SimT: Handling Open-set Noise for Domain Adaptive Semantic Segmentation

Xiaoqing Guo[1] **Jie Liu[1]** **Tongliang Liu[2]** **Yixuan Yuan[1†]**

[1] City University of Hong Kong **[2] The University of Sydney**

xqguo.ee@my.cityu.edu.hk jliu.ee@my.cityu.edu.hk tongliang.liu@sydney.edu.au yxyuan.ee@cityu.edu.hk

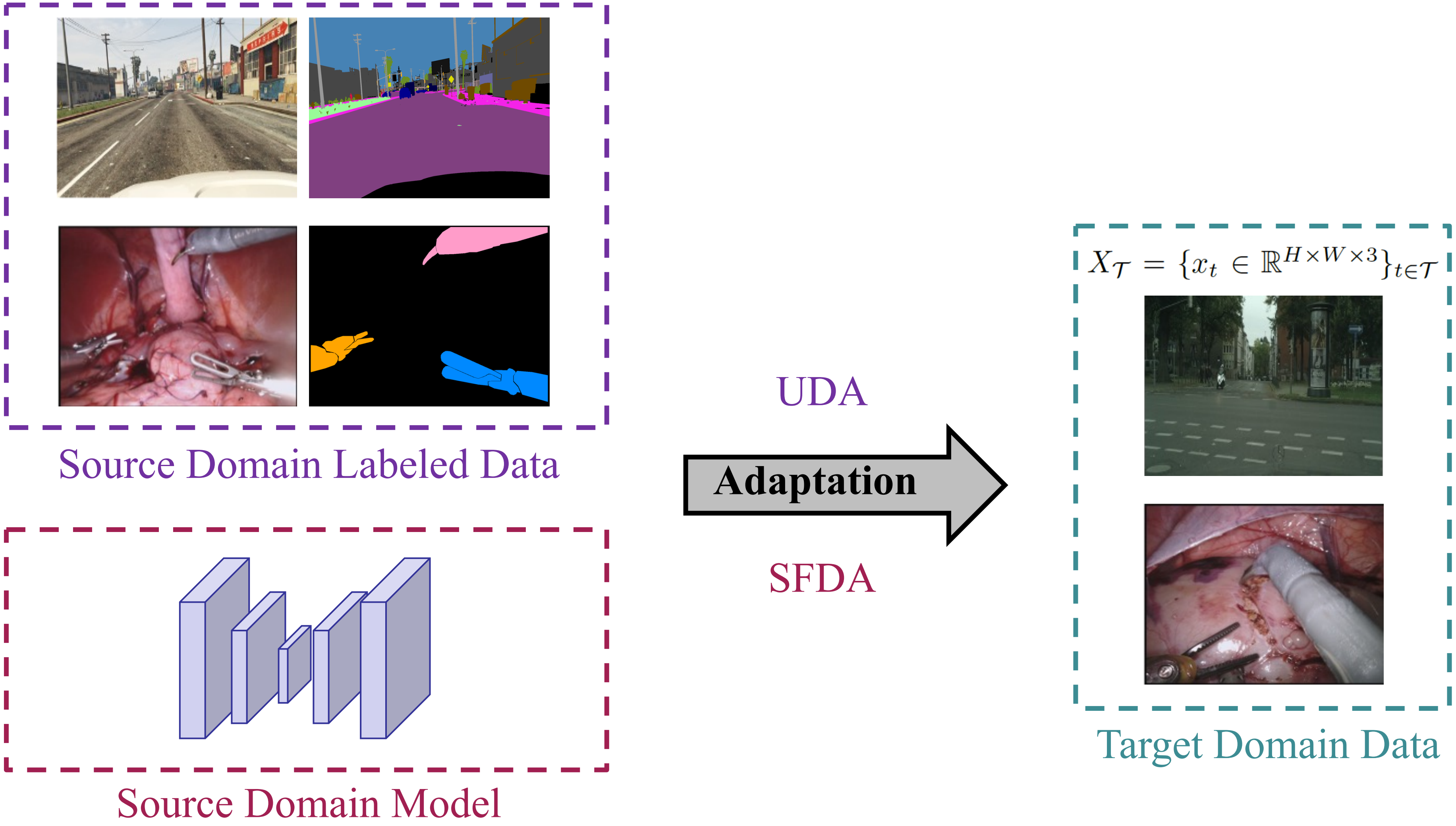
† Yixuan Yuan is the corresponding author

Content

- **Introduction**
  - Domain adaptation (DA) in semantic segmentation
  - A general and practical DA Task
  - Major challenge of DA task
  - Major method – self-training (ST) strategy
  - Major limitations of ST
  - Problem definition of ST in DA
  - Our solution - simplex noise transition matrix (SimT, T)
  - Problem reformulation
- **Method**
  - Geometry analysis of SimT
  - Volume minimization for estimating SimT
  - Three proposed regularizations for estimating SimT
- **Experiments**
  - Dataset
  - Quantitative results
  - Qualitative results
- **Summarization**

