Deep Unsupervised Domain Adaptation and its Application to Generative Models

LI Rui

City University of Hong Kong

ruili52-c@my.cityu.edu.hk

Outline

- Introduction to Domain Adaptation

- Main Idea for Recent Adaptation Works

- Proposed Works
    - Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation
    - Generating Target Image-Label Pairs for Unsupervised Domain Adaptation
    - Model Adaptation: Unsupervised Domain Adaptation without Source Data

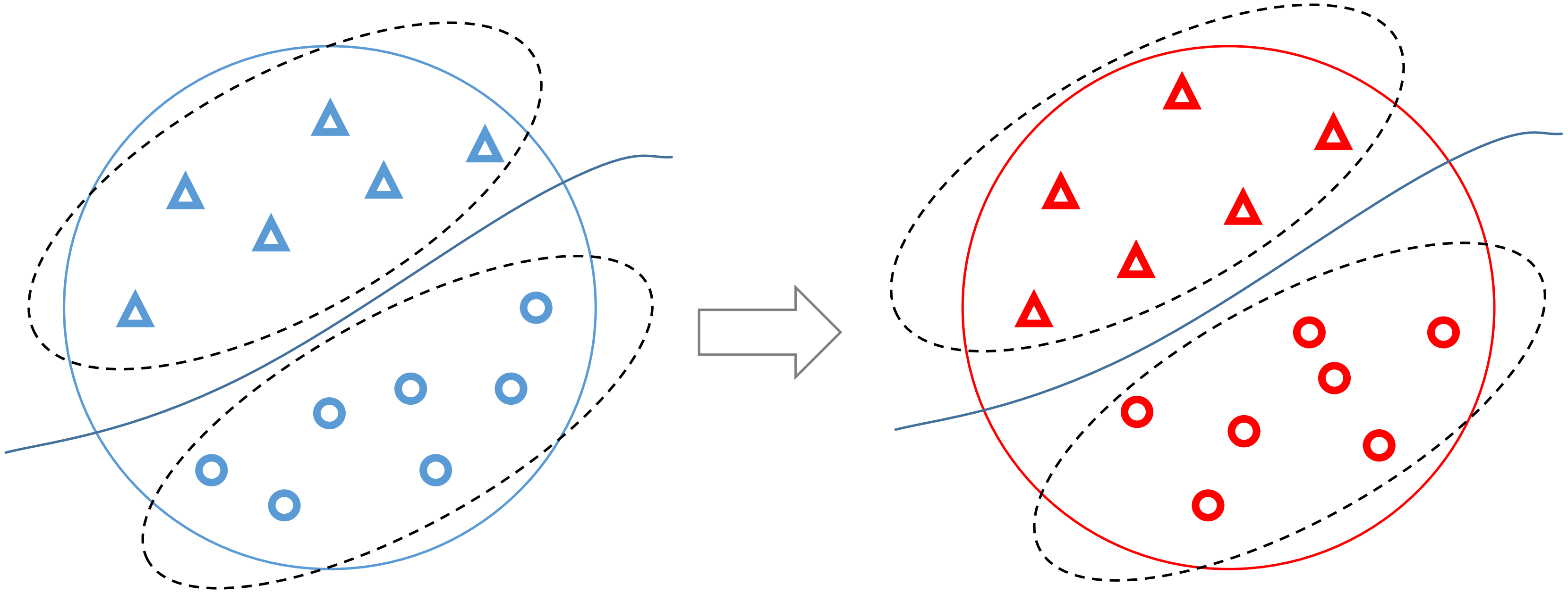
- Conclusions and Future Works

Outline

- Introduction to Domain Adaptation
- Main Idea for Recent Adaptation Works
- Proposed Works
  - Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation
  - Generating Target Image-Label Pairs for Unsupervised Domain Adaptation
  - Model Adaptation: Unsupervised Domain Adaptation without Source Data
- Conclusions and Future Works

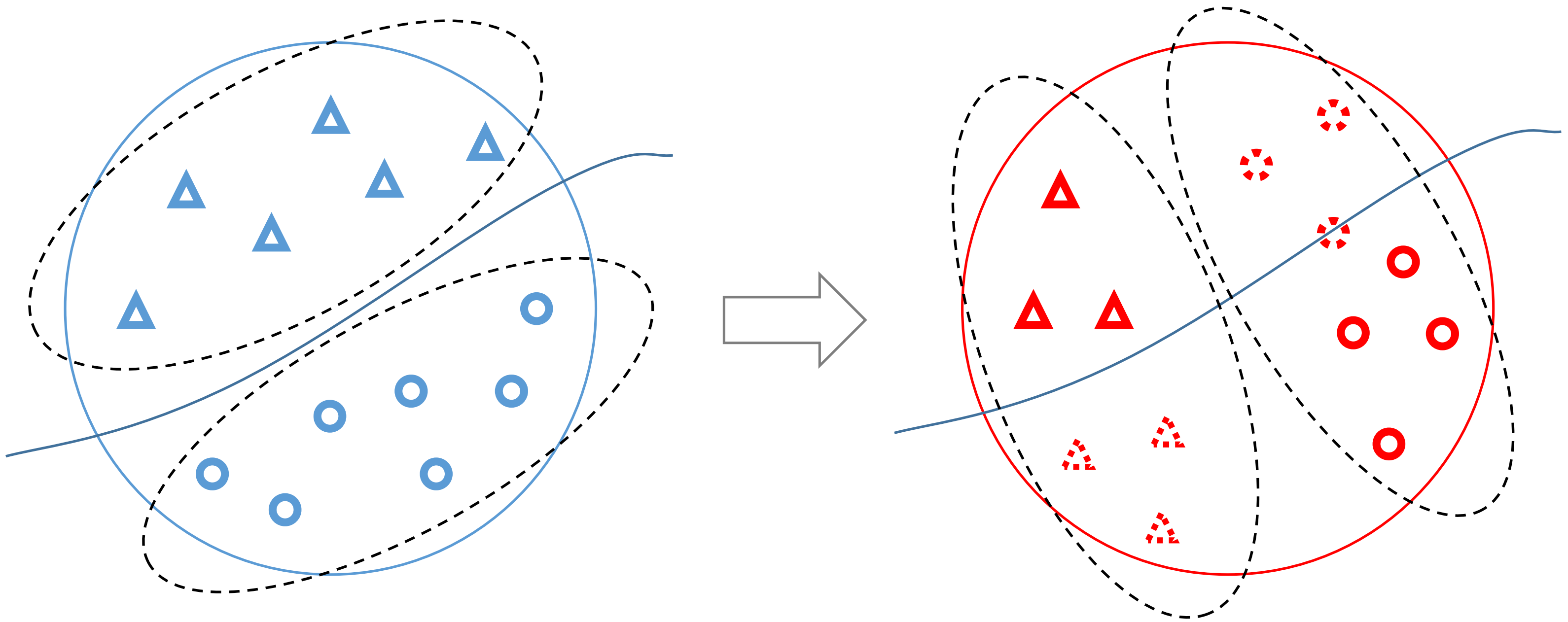
Domain Adaptation Background

- Traditional supervised learning works often implicitly assume that the training dataset and the test dataset have the same distribution, thus, the pretrained model can achieve reliable performance during testing.



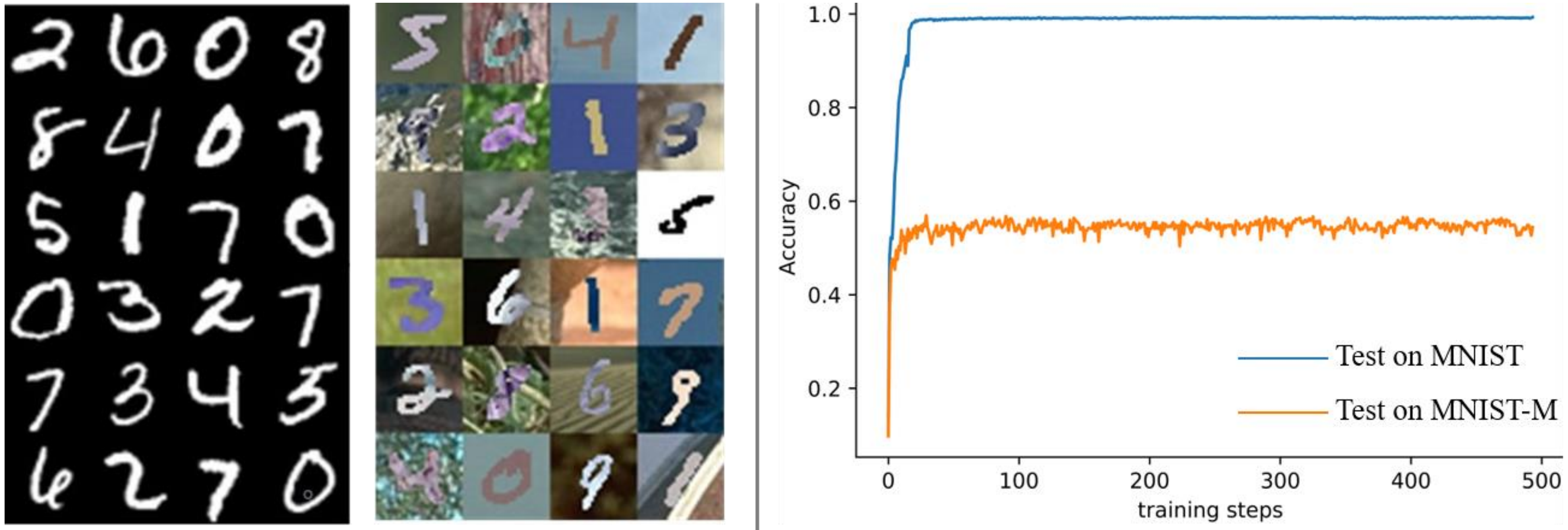
Domain Adaptation Background

- However, in real scenarios, the training and test datasets are from related but different distributions. The performance of source model may be degraded (called **domain shift**).



Real Domain Adaptation Scenarios

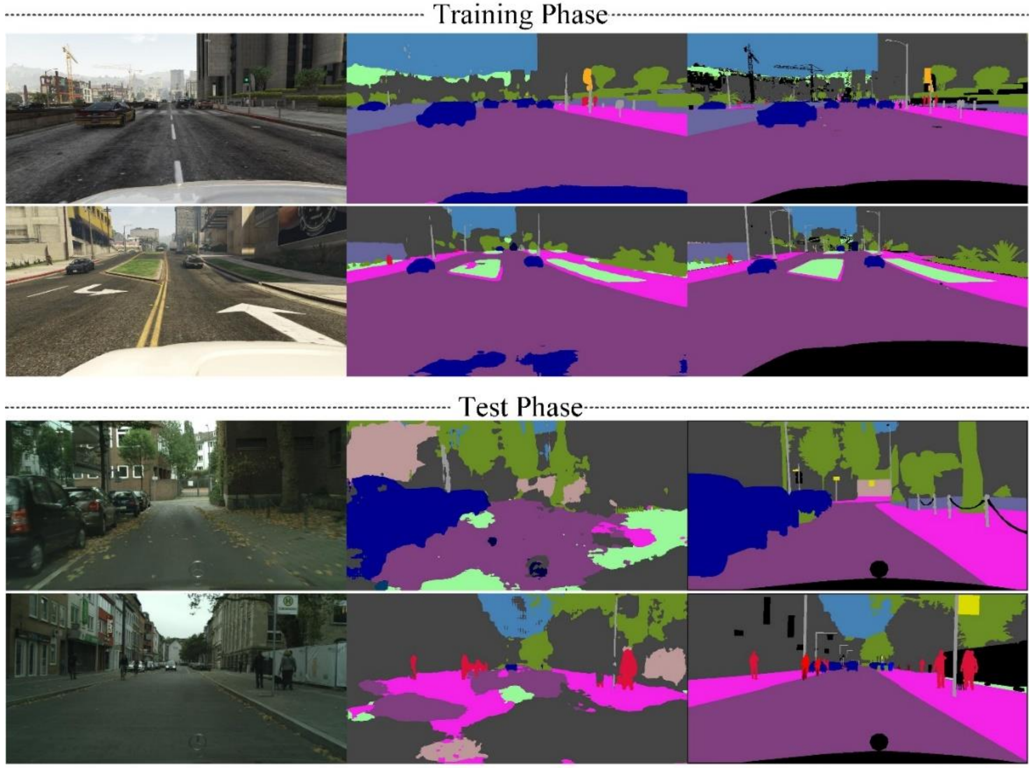
- In the image classification task (Source: MNIST and Target: MNIST-M):



