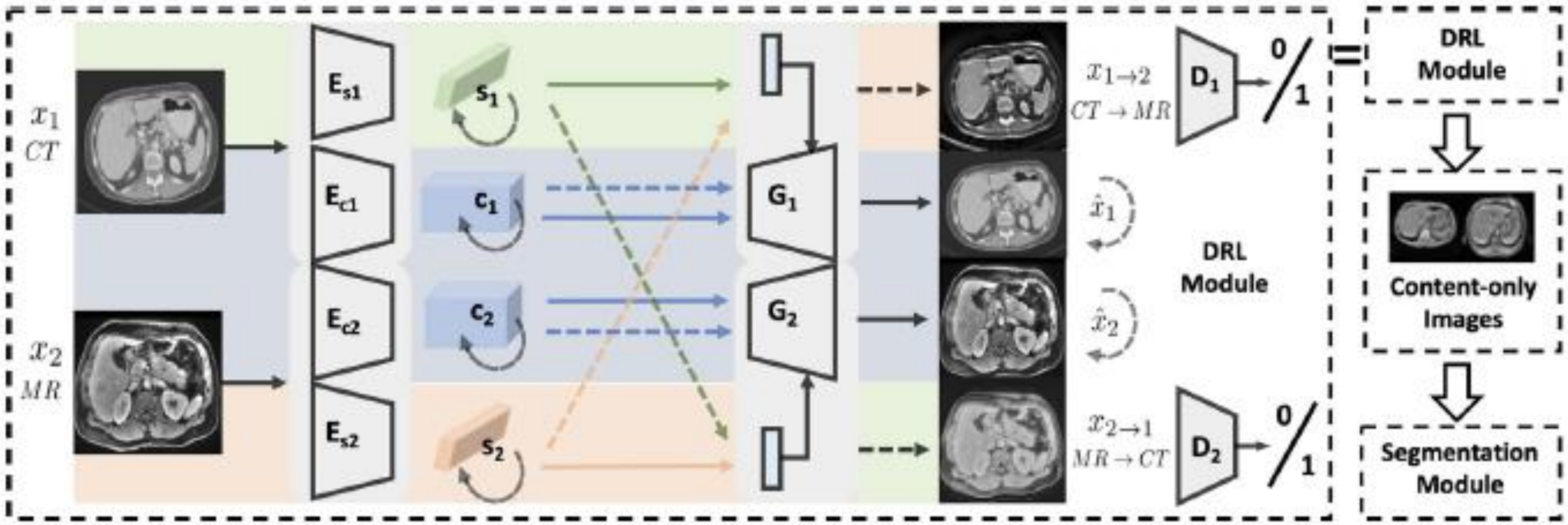



| | | | | | | | | | | | | | | | | | | | | MIOU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Baseline (ResNet) | 75.8 | 16.8 | 77.2 | 12.5 | 21.0 | 25.5 | 30.1 | 20.1 | 81.3 | 24.6 | 70.3 | 53.8 | 26.4 | 49.9 | 17.2 | 25.9 | 6.5 | 25.3 | **36.0** | 36.6 |
| Ours (feature) | 83.7 | 27.6 | 75.5 | 20.3 | 19.9 | **27.4** | 28.3 | **27.4** | 79.0 | 28.4 | 70.1 | 55.1 | 20.2 | 72.9 | 22.5 | **35.7** | **8.3** | 20.6 | 23.0 | 39.3 |
| Ours (single-level) | **86.5** | 25.9 | 79.8 | 22.1 | 20.0 | 23.6 | 33.1 | 21.8 | 81.8 | 25.9 | **75.9** | 57.3 | 26.2 | **76.3** | 29.8 | 32.1 | 7.2 | 29.5 | 32.5 | 41.4 |
| Ours (multi-level) | **86.5** | **36.0** | **79.9** | **23.4** | **23.3** | 23.9 | **35.2** | 14.8 | **83.4** | **33.3** | 75.6 | **58.5** | **27.6** | 73.7 | **32.5** | 35.4 | 3.9 | **30.1** | 28.1 | **42.4** |

**[4]. Yi-Hsuan Tsai ,et al Learning to Adapt Structured Output Space for Semantic Segmentation CVPR2018**

# 04 Improvement

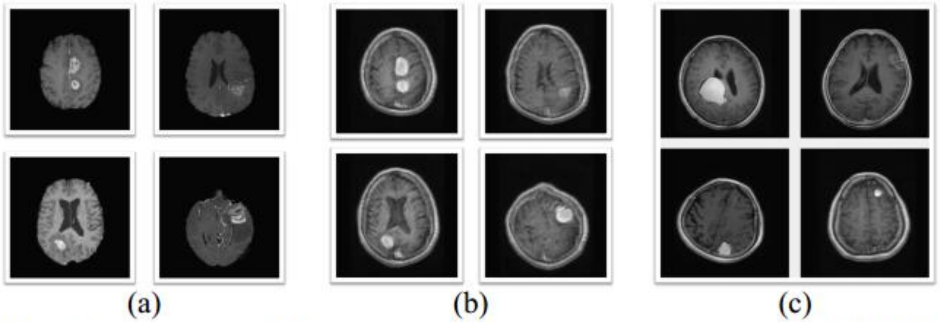
Pixel level adaptation vs distengled based adaptation：




[5]. Unsupervised Domain Adaptation via Disentangled Representations: Application to Cross-Modality Liver Segmentation, MICCAI 2019

# 04 Improvement

Adaptation+semi-supervised learning：




| amount of labeled data | prec | recall | dice | MIOU | AUC |
|---|---|---|---|---|---|
| 200 | 0.5553 | 0.5299 | 0.5061 | 0.4063 | 0.7618 |
| 400 | 0.6231 | 0.5664 | 0.5430 | 0.4359 | 0.7806 |
| 800 | 0.6330 | 0.6434 | **0.5990** | 0.4947 | 0.8149 |
| 1600 | 0.6350 | 0.6435 | 0.6034 | 0.4971 | 0.8188 |
| all(7400) | 0.6535 | 0.6325 | 0.606 | 0.4976 | 0.8137 |


**[6]. Hexin Dong et al,ANNOTATION-FREE GLIOMAS SEGMENTATION BASED ON A FEW LABELED GENERAL BRAIN TUMOR IMAGES ,ISBI 2020**

05

# Conclusions

## Multi source pixel level domain adaptation:

- CycleGAN based pixel adaptation method
- Task loss and Semantic Consistcy loss
- Feature level Adversarial learning

## Improvement:

- **Scalability —— starGAN**
- Feature Adversarial Training vs Output space Adversarial Training
- Pixel level adaptation vs distengled based adaptation
- Semi-supervised method

06

# Q&A