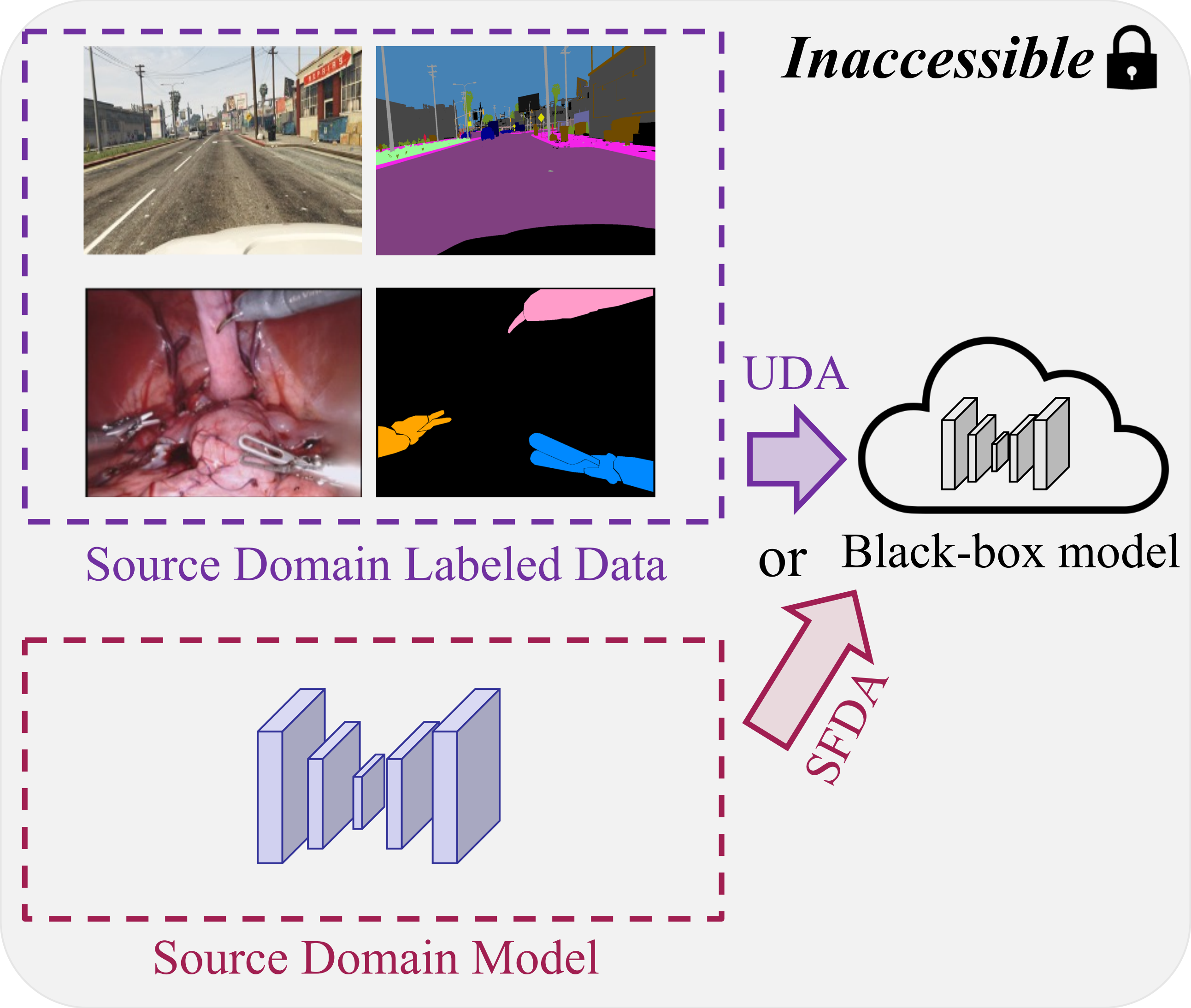
Introduction

- **Domain Adaptation (DA) in Semantic Segmentation**



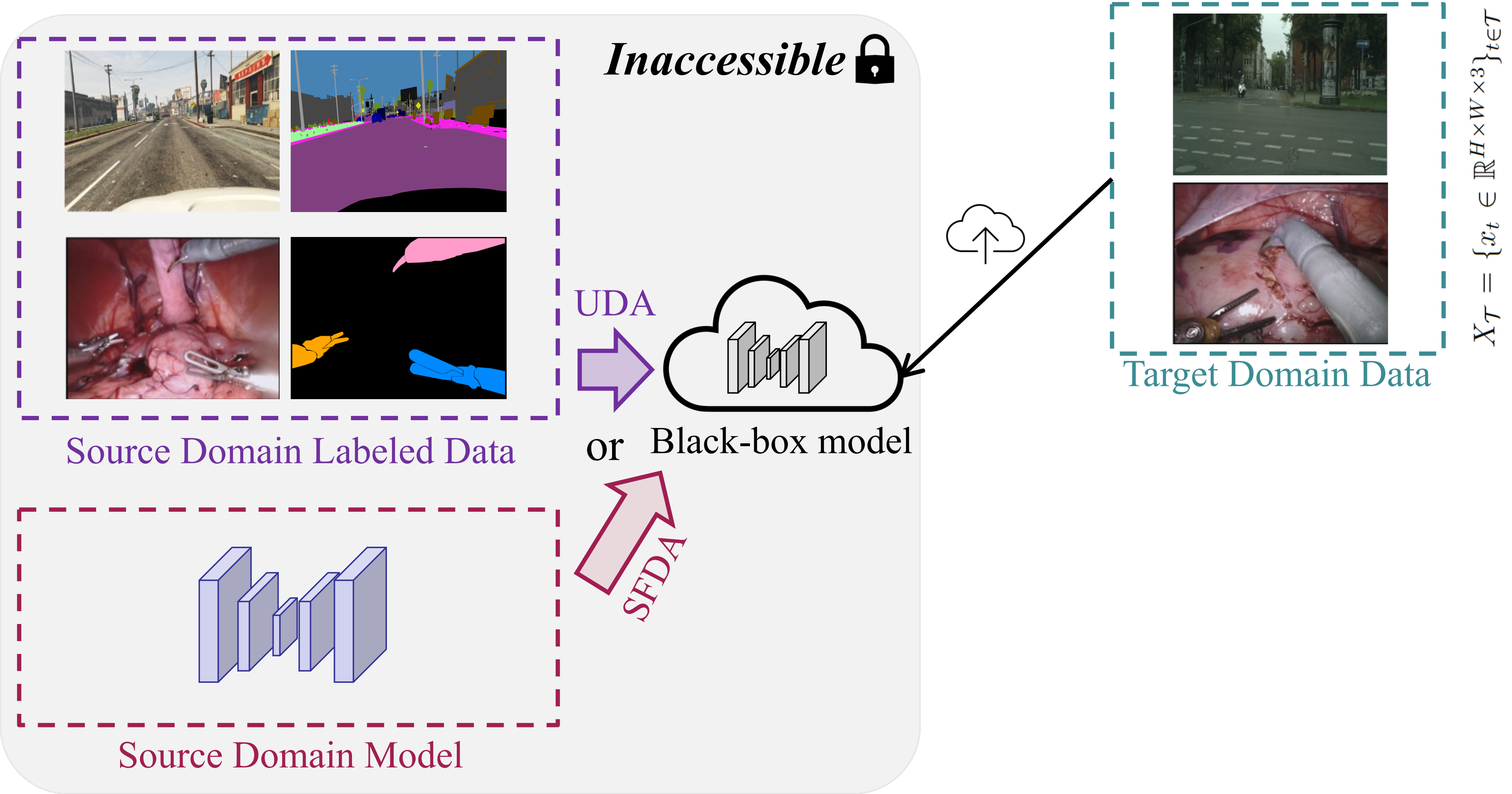
Introduction

- **A General and Practical DA Task**



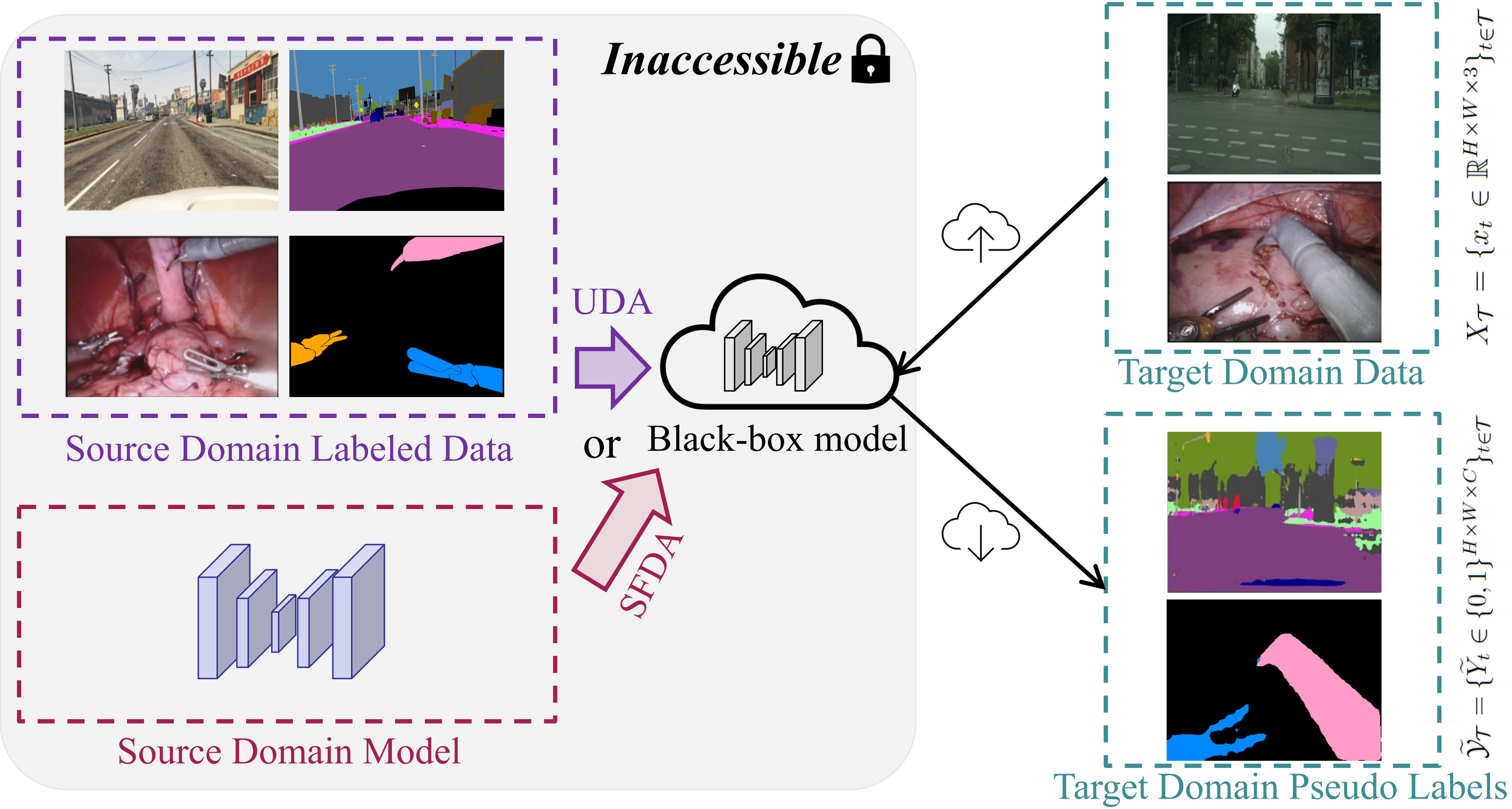
Introduction

- **A General and Practical DA Task**



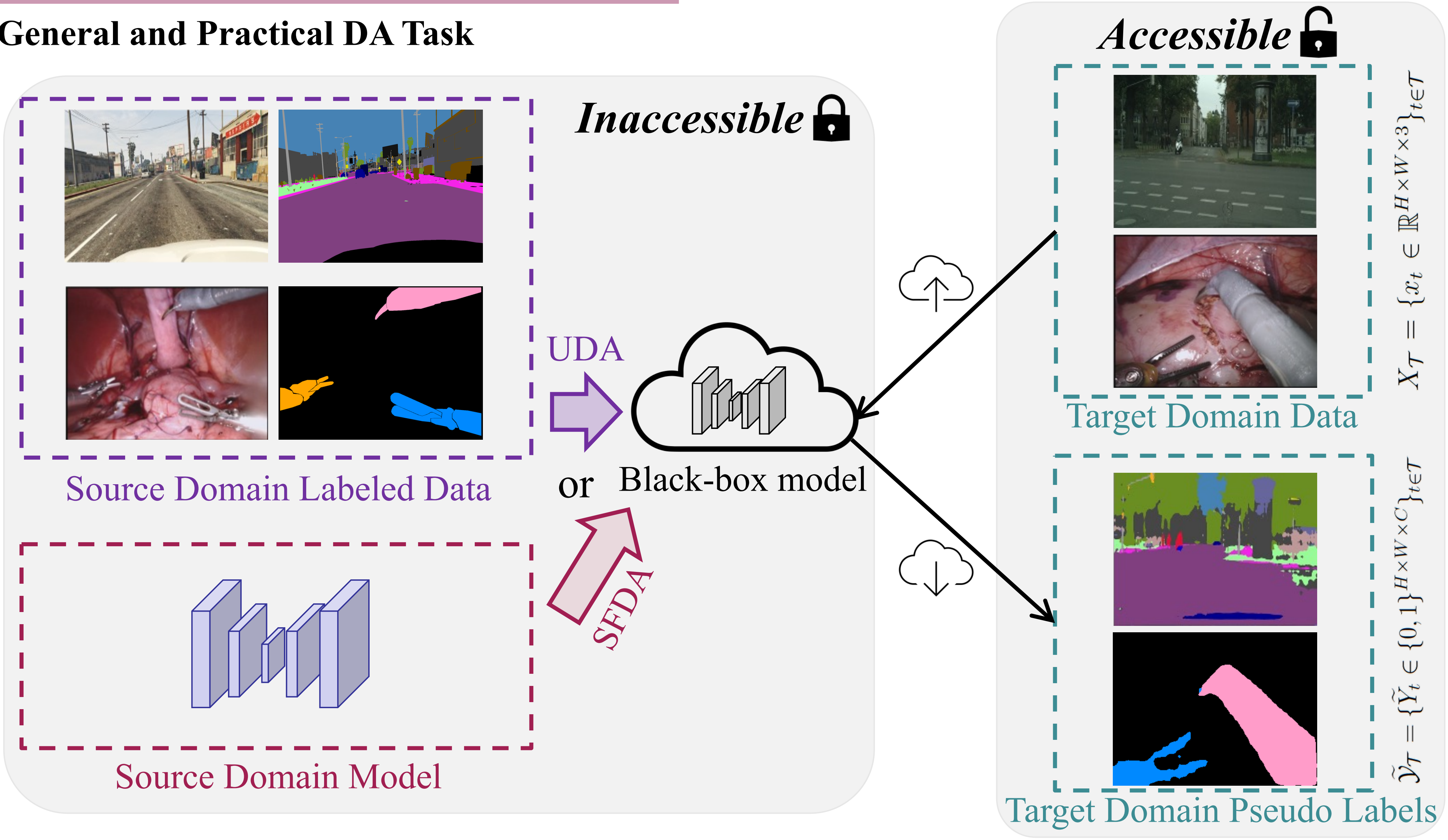
Introduction

- **A General and Practical DA Task**



Introduction

- **A General and Practical DA Task**



Introduction

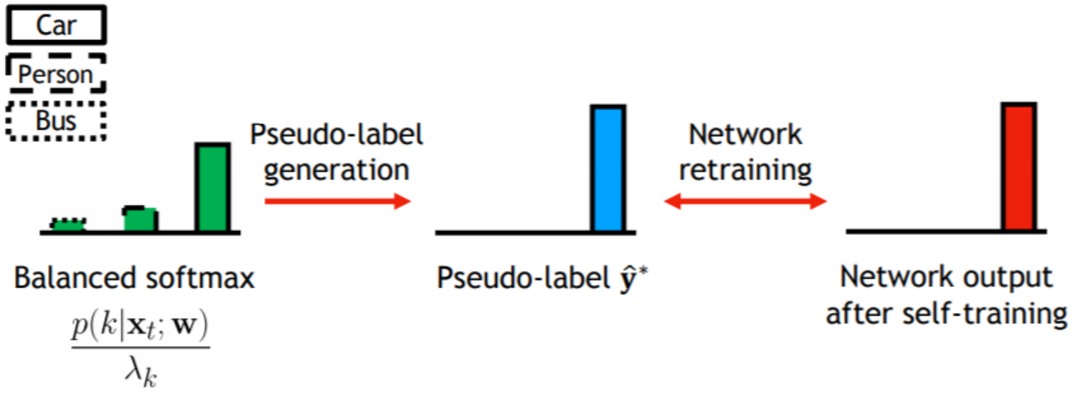
- **Major Challenge of DA Task**
- Label noise issues in pseudo-labeled target data

Introduction

- **Major Challenge of DA Task**
  - Label noise issues in pseudo-labeled target data

- **Major Method**
  - Self-training (ST) strategy:

$$\mathcal{L}_{ST} = -\sum_{t\in\mathcal{T}} \widetilde{Y}_t \log f\left(X_t\right)_{\mathbf{w}}$$



Self-training for DA in semantic segmentation [1]

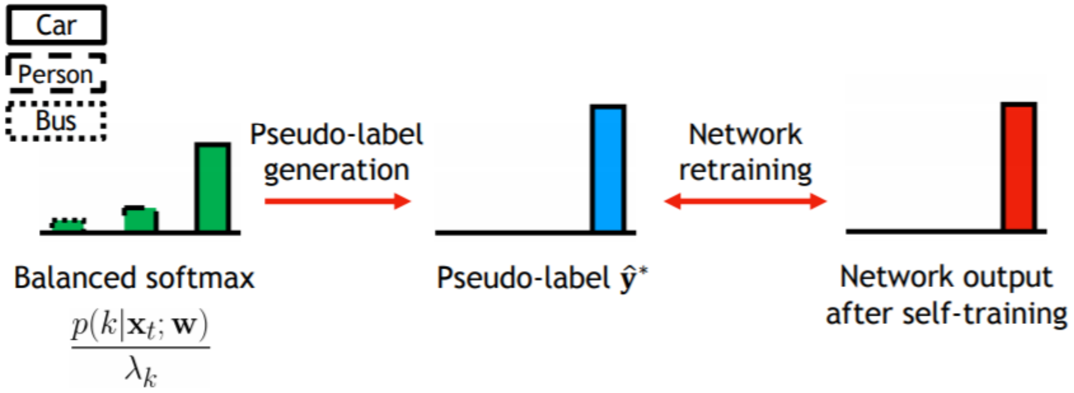
[1] Zou, Yang, et al. "Confidence regularized self-training." ICCV. 2019.

Introduction

- **Major Challenge of DA Task**