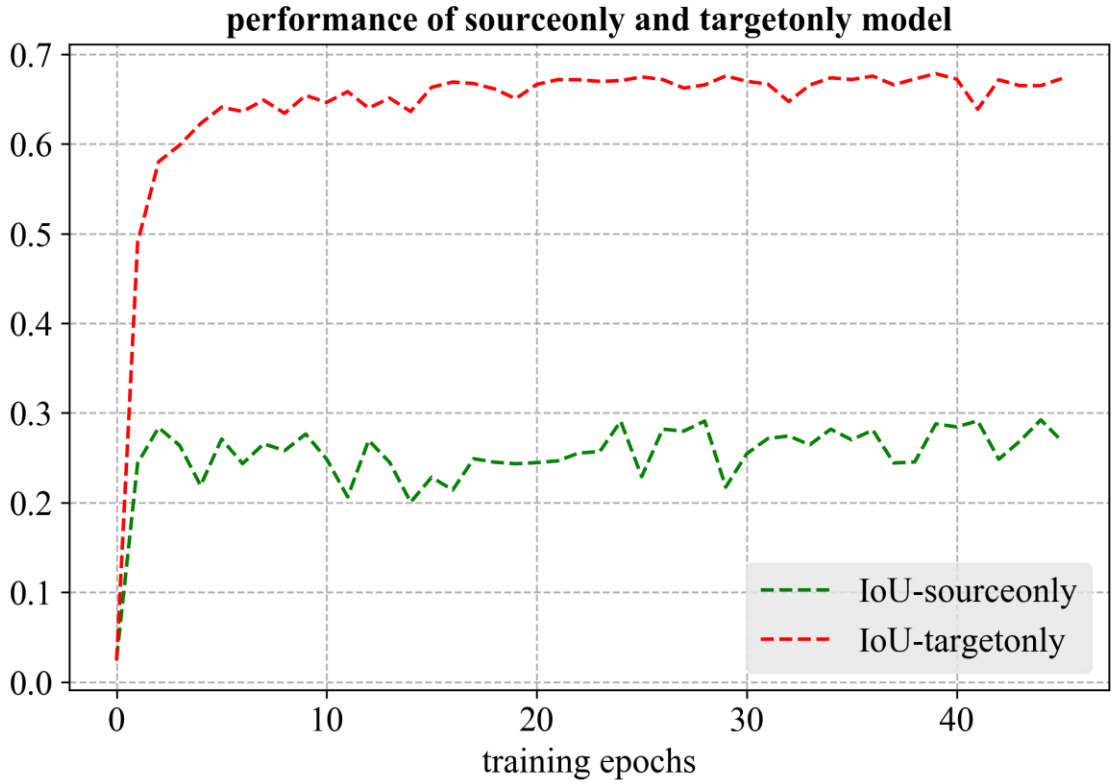
(a) MNIST (b) MNIST-M (c) Performance Comparison

Real Domain Adaptation Scenarios

- In the semantic segmentation task (Source: GTA5 and Target: Cityscapes):



(a) Input Image (b) Predicted Result (c) Ground Truth



Domain Adaptation Problem Settings

香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative
For The World

Source Data: $D_s = \{X_s, Y_s\}$ and Target Data: $D_t = \{X_t, Y_t\}$

Two Domains follow different data distributions, and the goal is to infer target labels as accurately as possible.

- If we have enough target labeled data: no need adaptation;
- If we have a few target labeled data: we can fine-tune the model trained on D_s with these target labeled data.
- If we have no labeled data, i.e., we only have $D_s = \{X_s, Y_s\}$ and $D_t = \{X_t\}$. This is referred to as **Unsupervised Domain Adaptation** (UDA), which is our main topic.

Domain Adaptation Benefits

Domain adaptation allows to adapt the model from a label-rich (domain) to a label-scarce (target) domain.

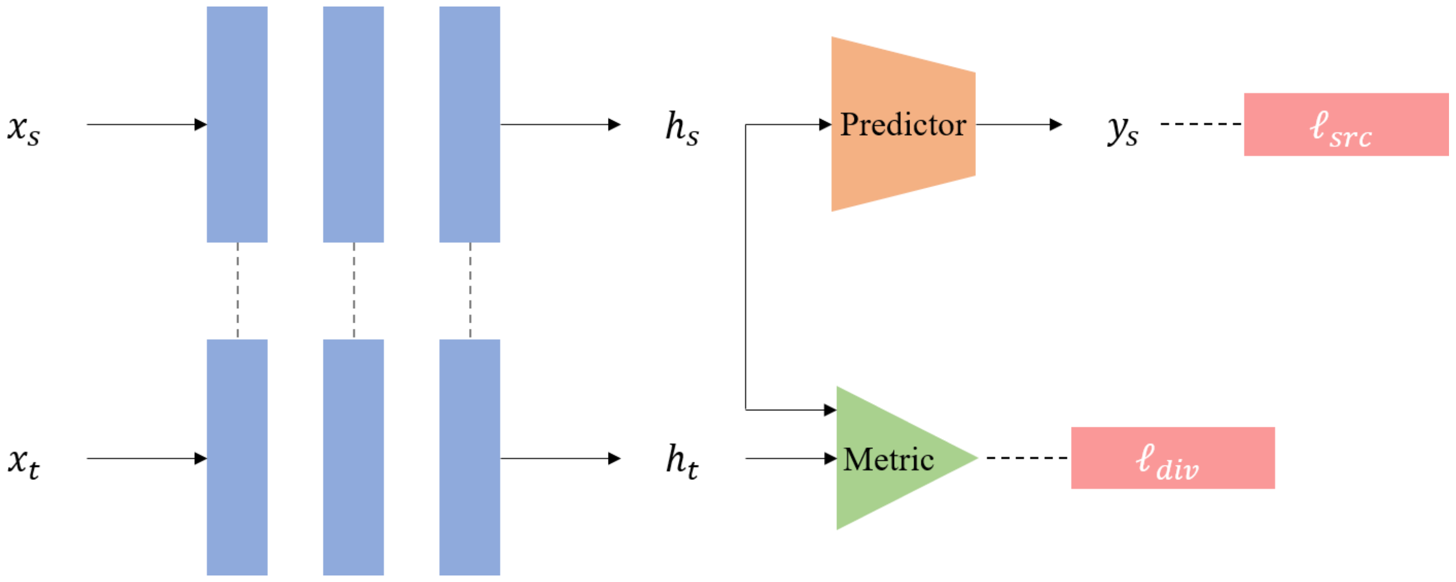
- Avoid learning from scratch.
- Transfer knowledge learnt from source data.
- Reduce the cost of annotation.

Outline

- Introduction to Domain Adaptation
- Main Idea for Recent Adaptation Works
- Proposed Works
    - Cross Domain Semantic Feature Learning via Adversarial Adaptation Networks
    - Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation
    - Generating Target Image-Label Pairs for Unsupervised Domain Adaptation
    - Model Adaptation: Unsupervised Domain Adaptation without Source Data
- Conclusions and Future Works

Main Idea for Domain Adaptation

- Reduce the domain discrepancy -- learning domain-invariant features (Feature-space alignment between two domains)

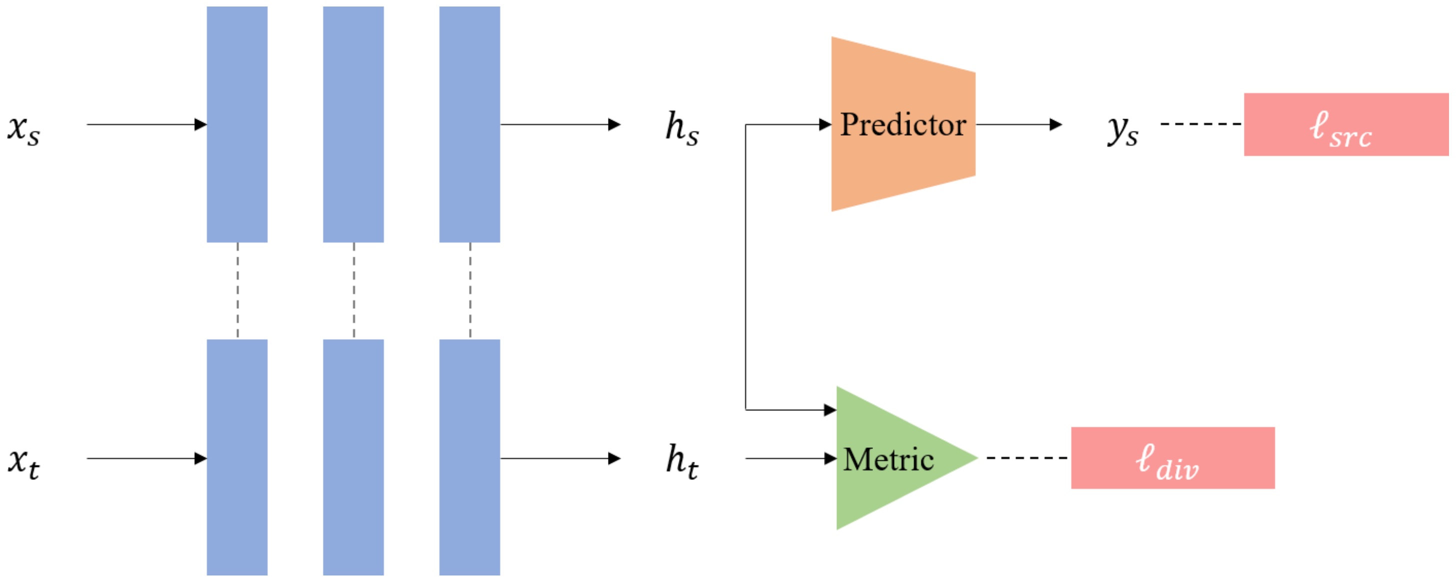


For feature extractor (blue blocks):

- Sharing layers or not
- Single or multiple layers

Main Idea for Domain Adaptation

- Reduce the domain discrepancy -- learning domain-invariant features (Feature-space alignment between two domains)



For distance metric (Green block):

- Statistics metrics, i.e., MMD, CORAL, etc.
- Adversarial training with an extra discriminator or classifier.

Main Idea for Domain Adaptation

香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative
For The World

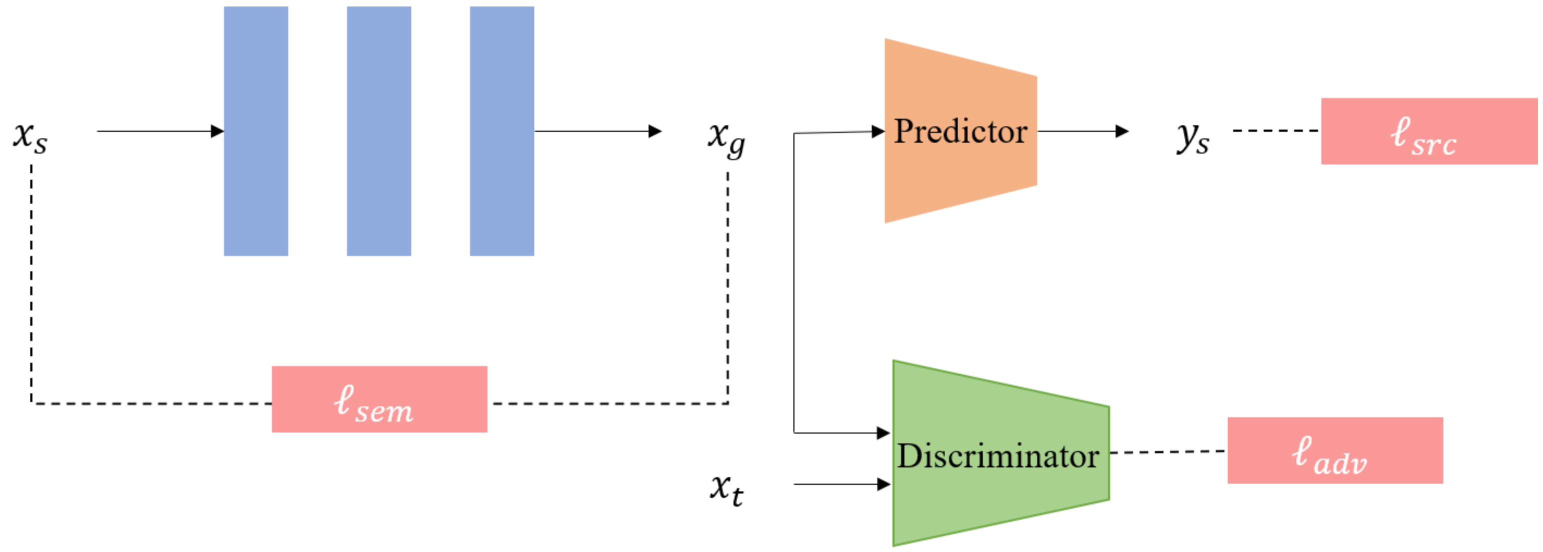
- Reduce the domain discrepancy -- Image-to-image translation (Pixel-space alignment between two domains)

  - Focus on generating reliable image-label pairs in the target domain.