  - Label noise issues in pseudo-labeled target data

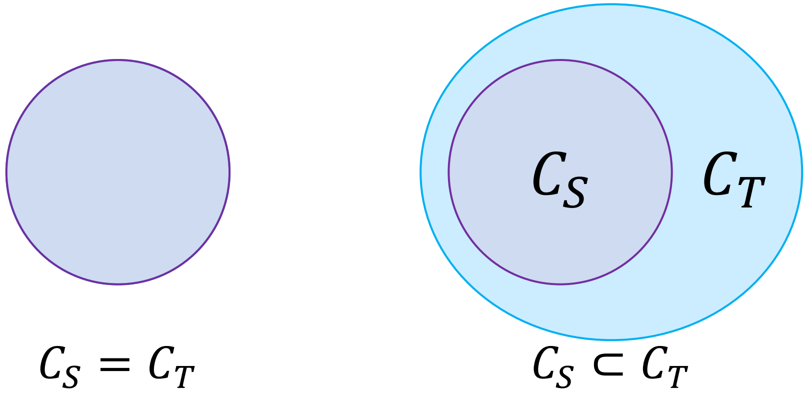
- **Major Method**
  - Self-training (ST) strategy:

$$\mathcal{L}_{ST} = -\sum_{t\in\mathcal{T}} \widetilde{Y}_t \log f\left(X_t\right)_{\mathbf{w}}$$



Self-training for DA in semantic segmentation [1]

- **Major Limitations of ST**
  - Ignore the open-set pseudo label noises in target domain



[1] Zou, Yang, et al. "Confidence regularized self-training." ICCV. 2019.

Introduction

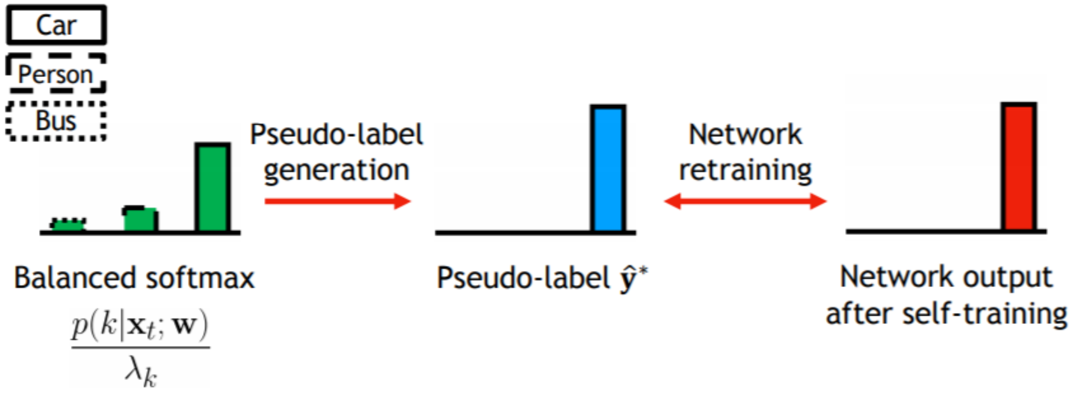
➢ Major Challenge of DA Task

- Label noise issues in pseudo-labeled target data

➢ Major Method

- Self-training (ST) strategy:

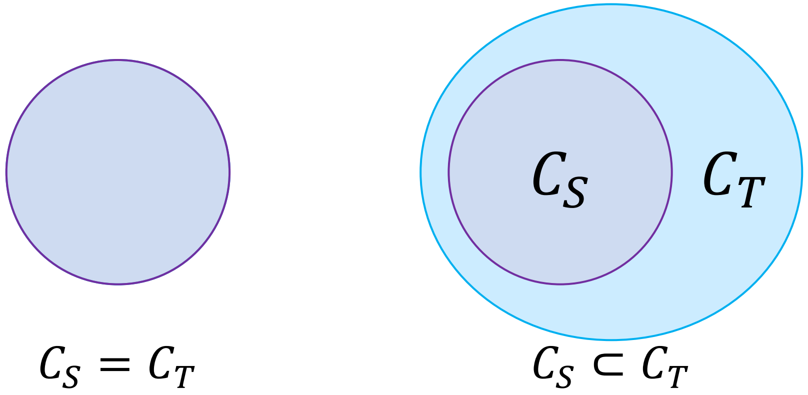
$$\mathcal{L}_{ST} = -\sum_{t\in\mathcal{T}} \widetilde{Y}_t \log f\left(X_t\right)_{\mathbf{w}}$$



Self-training for DA in semantic segmentation [1]

➢ Major Limitations of ST

- Ignore the open-set pseudo label noises in target domain
- Simply dropping confusing pixels leads to biased optimization



[1] Zou, Yang, et al. "Confidence regularized self-training." ICCV. 2019.

Introduction

- **Problem Definition of ST in DA**
  - Carefully address confusing pixels rather than drop them
  - Deal with open-set label noises

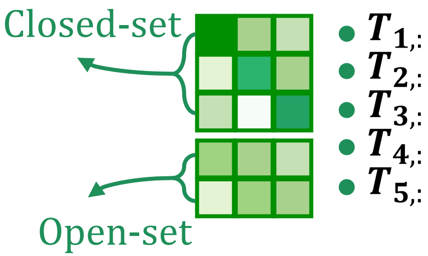
Introduction

➢ **Problem Definition of ST in DA**

- Carefully address confusing pixels rather than drop them
- Deal with open-set label noises

➢ **Our Solution - Simplex noise transition matrix (SimT, *T*)**

- Model the closed- and open-set label noise distributions



$$p(\widetilde{Y}_t = k \mid X_t; \mathbf{w}) = \sum_{j=1}^{C+n} \boldsymbol{T}_{jk} \cdot p(Y_t = j \mid X_t; \mathbf{w}),$$

$$\Rightarrow p(\widetilde{Y}_t \mid X_t; \mathbf{w}) = p(Y_t \mid X_t; \mathbf{w})\, \boldsymbol{T}.$$

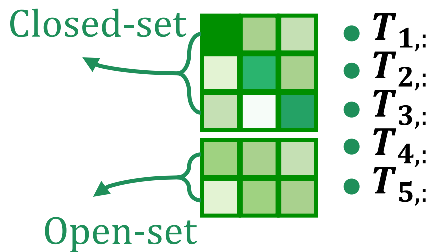
Introduction

➢ **Problem Definition of ST in DA**

- Carefully address confusing pixels rather than drop them
- Deal with open-set label noises

➢ **Our Solution - Simplex noise transition matrix (SimT, *T*)**

- Model the closed- and open-set label noise distributions



$$p(\widetilde{Y}_t = k \mid X_t; \mathbf{w}) = \sum_{j=1}^{C+n} \boldsymbol{T}_{jk} \cdot p(Y_t = j \mid X_t; \mathbf{w}),$$

$$\Rightarrow p(\widetilde{Y}_t \mid X_t; \mathbf{w}) = p(Y_t \mid X_t; \mathbf{w})\, \boldsymbol{T}.$$

- Rectify supervision signals (i.e. loss) derived from noisy pseudo labels, thereby using all target data

$$\mathcal{L}_{ST} = -\sum_{t \in \mathcal{T}} \widetilde{Y}_t \log f\,(X_t)_{\mathbf{w}} \quad \Rightarrow \quad \mathcal{L}_{LC} = -\sum_{t \in \mathcal{T}} \widetilde{Y}_t \log[f\,(X_t)_{\mathbf{w}}\, \boldsymbol{T}].$$

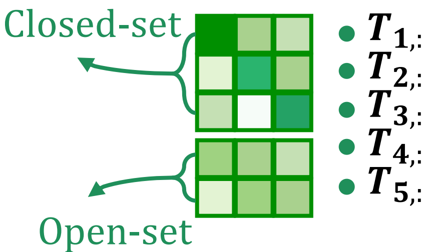
Introduction

➢ Problem Definition of ST in DA

- Carefully address confusing pixels rather than drop them
- Deal with open-set label noises

➢ Our Solution - Simplex noise transition matrix (SimT, *T*)

- Model the closed- and open-set label noise distributions



$$p(\widetilde{Y}_t = k \mid X_t; \mathbf{w}) = \sum_{j=1}^{C+n} \boldsymbol{T}_{jk} \cdot p(Y_t = j \mid X_t; \mathbf{w}),$$

$$\Rightarrow p(\widetilde{Y}_t \mid X_t; \mathbf{w}) = p(Y_t \mid X_t; \mathbf{w})\, \boldsymbol{T}.$$

- Rectify supervision signals (i.e. loss) derived from noisy pseudo labels, thereby using all target data

$$\mathcal{L}_{ST} = -\textstyle\sum_{t \in \mathcal{T}} \widetilde{Y}_t \log f(X_t)_{\mathbf{w}} \quad \Rightarrow \quad \mathcal{L}_{LC} = -\sum_{t \in \mathcal{T}} \widetilde{Y}_t \log[f(X_t)_{\mathbf{w}}\, \boldsymbol{T}].$$

➢ Problem Reformulation

- The problem of alleviating pseudo label noises in DA ⇒ the problem of estimating SimT

Geometry Analysis of SimT

- In the toy example of Fig. 1, SimT is defined as $\boldsymbol{T} \in [0, 1]^{5\times 3}$

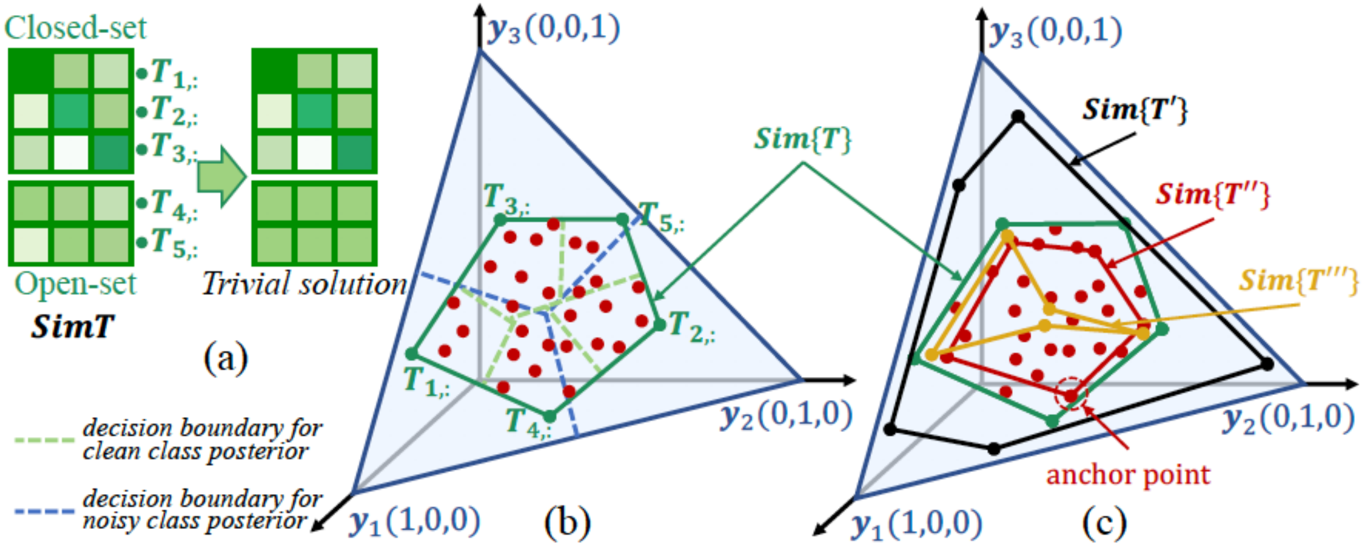


Figure 1. Geometric illustration of SimT with the setting of 3 pseudo label classes and 5 ground truth label classes. Note that coordinates (x, y, z) of any points in blue triangle satisfy conditions of $x + y + z = 1$ and $0 < x, y, z < 1$.

- **Geometry Analysis of SimT**
  - In the toy example of Fig. 1, SimT is defined as $\boldsymbol{T} \in [0, 1]^{5 \times 3}$
  - Rows $[\boldsymbol{T}_{1,:}, \boldsymbol{T}_{2,:}, \boldsymbol{T}_{3,:}]$ model closed-set label noise distribution, and rows $[\boldsymbol{T}_{4,:}, \boldsymbol{T}_{5,:}]$ model open-set one

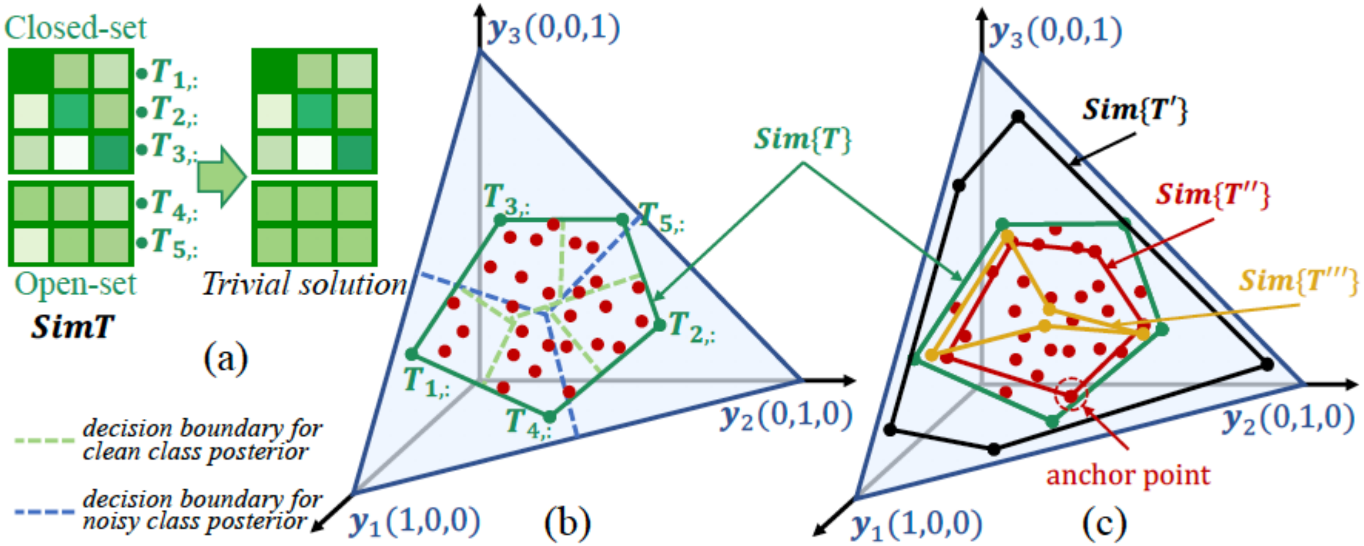


Figure 1. Geometric illustration of SimT with the setting of 3 pseudo label classes and 5 ground truth label classes. Note that coordinates (x, y, z) of any points in blue triangle satisfy conditions of $x + y + z = 1$ and $0 < x, y, z < 1$.