  - The common tool is the **Generative Adversarial Networks** (GAN) for simulating the target distributions.

Main Idea for Domain Adaptation

- Reduce the domain discrepancy -- Image-to-image translation (Pixel-space alignment between two domains)

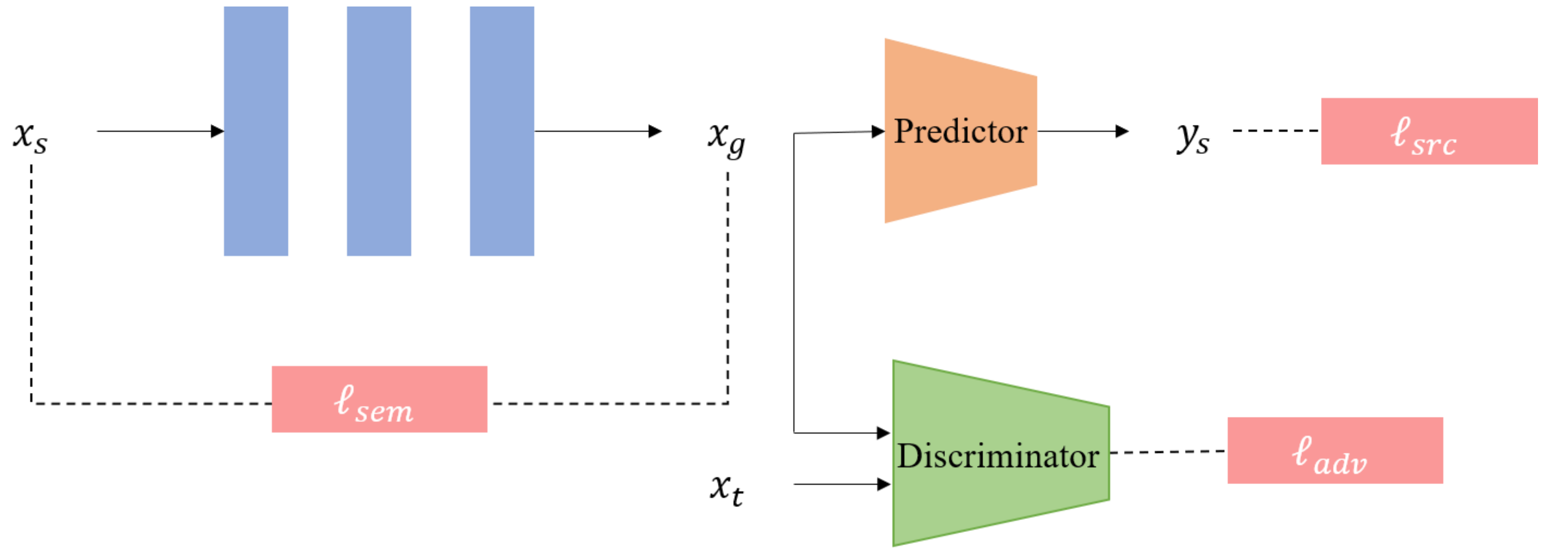


There are two requirements during the translation:

- Distribution matching: x_g and x_t should be undistinguishable by the discriminator.
- Semantic preservation: the semantic information between x_g and x_s should be the same.

Main Idea for Domain Adaptation

- Reduce the domain discrepancy -- Image-to-image translation (Pixel-space alignment between two domains)



These generative-based methods have two benefits:

- Interpretable: visualize x_g.
- Flexible: the predictor and the adaptor are decoupled.

Outline

- Introduction to Domain Adaptation
- Main Idea for Recent Adaptation Works
- Proposed Works
  - Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation
  - Generating Target Image-Label Pairs for Unsupervised Domain Adaptation
  - Model Adaptation: Unsupervised Domain Adaptation without Source Data
- Conclusions and Future Works

香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative
For The World

Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation

***Rui Li**, Wenming Cao, Qianfen Jiao, Si Wu, Hau-San Wong. Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation. Pattern Recognition Journal, PR, 2020.*

Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Motivation

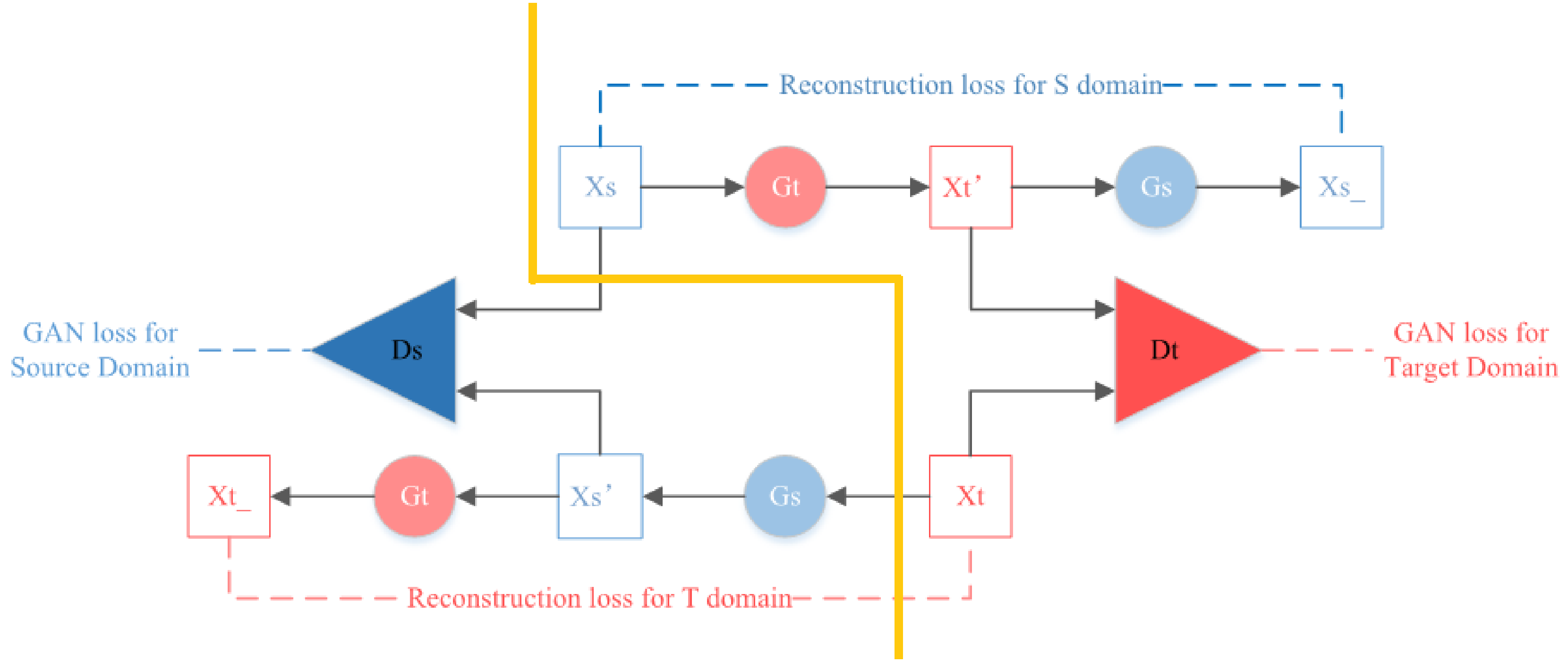
香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative
For The World

- Many feature-level alignment methods focus on the image classification, which may not be effective on the semantic segmentation which requires accurate spatial information.

- Therefore, recent works tend to image-to-image translation for domain adaptation, which is the pixel-level alignment.

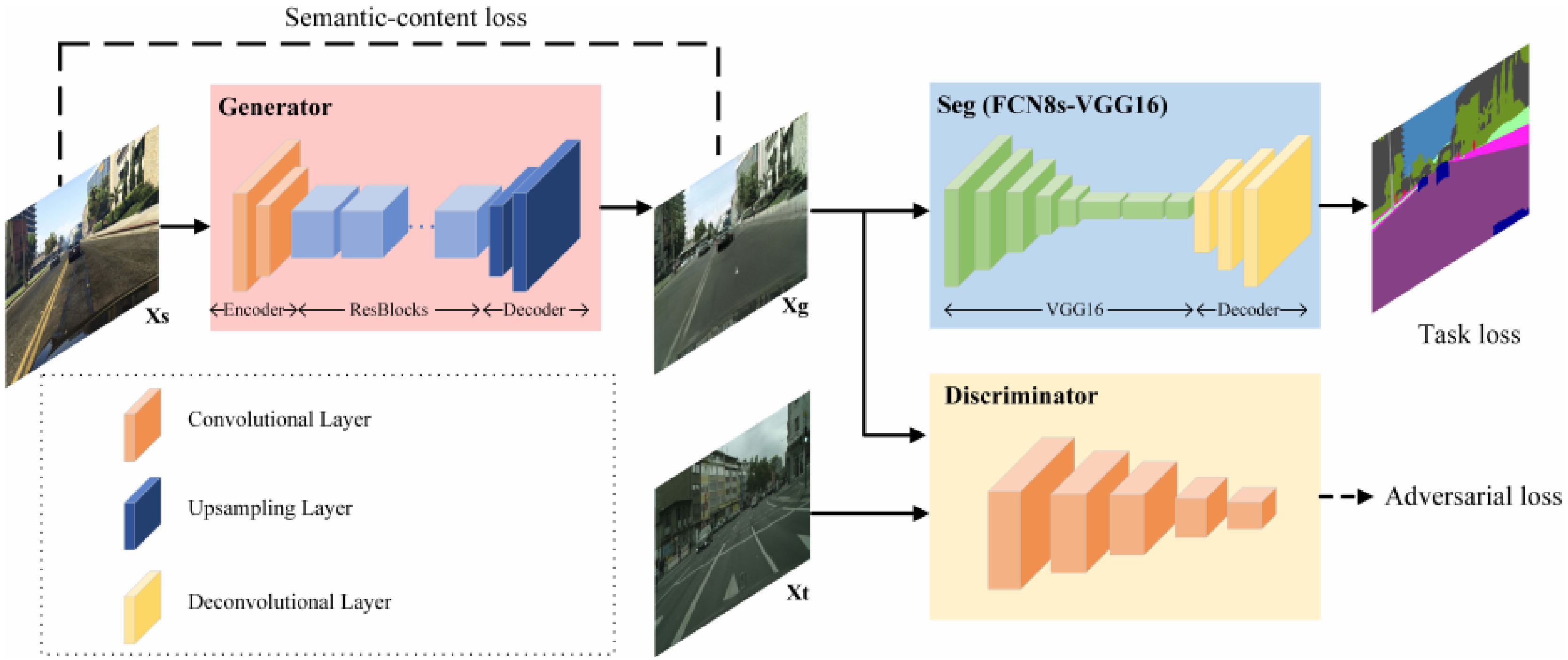
Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Motivation

- However, for <u>unsupervised high-resolution image-to-image translation</u>, Cycle-GAN is often used:



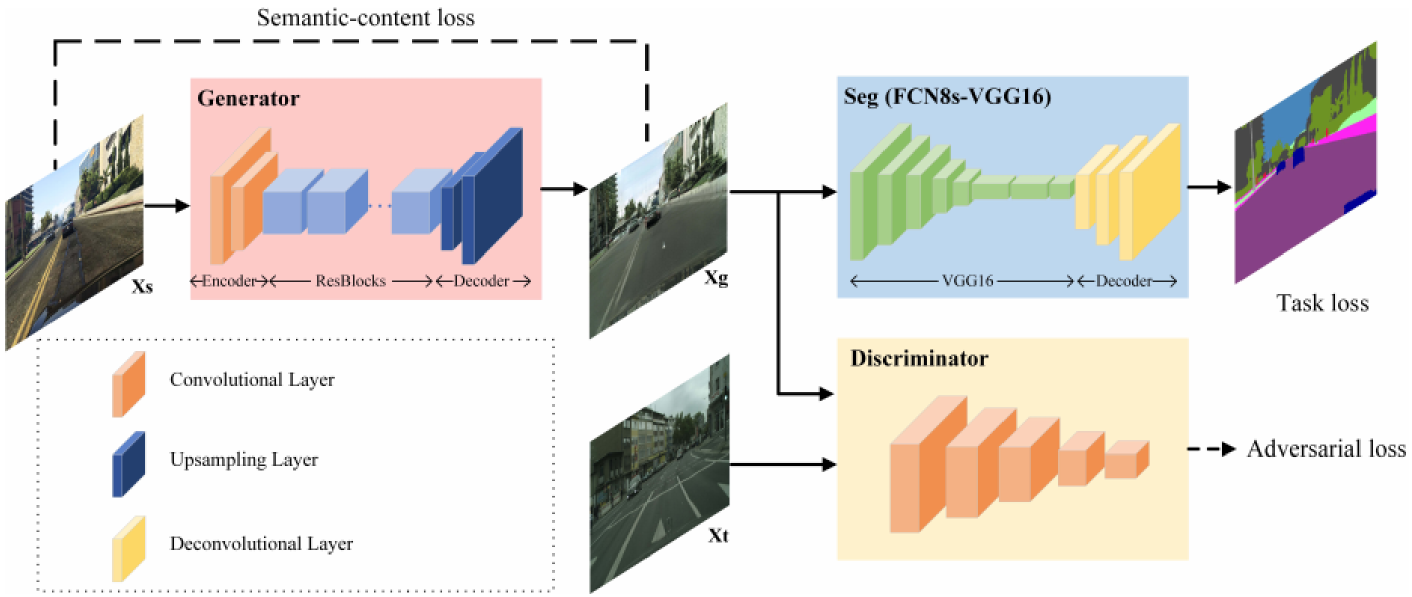
Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Motivation

- We aims to propose a simplified one-way translation method, designed for the semantic segmentation task:



Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Method

香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional·Creative
For The World

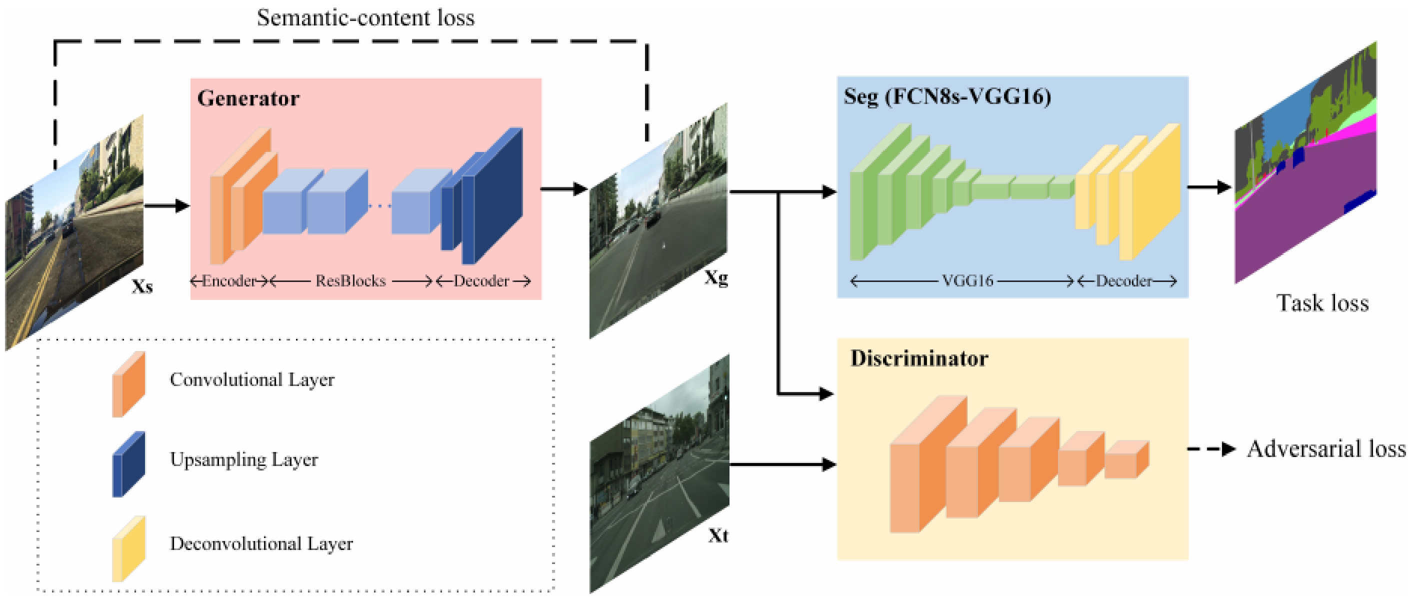


Adversarial Loss: aims to match the target distribution.

$$L_{adv}(D,G) = \mathbb{E}_{X_t}[\log D(X_t)] + \mathbb{E}_{X_s}[\log(1 - D(G(X_s)))]$$

$$\min_{G,Seg} \max_{D} L_{adv}(D,G) + L_{semCon}(G) + L_{task}(Seg)$$

Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Method



$$\min_{G,Seg} \max_{D} L_{adv}(D,G) + L_{semCon}(G) + L_{task}(Seg)$$

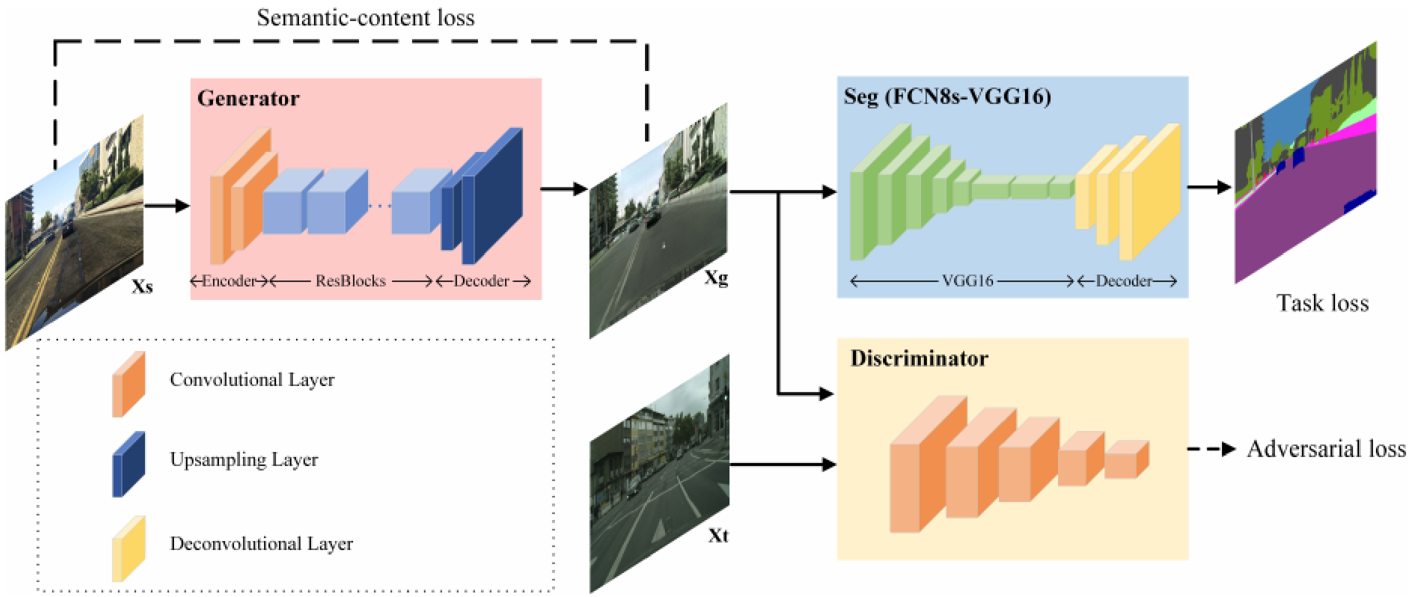
Semantic-Content Loss: aims to preserve the semantic and content information.

$$L_{semCon}(G) = L_{sem} + L_{con}$$

Where $L_{sem} = \mathbb{E}_{x_s}\mathcal{H}[Seg(G(X_s)), Y_s]$

$$L_{con} = \mathbb{E}_{x_s}[\|f(G(X_s)) - f(X_s)\|^2]$$

Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Method



Task Loss: computes cross entropy for each pixel with source and generated images.

$$L_{task}(Seg) = \mathbb{E}_{x_s}\{\mathcal{H}[Seg(X_s), Y_s] + \mathcal{H}[Seg(G(X_s)), Y_s]\}$$

$$\min_{G,Seg} \max_{D} L_{adv}(D,G) + L_{semCon}(G) + L_{task}(Seg)$$

Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Experiments

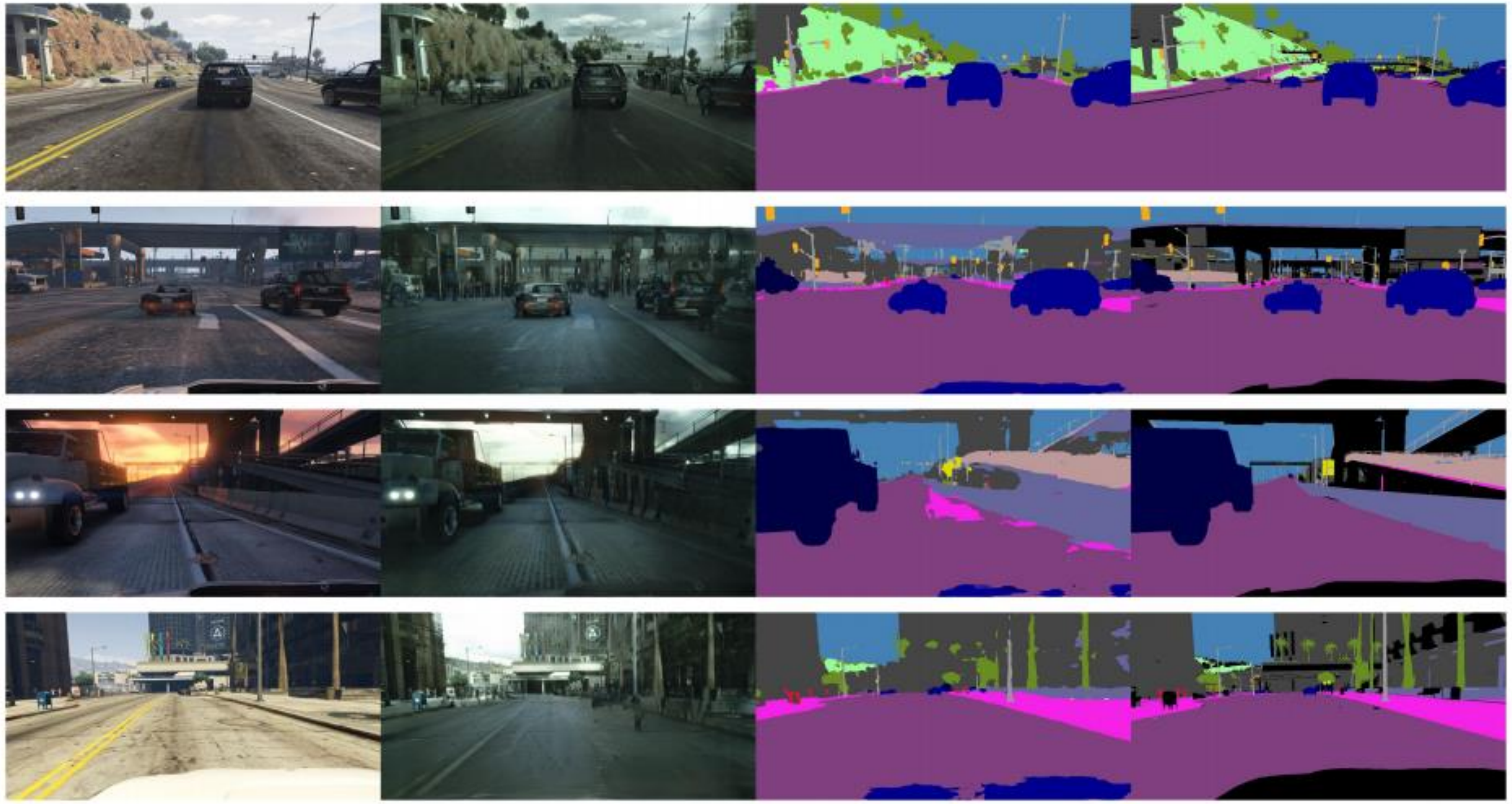
香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative For The World

- Dataset Settings: focus on adaptation from synthetic domain to real domain:
  - GTA5: is a large-scale urban scene dataset that includes 24966 synthetic high-quality images with pixel-level annotations. All these images and annotations are collected from Grand Theft Auto V (GTA5).
  - SYNTHIA: is a dataset with large well-annotated synthetic images collected from virtual worlds across various environments.
  - Cityscapes: is a real-world urban scene dataset that includes 2975 training samples and 500 validation samples with fine pixel-level annotation.

Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Experiments

香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative
For The World

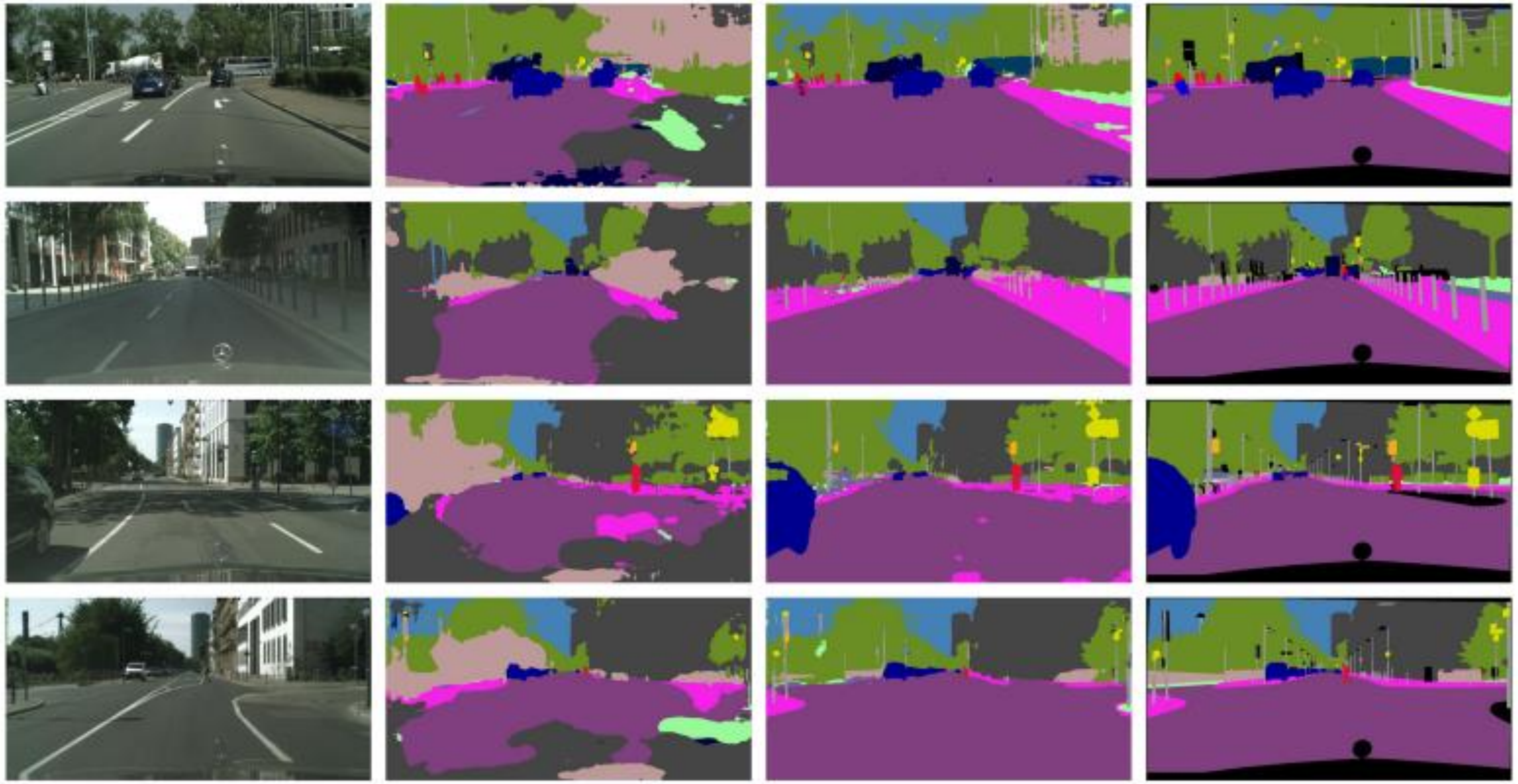
- GTA5 to Cityscapes (Generated images)



(a) Source Images (b) Generated Images (c) Segmentation Map (d) Source Labels

Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Experiments

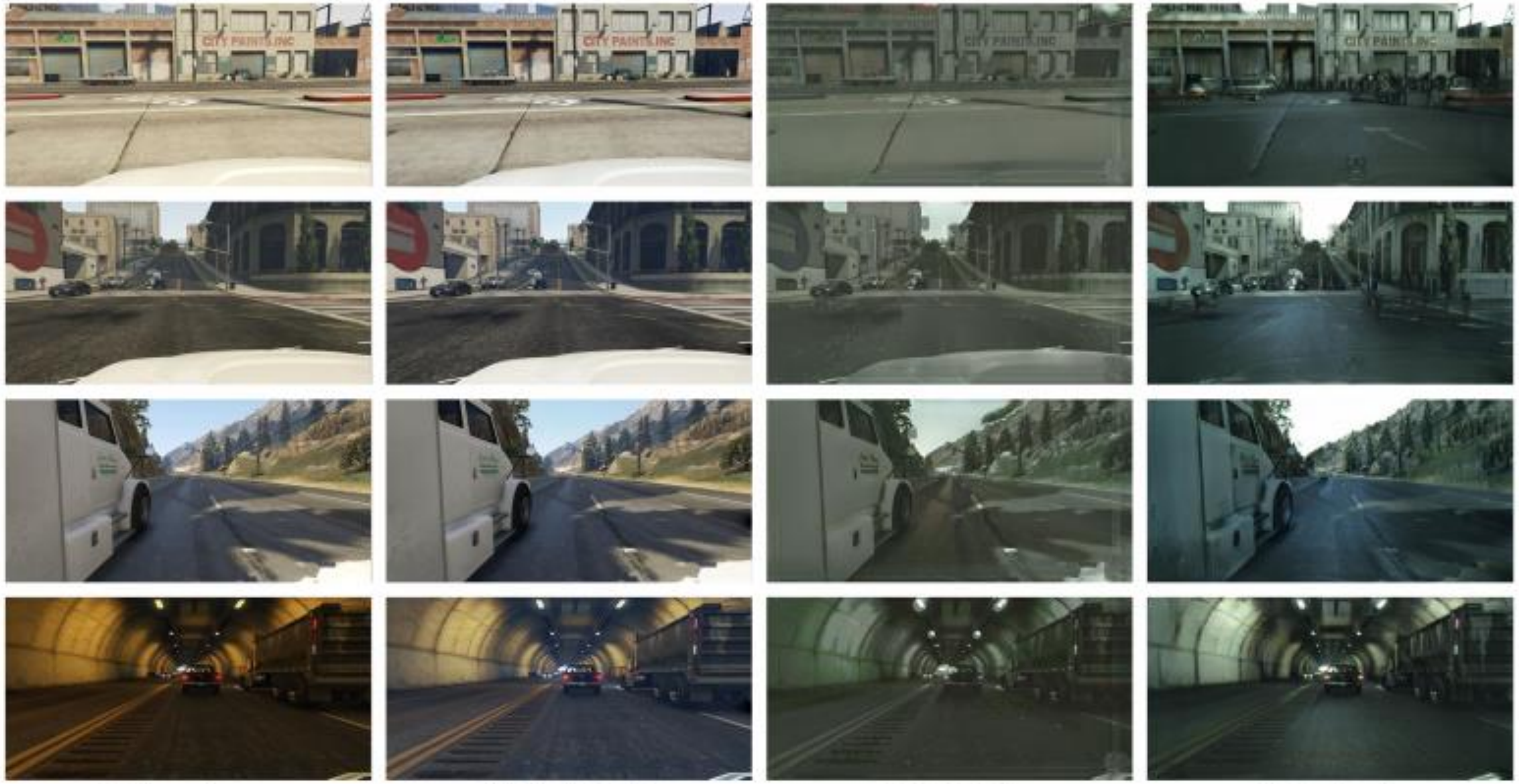
- GTA5 to Cityscapes (Comparison before and after adaptation)



(a) Test Images (b) SourceOnly Results (c) Ours Results (d) Ground Truth

Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Experiments

- GTA5 to Cityscapes (Generation comparison)



(a) Source Images (b) Vanilla GAN (c) CyCADA (d) Our Results

Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Experiments

- GTA5 to Cityscapes (Segmentation comparison)

Method	SourceOnly	FCNWild [42]	CDA [154]	MCD [109]	CBST [161]	OutputSpaceAdapt [128]	CyCADA [41]	DCAN [143]	SUIT	SUIT w/ Avg.W	SourceOnly	OutputSpaceAdapt [128]	DLOW [32]	SUIT	SUIT w/ Avg.W
Base Net	FCN8s-VGG16										Deeplab-v2				
mIoU (%)	26.4	27.1	28.9	28.8	30.9	35.0	35.4	36.2	38.9	40.6	34.2	41.4	42.3	43.7	**45.3**

Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation --- Experiments

香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative
For The World

- Transplant the pre-trained generator (with FCN-VGG16) to the other segmentation backbone.

Method	Seg Net	From GTA-5	From SYNTHIA
SourceOnly	Deeplab-v2 [11]	34.2	37.4
OutputSpaceAdapt [128]		41.4	45.9
SUIT		42.1	44.3
SUIT w/ Avg.W		**44.5**	**46.5**
SourceOnly	Deeplab-v3 [12]	35.2	38.8
SUIT		44.6	46.0
SUIT w/ Avg.W		**46.2**	**47.1**

Generating Target Image-Label Pairs for Unsupervised Domain Adaptation

***Rui Li,** Wenming Cao, Si Wu, Hau-San Wong. Generating Target Image-Label Pairs for Unsupervised Domain Adaptation. IEEE Transaction on Image Processing, TIP, 2020.*

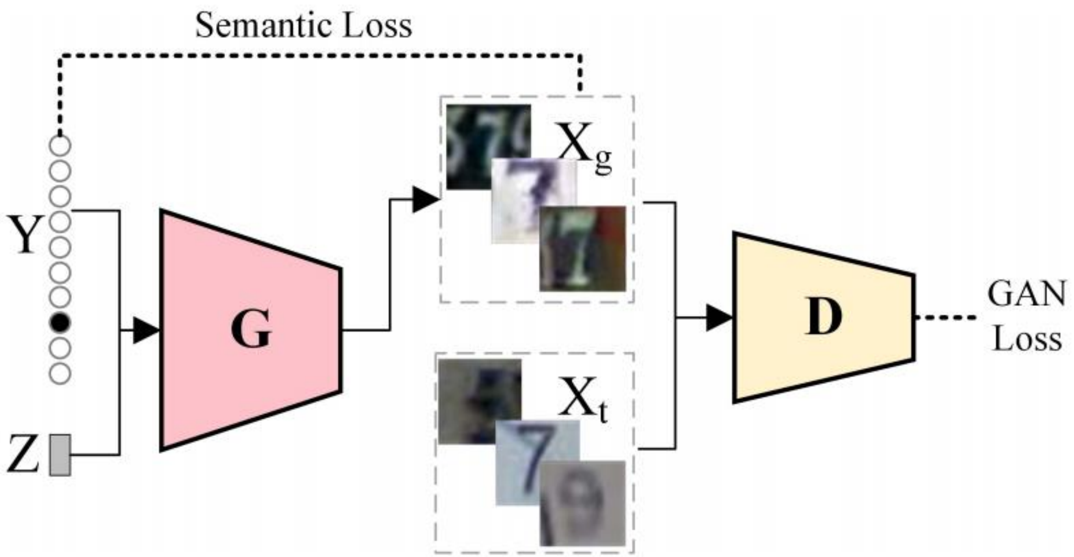
Generating Target Image-Label Pairs for Unsupervised Domain Adaptation --- Motivation

香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative
For The World

- Most domain adaptation works focus on reducing the discrepancy between two domains, which becomes difficult in case of the large domain gap.