Method

Geometry Analysis of SimT

- In the toy example of Fig. 1, SimT is defined as $\boldsymbol{T} \in [0, 1]^{5\times 3}$
- Rows $[\boldsymbol{T}_{1,:}, \boldsymbol{T}_{2,:}, \boldsymbol{T}_{3,:}]$ model closed-set label noise distribution, and rows $[\boldsymbol{T}_{4,:}, \boldsymbol{T}_{5,:}]$ model open-set one
- Each row $\boldsymbol{T}_{c,:}$ is a 3D point within the blue triangle

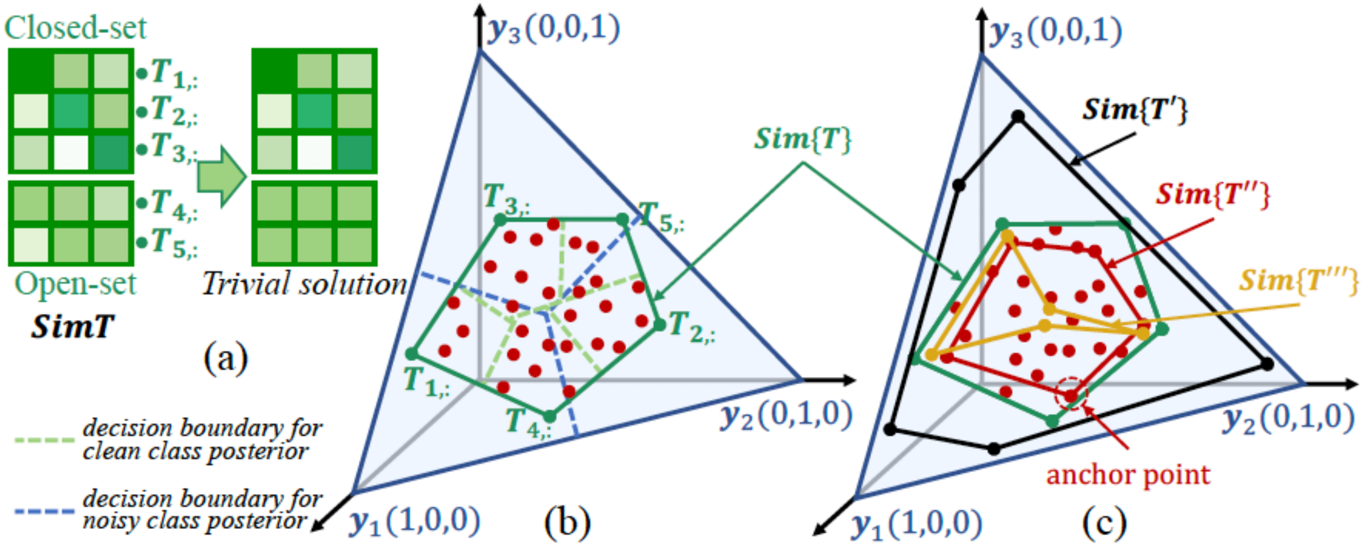


Figure 1. Geometric illustration of SimT with the setting of 3 pseudo label classes and 5 ground truth label classes. Note that coordinates (x, y, z) of any points in blue triangle satisfy conditions of $x + y + z = 1$ and $0 < x, y, z < 1$.

Department of Electrical Engineering 香港城市大學 City University of Hong Kong
THE UNIVERSITY OF SYDNEY

- **Geometry Analysis of SimT**
  - Given an input $\boldsymbol{x}$, the noisy class posterior $p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x})$ is a red point scattered in the blue triangle

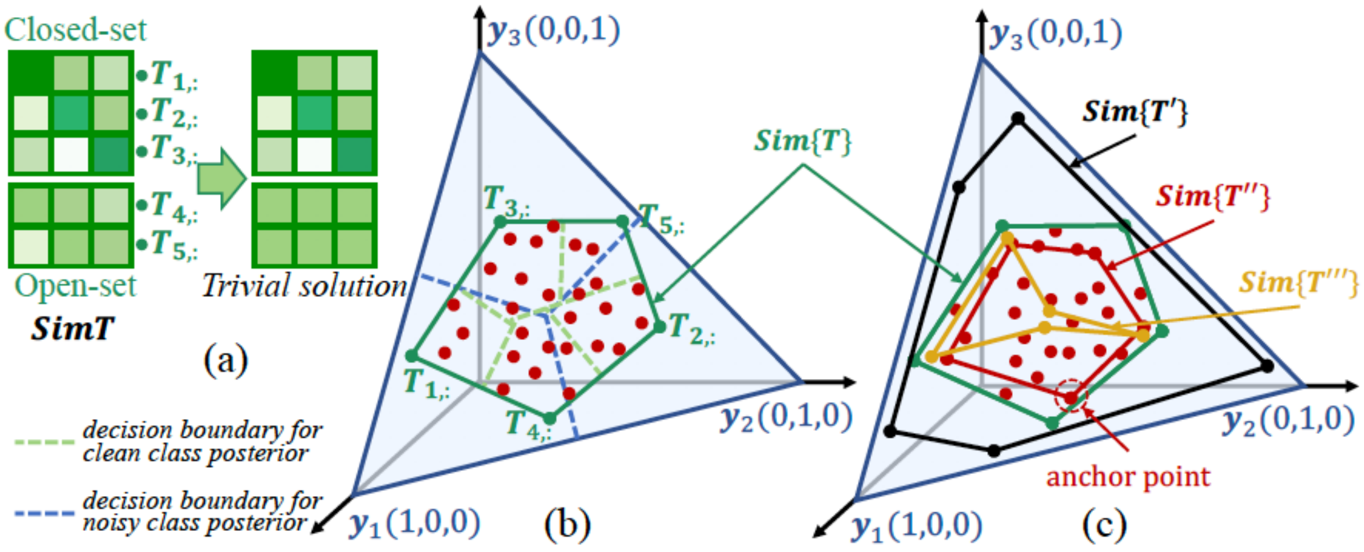


Figure 1. Geometric illustration of SimT with the setting of 3 pseudo label classes and 5 ground truth label classes. Note that coordinates (x, y, z) of any points in blue triangle satisfy conditions of $x + y + z = 1$ and $0 < x, y, z < 1$.

Noisy class posterior:

$$p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x}) = [p(\tilde{y} = 1 \mid \boldsymbol{x}), p(\tilde{y} = 2 \mid \boldsymbol{x}), p(\tilde{y} = 3 \mid \boldsymbol{x})]$$

Clean class posterior:

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = [p(y = 1 \mid \boldsymbol{x}), p(y = 2 \mid \boldsymbol{x}), p(y = 3 \mid \boldsymbol{x}), p(y = 4 \mid \boldsymbol{x}), p(y = 5 \mid \boldsymbol{x})]$$

Method

➢ Geometry Analysis of SimT

- Given an input $\boldsymbol{x}$, the noisy class posterior $p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x})$ is a red point scattered in the blue triangle

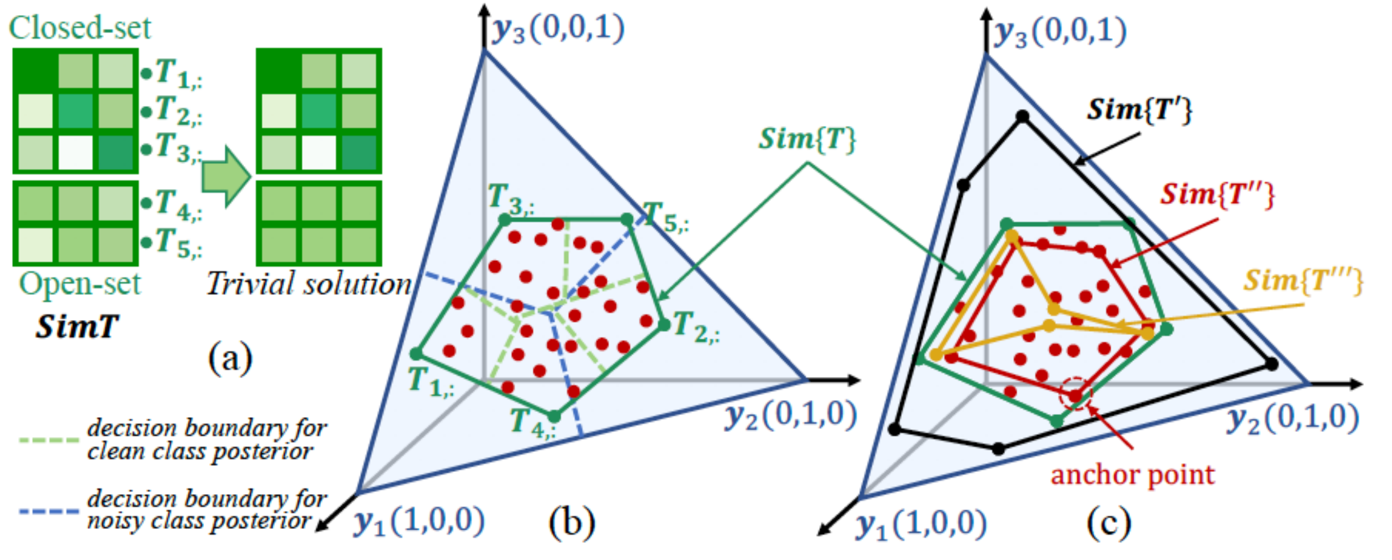


Figure 1. Geometric illustration of SimT with the setting of 3 pseudo label classes and 5 ground truth label classes. Note that coordinates (x, y, z) of any points in blue triangle satisfy conditions of $x + y + z = 1$ and $0 < x, y, z < 1$.

Noisy class posterior:

$$p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x}) = [p(\tilde{y} = 1 \mid \boldsymbol{x}), p(\tilde{y} = 2 \mid \boldsymbol{x}), p(\tilde{y} = 3 \mid \boldsymbol{x})]$$

Clean class posterior:

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = [p(y = 1 \mid \boldsymbol{x}), p(y = 2 \mid \boldsymbol{x}), p(y = 3 \mid \boldsymbol{x}), p(y = 4 \mid \boldsymbol{x}), p(y = 5 \mid \boldsymbol{x})]$$

Bridge of SimT:

$$p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x}) = p(\boldsymbol{y} \mid \boldsymbol{x})\, \boldsymbol{T}$$

- **Geometry Analysis of SimT**
  - Given an input $\boldsymbol{x}$, the noisy class posterior $p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x})$ is a red point scattered in the blue triangle
  - $p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x})$ is the convex combination among rows of $\boldsymbol{T}$ with the factor of clean class posterior $p(\boldsymbol{y} \mid \boldsymbol{x})$.

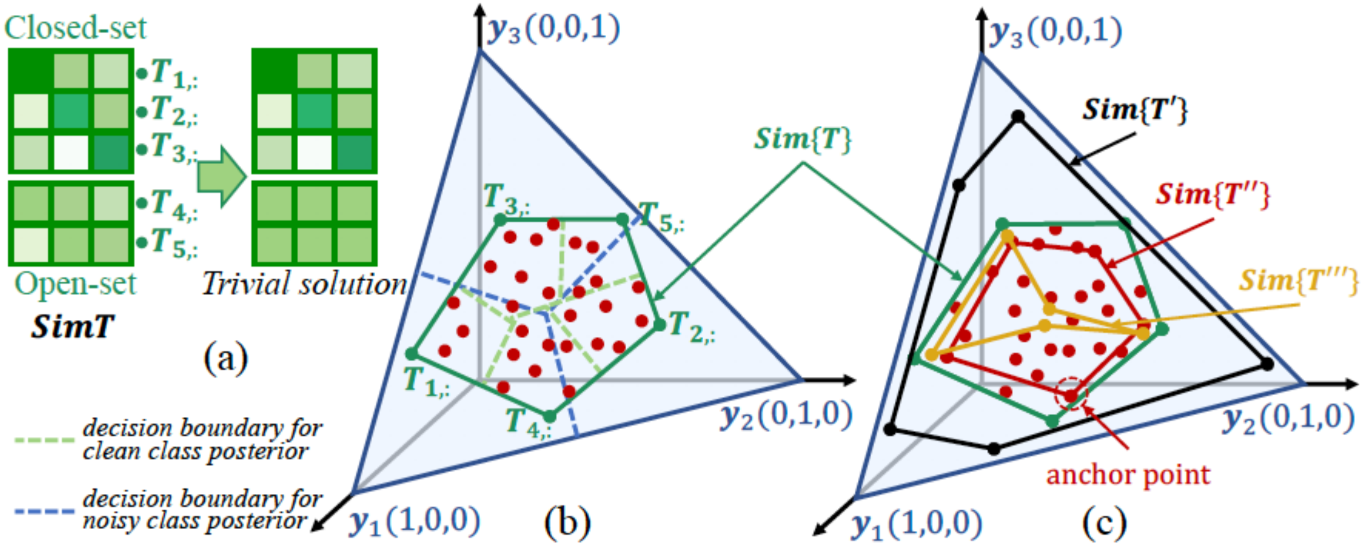


Figure 1. Geometric illustration of SimT with the setting of 3 pseudo label classes and 5 ground truth label classes. Note that coordinates (x, y, z) of any points in blue triangle satisfy conditions of $x + y + z = 1$ and $0 < x, y, z < 1$.

Noisy class posterior:

$$p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x}) = [p(\tilde{y} = 1 \mid \boldsymbol{x}), p(\tilde{y} = 2 \mid \boldsymbol{x}), p(\tilde{y} = 3 \mid \boldsymbol{x})]$$

Clean class posterior:

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = [p(y = 1 \mid \boldsymbol{x}), p(y = 2 \mid \boldsymbol{x}), p(y = 3 \mid \boldsymbol{x}), p(y = 4 \mid \boldsymbol{x}), p(y = 5 \mid \boldsymbol{x})]$$

Bridge of SimT:

$$p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x}) = p(\boldsymbol{y} \mid \boldsymbol{x})\, \boldsymbol{T}$$

Method

Geometry Analysis of SimT

- In blue triangle, the closer $p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x})$ is to $\boldsymbol{T}_{c,:}$ of Sim{$\boldsymbol{T}$}, the more likely $\boldsymbol{x}$ belongs to class c.

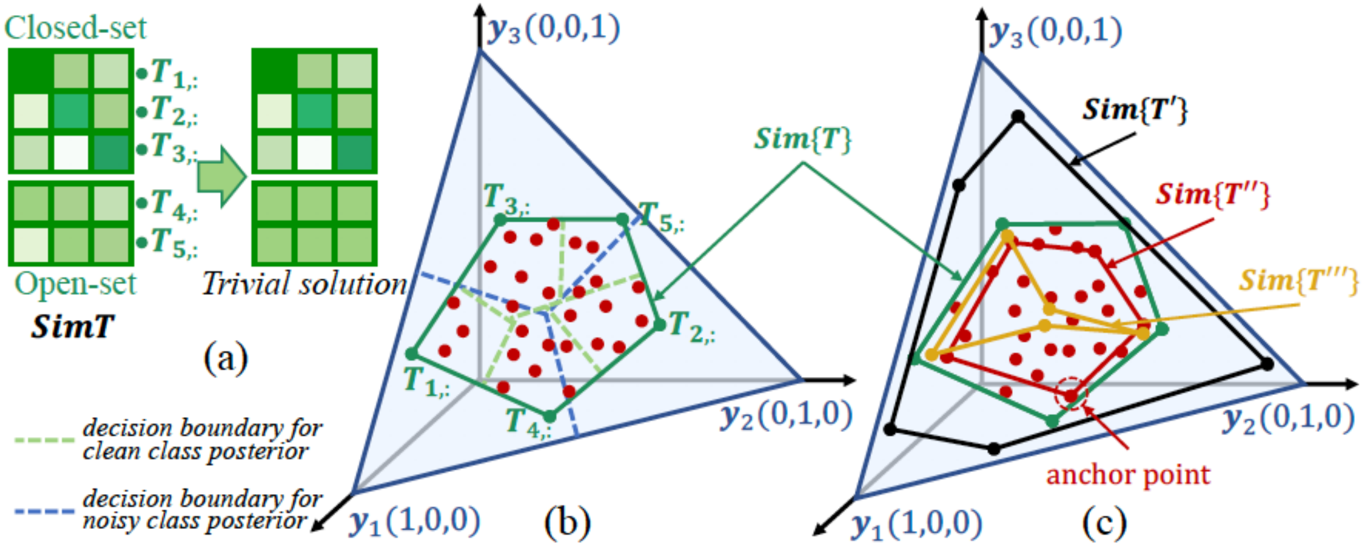


Figure 1. Geometric illustration of SimT with the setting of 3 pseudo label classes and 5 ground truth label classes. Note that coordinates (x, y, z) of any points in blue triangle satisfy conditions of $x + y + z = 1$ and $0 < x, y, z < 1$.

Noisy class posterior:

$$p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x}) = [p(\tilde{y} = 1 \mid \boldsymbol{x}), \; p(\tilde{y} = 2 \mid \boldsymbol{x}), \; p(\tilde{y} = 3 \mid \boldsymbol{x})]$$

Clean class posterior:

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = [p(y = 1 \mid \boldsymbol{x}), \; p(y = 2 \mid \boldsymbol{x}), \; p(y = 3 \mid \boldsymbol{x}), \; p(y = 4 \mid \boldsymbol{x}), \; p(y = 5 \mid \boldsymbol{x})]$$

Bridge of SimT:

$$p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x}) = p(\boldsymbol{y} \mid \boldsymbol{x})\, \boldsymbol{T}$$

Method

Geometry Analysis of SimT

- In blue triangle, the closer $p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x})$ is to $\boldsymbol{T}_{c,:}$ of Sim{$\boldsymbol{T}$}, the more likely $\boldsymbol{x}$ belongs to class c.
- Hence, the simplex formed by rows of $\boldsymbol{T}$ should enclose $p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x})$ for any $\boldsymbol{x}$.

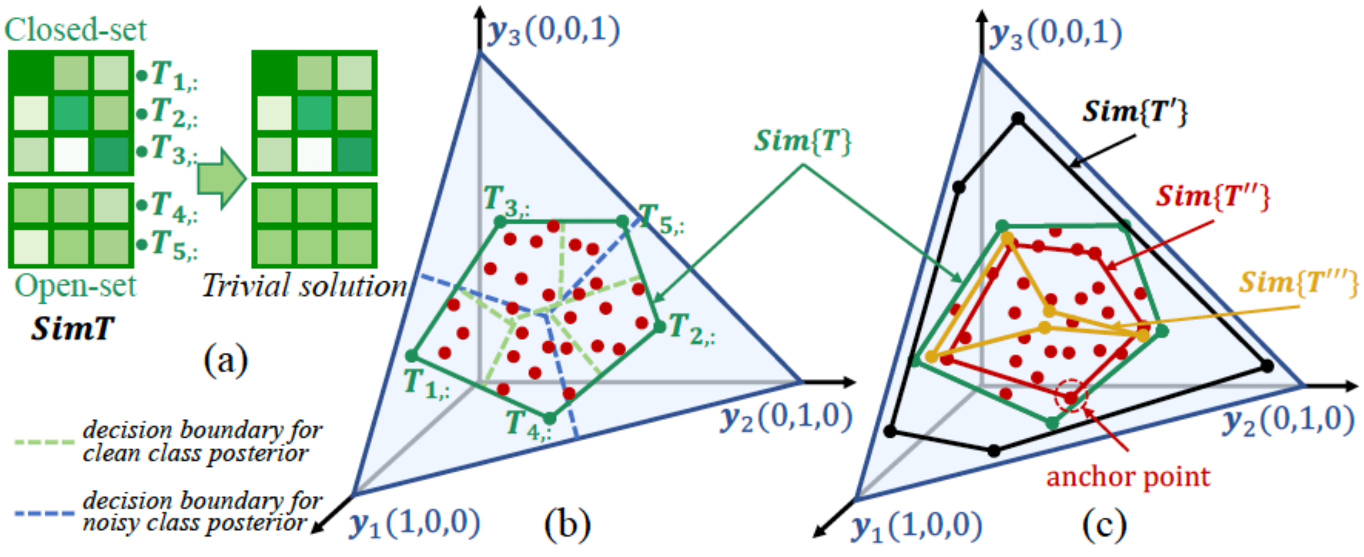


Figure 1. Geometric illustration of SimT with the setting of 3 pseudo label classes and 5 ground truth label classes. Note that coordinates (x, y, z) of any points in blue triangle satisfy conditions of $x + y + z = 1$ and $0 < x, y, z < 1$.