- We propose for target generation from the shared label space, which can alleviate the large domain gap.

Generating Target Image-Label Pairs for Unsupervised Domain Adaptation --- Motivation



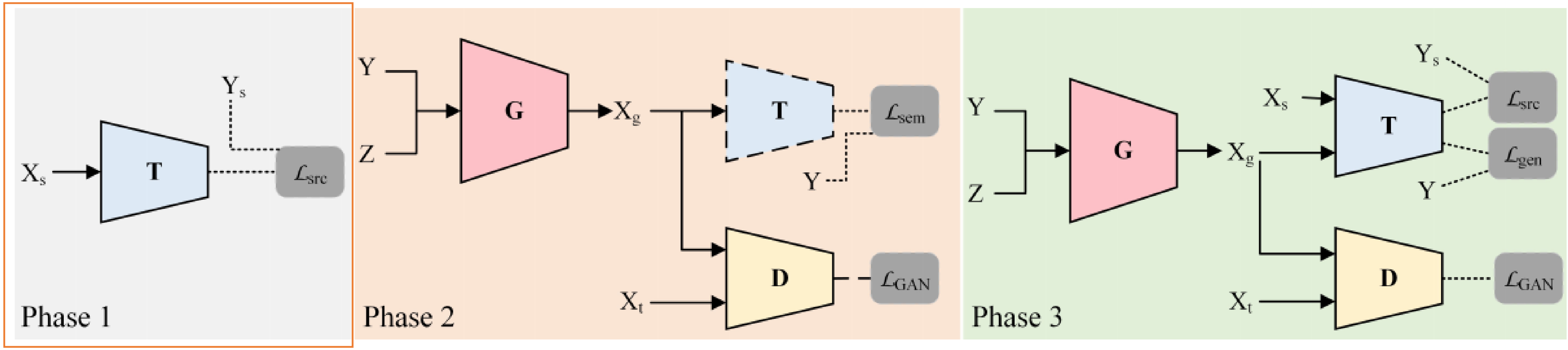
(a) Idea of label-to-image generation process

(b) Examples sampled from real target domain and generated by our model

Generating Target Image-Label Pairs for Unsupervised Domain Adaptation --- Method

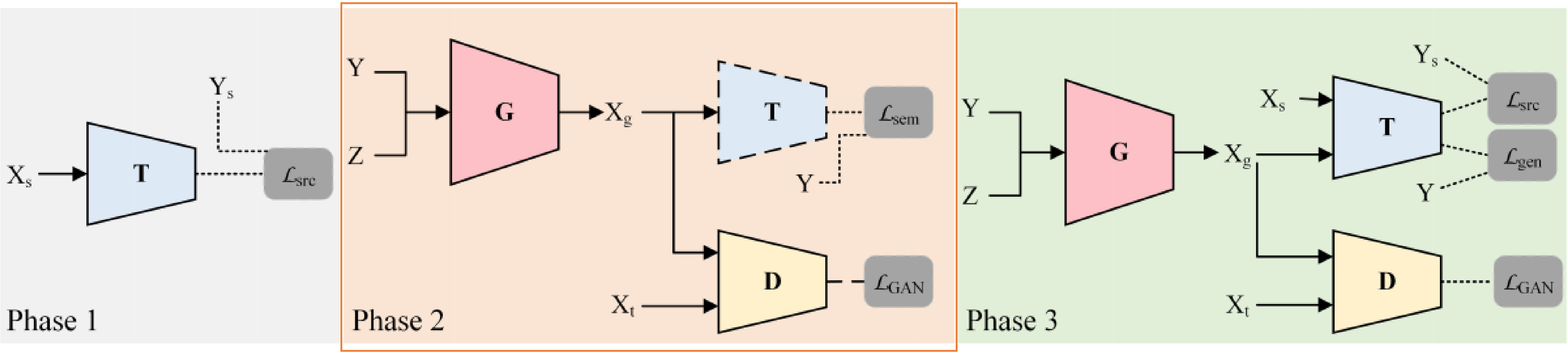
香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative For The World

- (Phase 1) Training the task prediction model (T) with the source data to obtain a weak T in the target domain.



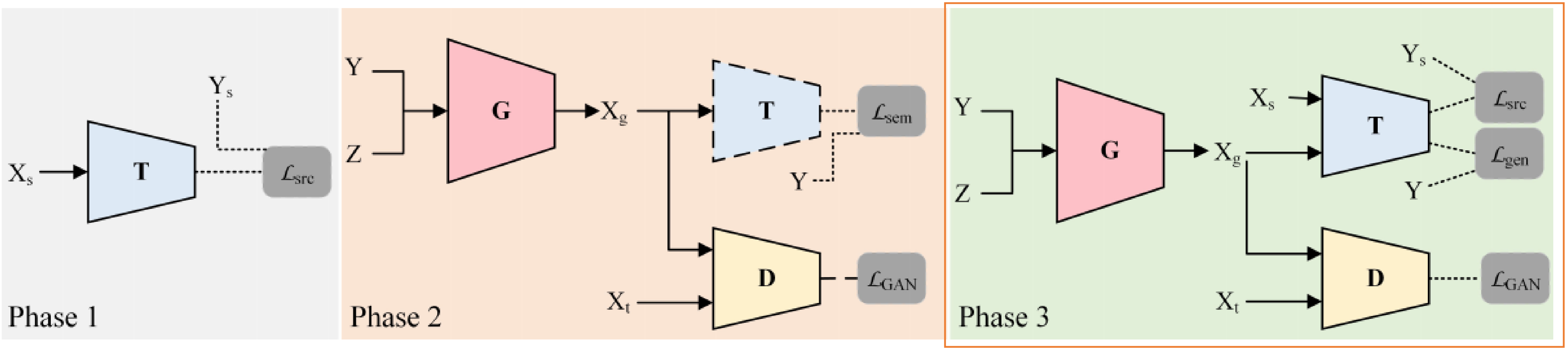
Generating Target Image-Label Pairs for Unsupervised Domain Adaptation --- Method

- (Phase 2) Training the generator G and the discriminator D in adversarial manner to obtain target training pairs.



Generating Target Image-Label Pairs for Unsupervised Domain Adaptation --- Method

- (Phase 3) Adding the generated data into training of T.



Generating Target Image-Label Pairs for Unsupervised Domain Adaptation --- Experiments

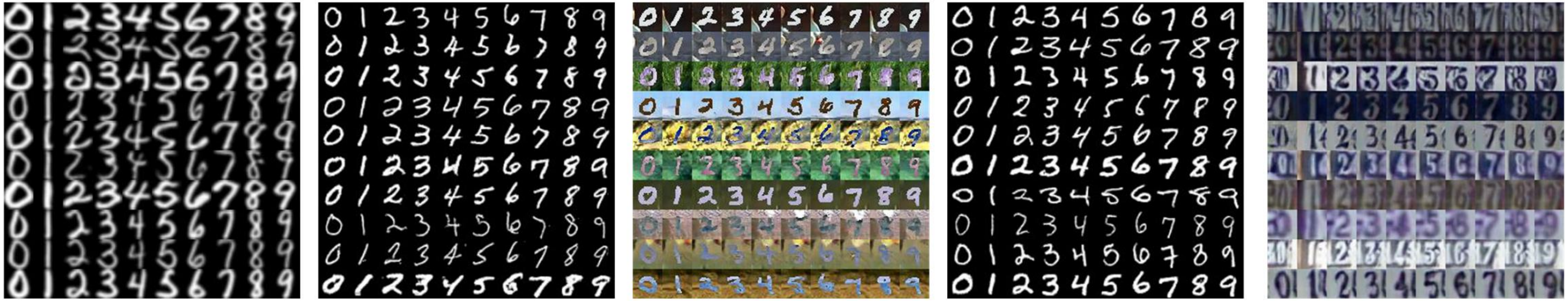
- Quantitative results.

Method	MNIST→USPS	USPS→MNIST	MNIST→MNIST-M	SVHN→MNIST	MNIST→SVHN
Not GAN-based Adaptation					
MMD [7]	81.1	-	76.9	71.1	-
CORAL [5]	81.7	-	57.7	63.1	-
DA$_{assoc}$ [33]	-	-	89.5	97.6	-
DANN [43]	85.1	73.0	77.4	73.9	35.7
ADDA [44]	89.4±0.2	90.1±0.8	-	76.0±1.8	-
ATT [10]	-	-	94.2	86.2	52.8
DSN [16]	91.3	-	83.2	82.7	-
DRCN [32]	91.8±0.1	73.7±0.1	-	82.0±0.2	40.1±0.1
VADA [45]	-	-	97.7	97.9	73.3
DIRT-T [45]	-	-	98.9	99.4	76.5
GAN-based Adaptation					
UNIT [53]	95.9	93.5	-	90.5	-
DuplexGAN [48]	96.01	98.75	-	92.46	62.65
GenToAdapt [47]	95.3±0.7	90.8±1.3	-	92.4±0.9	-
CoGAN [52]	91.2±0.8	89.1±0.8	62.0	-	-
PixelDA [12]	95.9	-	98.2	-	-
SBADA-GAN C_t [55]	96.7	94.4	99.1	72.2	59.2
SBADA-GAN [55]	97.6	95.0	**99.4**	76.1	61.1
CyCADA pixel only [26]	95.6±0.2	96.4±0.1	-	70.3±0.2	-
CyCADA pixel+feat [26]	95.6±0.2	96.5±0.1	-	90.4±0.4	-
Our results					
SourceOnly	92.4±1.7	86.1±1.3	54.2±0.9	76.4±1.5	57.3±2.1
Label2Image-DA	96.9±0.5	98.9±0.1	97.1±0.2	99.0±0.1	-
Label2Image-DA with $\mathcal{L}_{reg}$	**98.1**±0.3	**99.4**±0.1	99.2±0.1	**99.5**±0.03	**91.3**±0.2

Generating Target Image-Label Pairs for Unsupervised Domain Adaptation --- Experiments

- Generation results.

Each column shares the same label and each row shares the same noise variable

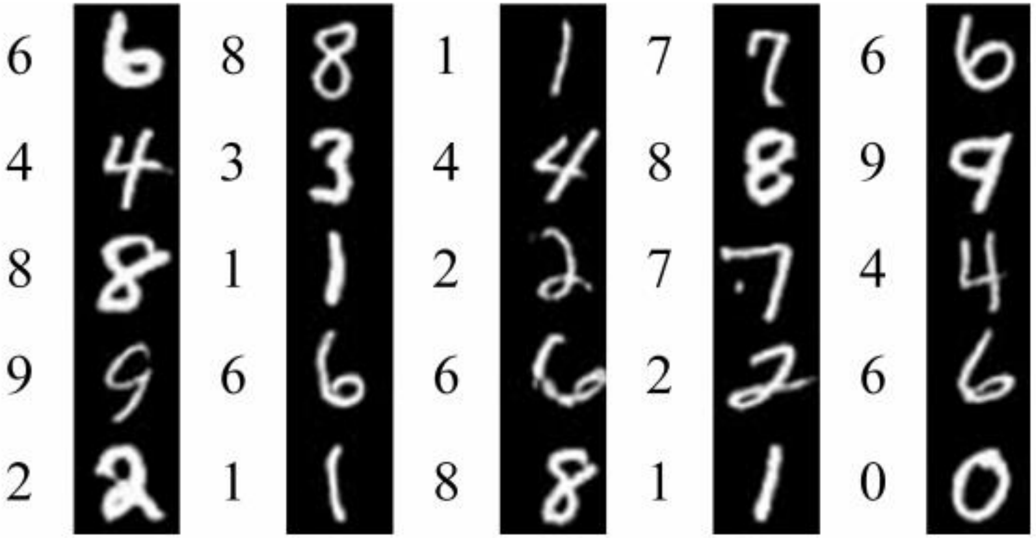


(a) MNIST → USPS (b) USPS → MNIST (c) MNIST → MNIST-M (d) SVHN → MNIST (d) MNIST → SVHN

Generating Target Image-Label Pairs for Unsupervised Domain Adaptation --- Experiments

- Comparison between Image-to-Image Translation and Label-to-Image Translation.

(a) Image-to-Image Translation



(b) Label-to-Image Translation

Generating Target Image-Label Pairs for Unsupervised Domain Adaptation --- Experiments

- Our model is also applicable for the semantic segmentation task (Generation).

(a) Input Labels (b) Generated Images (target-style) (c) Corresponding Source Images

Generating Target Image-Label Pairs for Unsupervised Domain Adaptation --- Experiments

- Our model is also applicable for the semantic segmentation task (Segmentation).

Method	SourceOnly	FCNWild [42]	CDA [154]	MCD [109]	CyCADA pixel-only [41]	CyCADA [41]	**Label2Image-DA**	SourceOnly	CycleGAN [159]	OutputSpaceAdapt [128]	DLOW [32]	**Label2Image-DA**
Base Net	FCN8s-VGG16							Deeplab-v2				
mIoU (%)	26.4	27.1	28.9	28.8	34.8	35.4	**37.5**	34.2	41.0	41.4	42.3	**43.8**

Model Adaptation: Unsupervised Domain Adaptation without Source Data

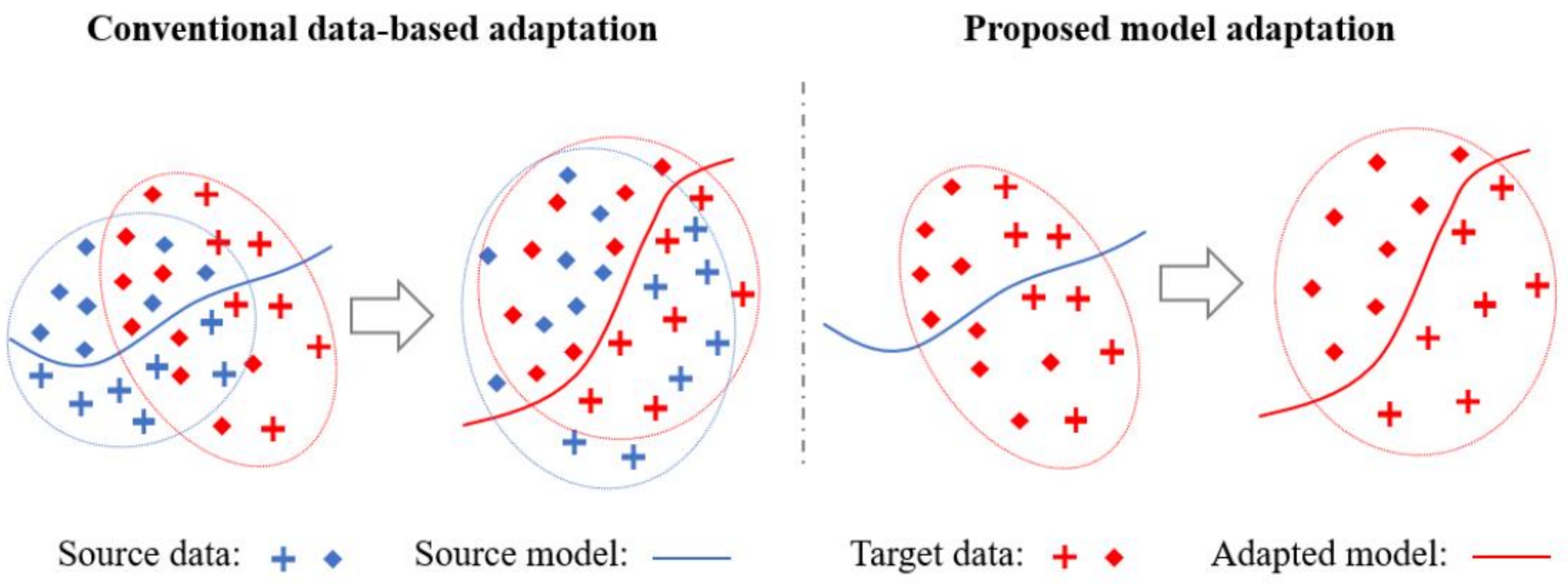
Rui Li, *Qianfen Jiao, Wenming Cao, Hau-San Wong, Si Wu. Model Adaptation: Unsupervised Domain Adaptation without Source Data. IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2020.*

Model Adaptation: Unsupervised Domain Adaptation without Source Data --- Motivation

- Considering source data are not always available:
  - For many companies, they will only provide the learned models instead of their customer data due to the data privacy and security issues.
  - The source datasets like videos or high-resolution images may be so large that it is not practical or convenient to transfer to different platforms.

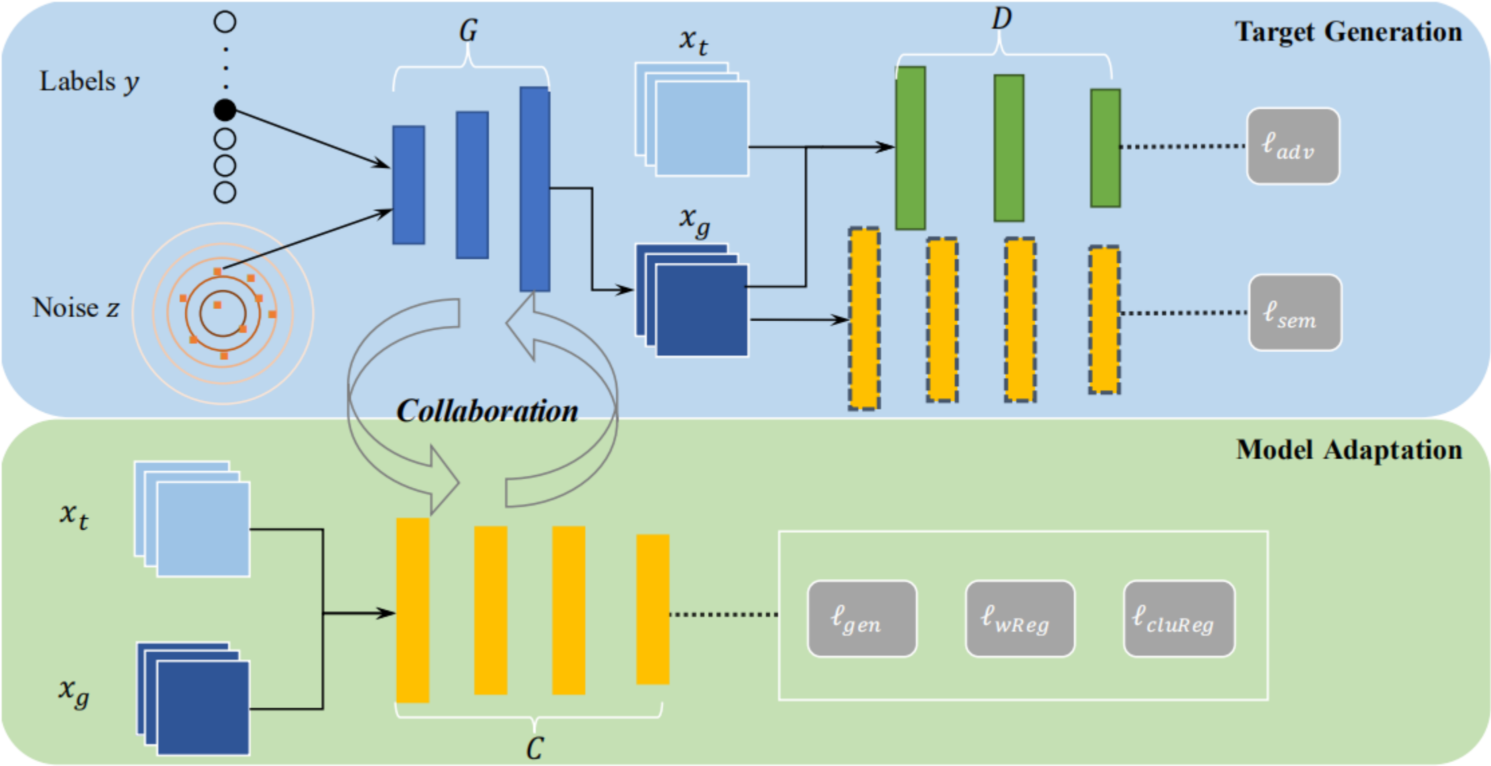
- We focus on a more challenging setting – Model Adaptation, where we only have the pre-trained source model and the unlabeled target dataset during adaptation process.

Model Adaptation: Unsupervised Domain Adaptation without Source Data --- Motivation



Model Adaptation: Unsupervised Domain Adaptation without Source Data --- Method

- Collaborative Class Conditional GAN (Generation)

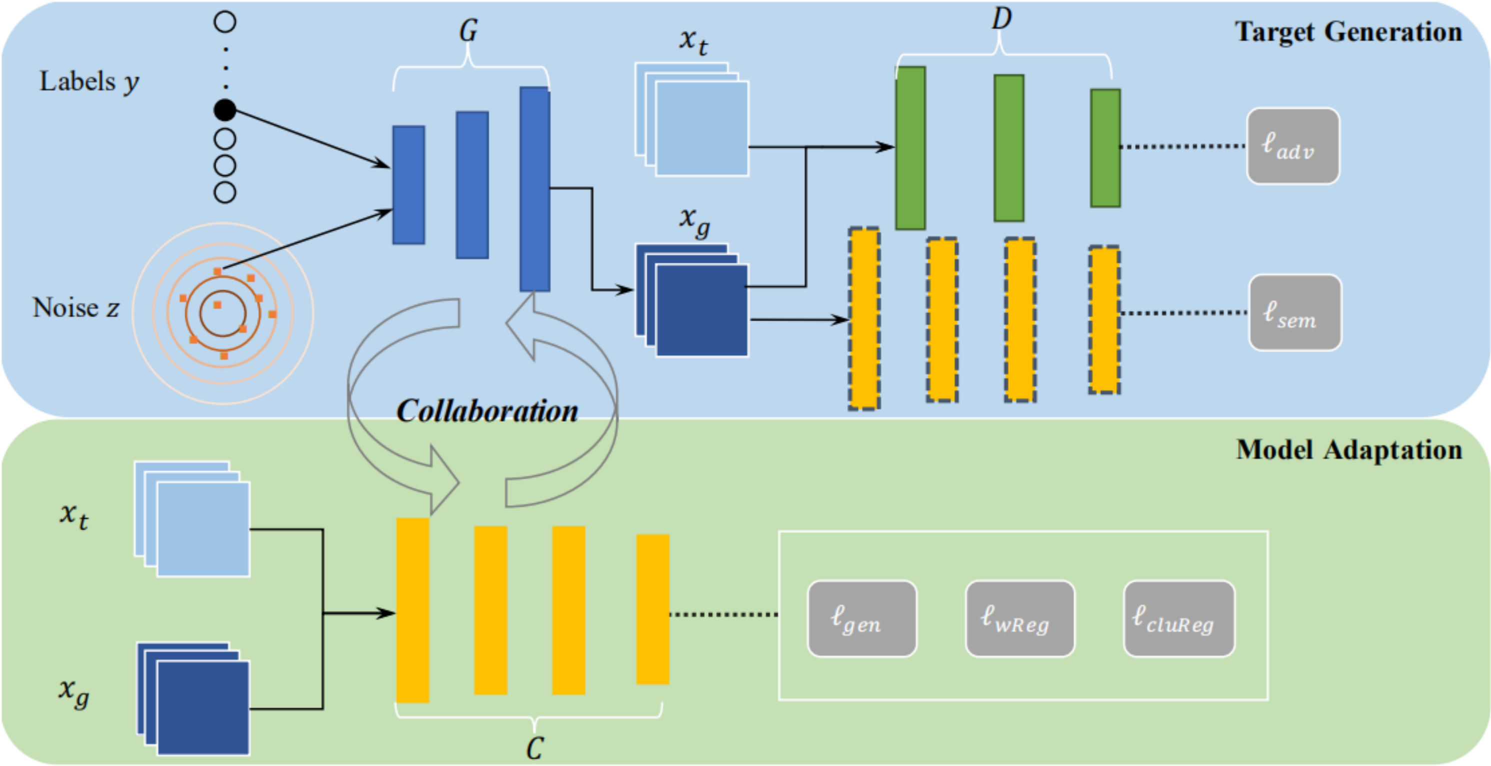


$$\min_G \max_D \ell_{adv} + \ell_{sem}$$

$$\ell_{sem}(G) = \mathbb{E}_{y,z}[-y \log C(G(y,z))]$$

Model Adaptation: Unsupervised Domain Adaptation without Source Data --- Method

- Collaborative Class Conditional GAN (Adaptation)