Noisy class posterior:

$$p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x}) = [p(\tilde{y} = 1 \mid \boldsymbol{x}), p(\tilde{y} = 2 \mid \boldsymbol{x}), p(\tilde{y} = 3 \mid \boldsymbol{x})]$$

Clean class posterior:

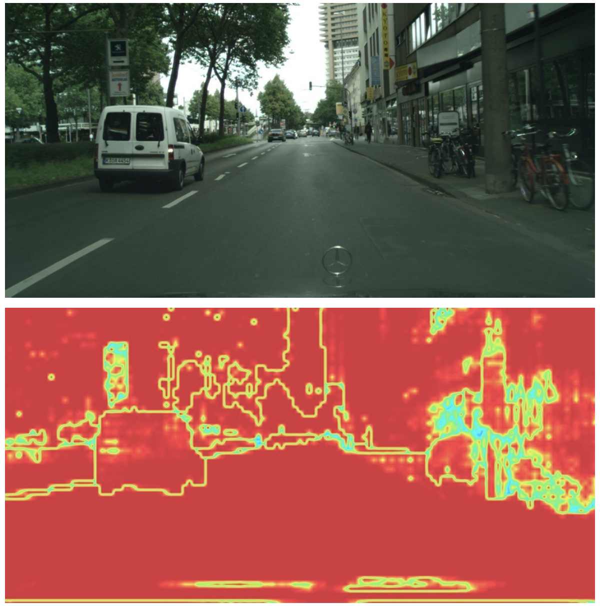
$$p(\boldsymbol{y} \mid \boldsymbol{x}) = [p(y = 1 \mid \boldsymbol{x}), p(y = 2 \mid \boldsymbol{x}), p(y = 3 \mid \boldsymbol{x}), p(y = 4 \mid \boldsymbol{x}), p(y = 5 \mid \boldsymbol{x})]$$

Bridge of SimT:

$$p(\tilde{\boldsymbol{y}} \mid \boldsymbol{x}) = p(\boldsymbol{y} \mid \boldsymbol{x})\, \boldsymbol{T}$$

Method

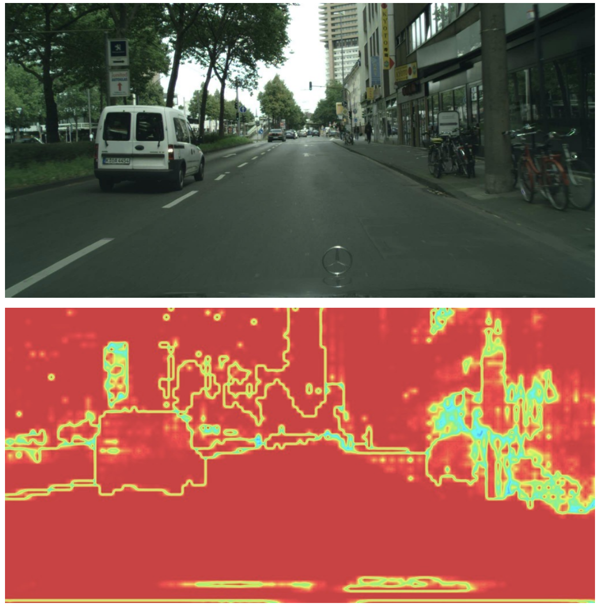
- **Volume Minimization (VM) for Estimating SimT**
  - VM is based on the sufficiently scattered assumption [2], which is easily satisfied in segmentation



[2] Xuefeng Li,Tongliang Liu,Bo Han,Gang Niu,and Masashi Sugiyama. Provably end-to-end label-noise learning without anchor points. In ICML, pages 6403–6413, 2021.

Method

- **Volume Minimization (VM) for Estimating SimT**
  - VM is based on the sufficiently scattered assumption [2], which is easily satisfied in segmentation
  - Directly applying VM that constrains a square $\boldsymbol{T}$ to SimT is problematic:

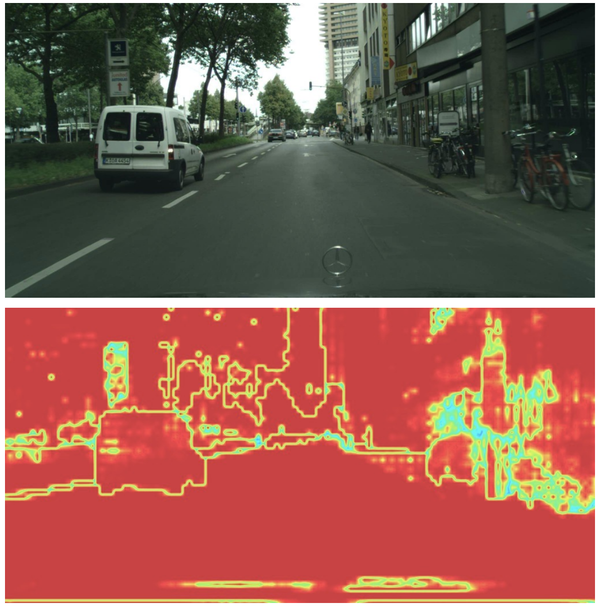


[2] Xuefeng Li,Tongliang Liu,Bo Han,Gang Niu,and Masashi Sugiyama. Provably end-to-end label-noise learning without anchor points. In ICML, pages 6403–6413, 2021.

Method

- **Volume Minimization (VM) for Estimating SimT**
  - VM is based on the sufficiently scattered assumption [2], which is easily satisfied in segmentation
  - Directly applying VM that constrains a square $\boldsymbol{T}$ to SimT is problematic:

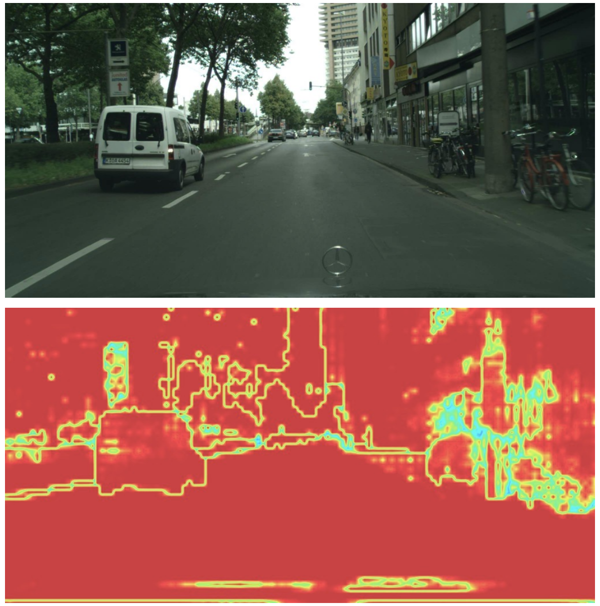
    (1) Volume of non-square SimT cannot be computed by its determination

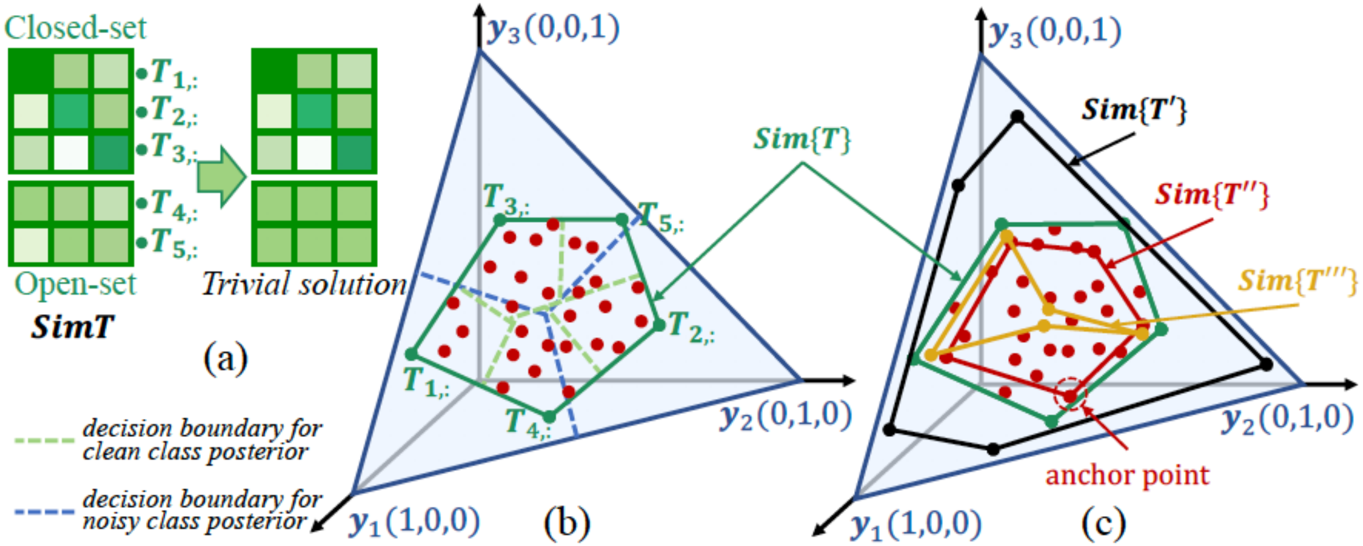


[2] Xuefeng Li,Tongliang Liu,Bo Han,Gang Niu,and Masashi Sugiyama. Provably end-to-end label-noise learning without anchor points. In ICML, pages 6403–6413, 2021.