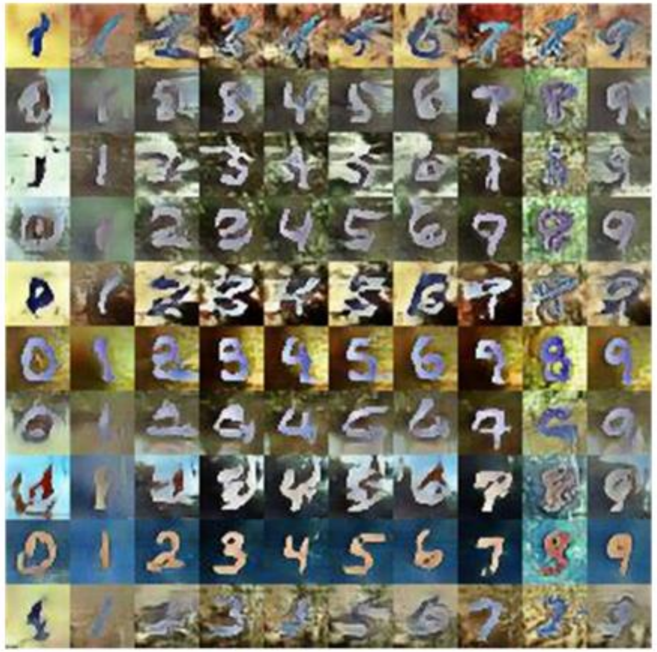
$$\ell_{gen} = \mathbb{E}_{x_g,y}\left[-y \log C(x_g)\right]$$

$$\ell_{wReg} = \left\|\theta_C - \theta_{C_s}\right\|$$

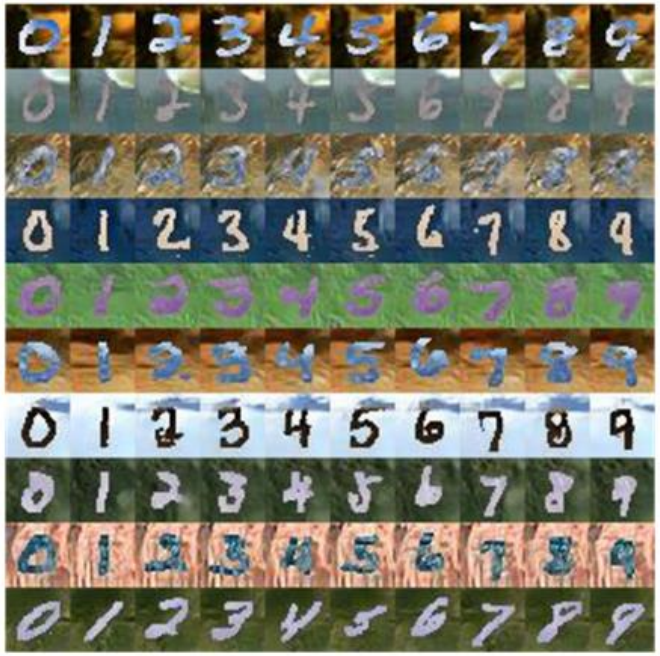
$$\ell_{cluReg} = \mathbb{E}_{x_t}\left[-C(x_t)\log C(x_t)\right] + \left[\mathrm{KL}\left(C(x_t)\|C(x_t+\tilde{r})\right)\right]$$

Model Adaptation: Unsupervised Domain Adaptation without Source Data --- Experiments

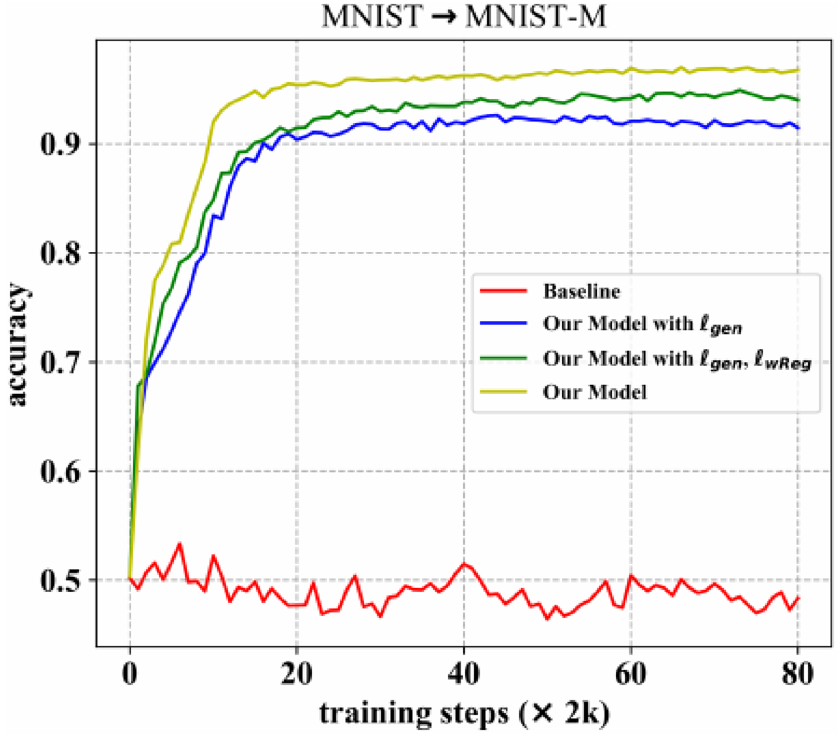
- Collaborative behavior



(a) Before Adaptation



(b) After Adaptation



Model Adaptation: Unsupervised Domain Adaptation without Source Data --- Experiments

- Adaptation results from synthetic domain to real domain

Method	Syn.Digits → SVHN	Syn.Sign → GTSRB
DANN	91.1	88.7
DSN	91.2	93.1
AssocDA	91.8	97.6
VADA	94.8	98.8
DIRT-T	**96.1**	99.5
Source-Only	86.2	78.3
3C-GAN (Ours)	95.9	**99.6**

Model Adaptation: Unsupervised Domain Adaptation without Source Data --- Experiments

- Visualization of the generator

(a) Syn.Digits→SVHN

(b) Syn.Sign→GTSRB

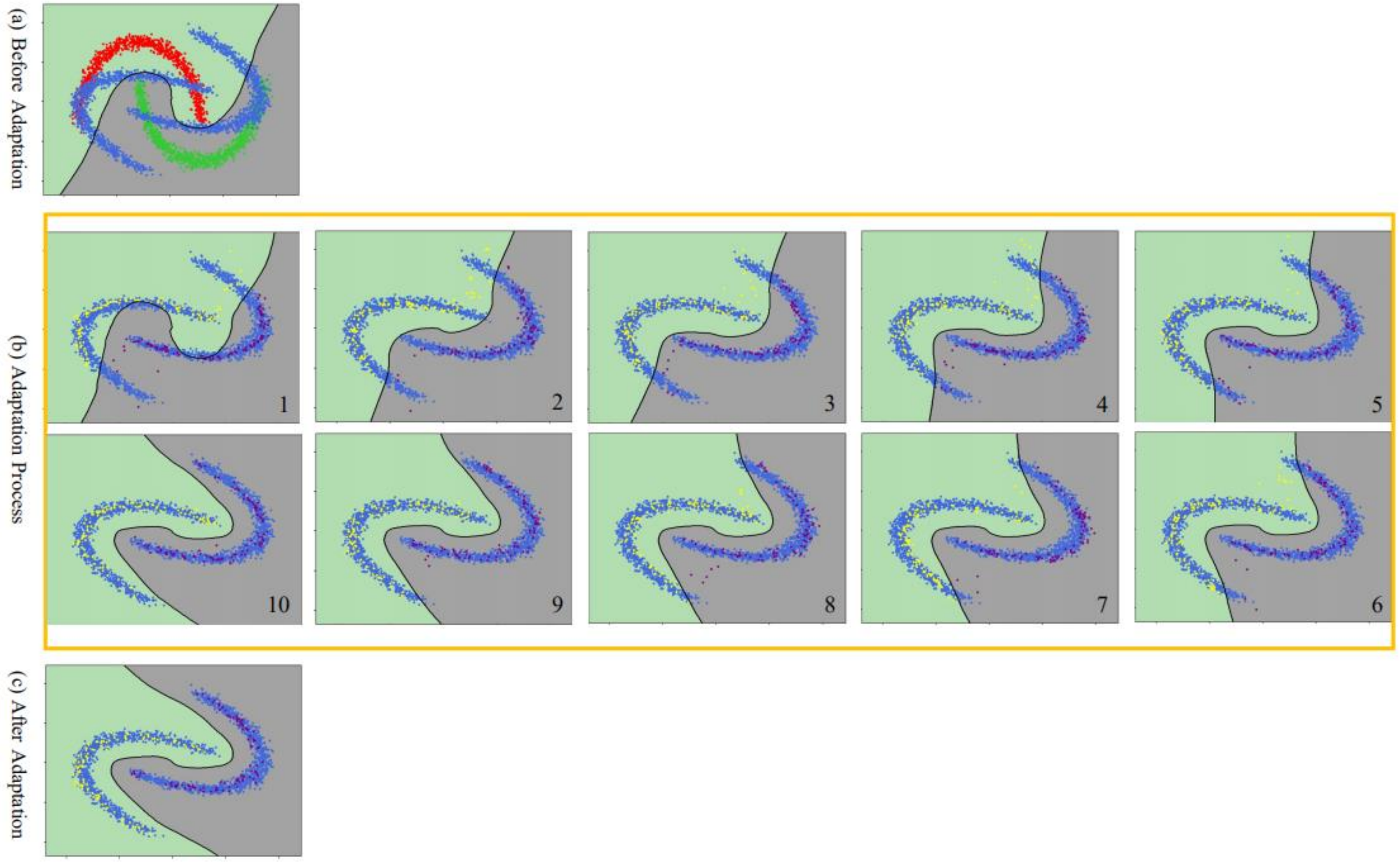
Model Adaptation: Unsupervised Domain Adaptation without Source Data --- Experiments

- Adaptation results in Office31 dataset (based on ResNet50)

Method	A $\downarrow$ W	D $\downarrow$ W	W $\downarrow$ D	A $\downarrow$ D	D $\downarrow$ A	W $\downarrow$ A	Avg
ResNet50 [40]	68.4±0.2	96.7±0.1	99.3±0.1	68.9±0.2	65.2±0.3	60.7±0.3	76.1
DAN [72]	80.5±0.4	97.1±0.2	99.6±0.1	78.6±0.2	63.6±0.3	62.8±0.2	80.4
RTN [75]	84.5±0.2	96.8±0.1	99.4±0.1	77.5±0.3	66.2±0.2	64.8±0.3	81.6
DANN [29]	82.6±0.4	96.9±0.2	99.3±0.2	81.5±0.4	68.4±0.5	67.5±0.5	82.7
ADDA [131]	86.2±0.5	96.2±0.3	98.4±0.3	77.8±0.3	69.5±0.4	68.9±0.5	82.9
JAN [74]	86.0±0.4	96.7±0.3	99.7±0.1	85.1±0.4	69.2±0.4	70.7±0.5	84.6
MADA [92]	90.0±0.2	97.4±0.1	99.6±0.1	87.8±0.2	70.3±0.3	66.4±0.3	85.2
GenToAdapt [110]	89.5±0.5	97.9±0.3	99.8±0.2	87.7±0.5	72.8±0.3	71.4±0.4	86.5
Our Model	**93.7**±0.2	**98.5**±0.1	**99.8**±0.2	**92.7**±0.4	**75.3**±0.5	**77.8**±0.1	**89.6**

Model Adaptation: Unsupervised Domain Adaptation without Source Data --- Experiments

- Adaptation process with a toy example



Outline

- Introduction to Domain Adaptation
- Main Idea for Recent Adaptation Works
- Proposed Works
  - Simplified Unsupervised Image Translation for Semantic Segmentation Adaptation
  - Generating Target Image-Label Pairs for Unsupervised Domain Adaptation
  - Model Adaptation: Unsupervised Domain Adaptation without Source Data
- Conclusions and Future Works

Conclusions

- We propose a simplified one-way translation method for flexible adaptation on semantic segmentation.
- We introduce a Label2ImageDA approaches to avoid cross-domain discrepancy reduction for challenging adaptation tasks.
- We develop a collaborative learning scheme with GAN, which enables unsupervised domain adaptation without source data.