Method

- **Volume Minimization (VM) for Estimating SimT**
  - VM is based on the sufficiently scattered assumption [2], which is easily satisfied in segmentation
  - Directly applying VM that constrains a square ***T*** to SimT is problematic:
    (1) Volume of non-square SimT cannot be computed by its determination
    (2) By minimizing volume of SimT, open-set part of SimT will collapse to a trivial solution



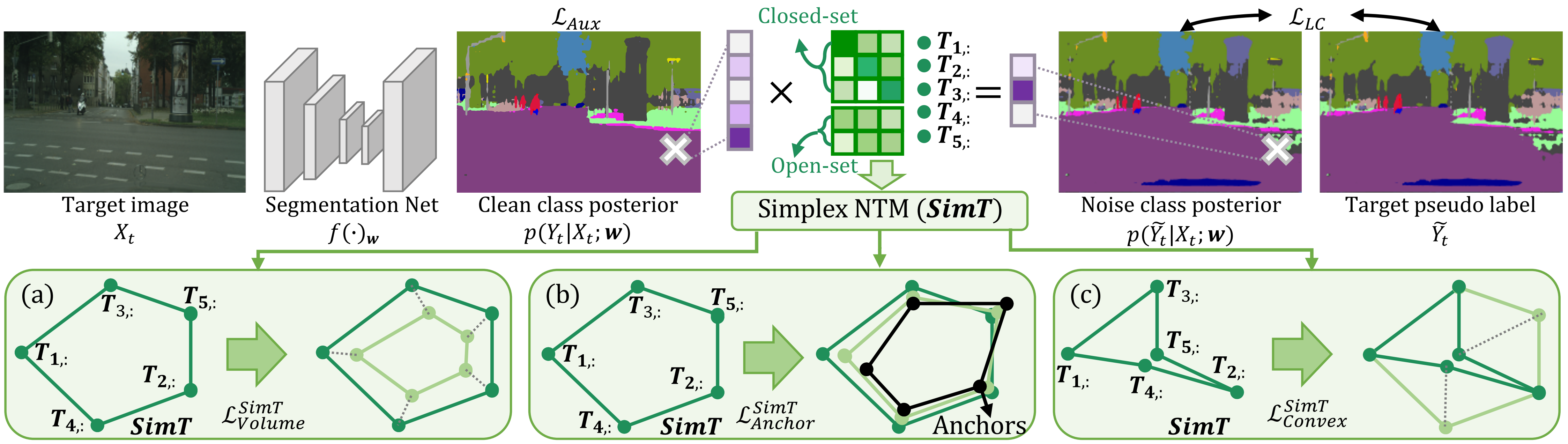


[2] Xuefeng Li,Tongliang Liu,Bo Han,Gang Niu,and Masashi Sugiyama. Provably end-to-end label-noise learning without anchor points. In ICML, pages 6403–6413, 2021.

Method

- **Three proposed regularizations for estimating SimT:**
  - Volume regularization:

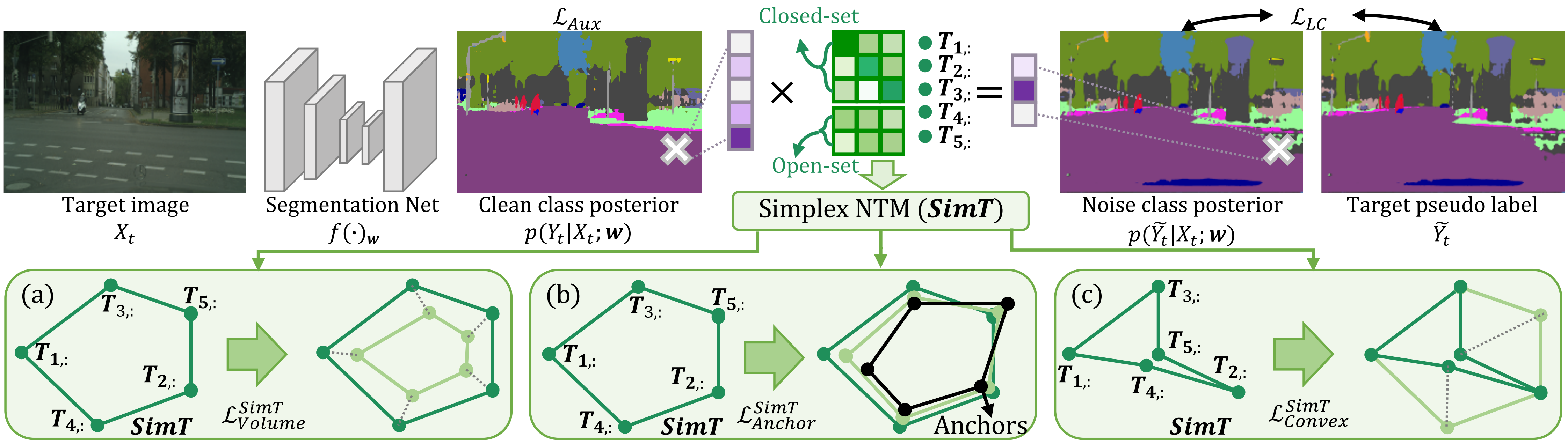
$$\mathcal{L}_{Volume}^{SimT} = \log[\mathrm{Vol}(\boldsymbol{T})] = \log\sqrt{\det(\boldsymbol{T}^{\top}\boldsymbol{T})/(N!)}$$



Method

- **Three proposed regularizations for estimating SimT:**
  - Anchor guidance:

anchor point of class c: $\boldsymbol{x}^c = \operatorname*{argmax}_{\boldsymbol{x}} p(y = c \mid \boldsymbol{x}; \mathbf{w})$.

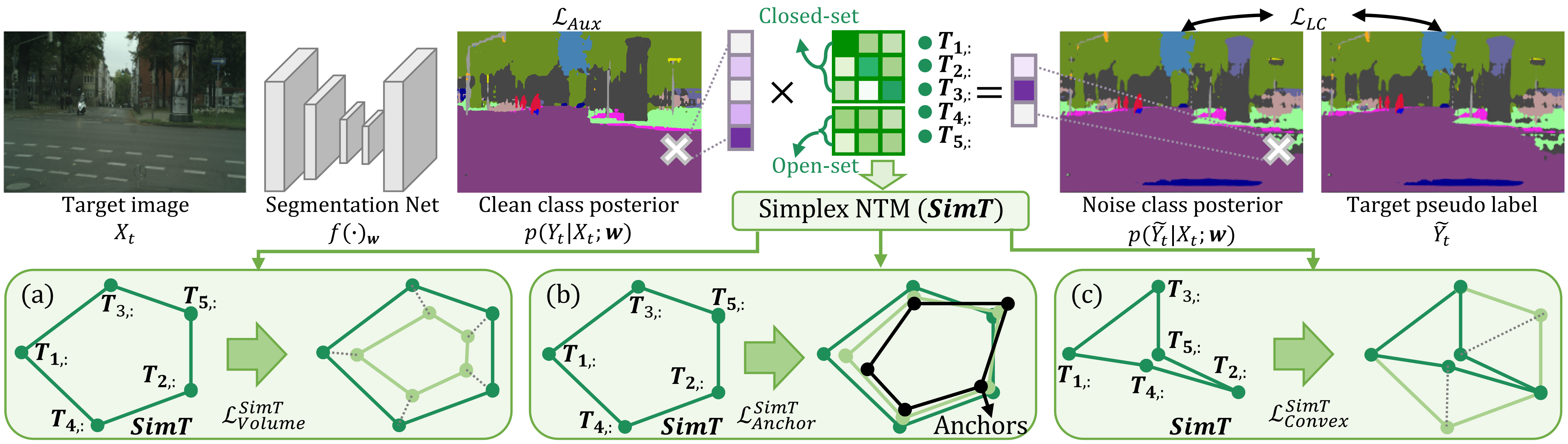


Method

- **Three proposed regularizations for estimating SimT:**
  - Anchor guidance:

anchor point of class c: $\boldsymbol{x}^c = \underset{\boldsymbol{x}}{\operatorname{argmax}}\, p(y = c \mid \boldsymbol{x}; \mathbf{w})$.

$$\mathcal{L}^{SimT}_{Anchor} = \sum_{c=1}^{C+n} \sum_{k=1}^{C} \left\| \boldsymbol{T}_{ck} - p(\widetilde{y} = k \mid \boldsymbol{x}^c; \mathbf{w}^{fixed}) \right\|^2$$



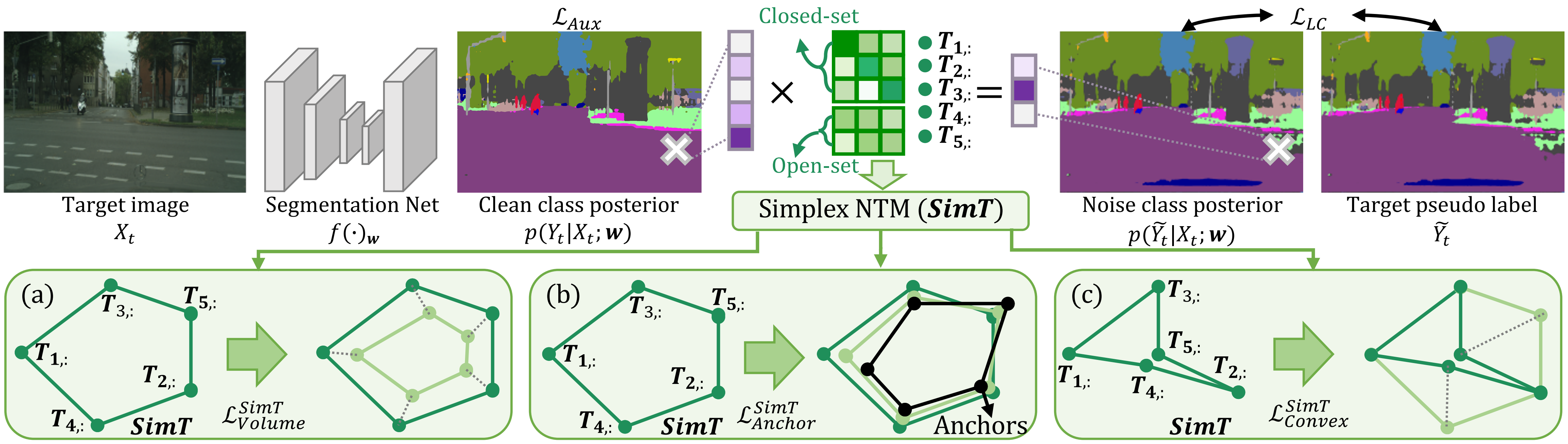
Method

- **Three proposed regularizations for estimating SimT:**
  - Anchor guidance:

anchor point of class c: $\boldsymbol{x}^c = \underset{\boldsymbol{x}}{\operatorname{argmax}}\, p(y = c \mid \boldsymbol{x}; \mathbf{w})$.

$$\mathcal{L}_{Aux} = -\sum_{i \in \mathcal{K} \& \mathcal{U}} \tilde{\boldsymbol{y}}_i \log f(\boldsymbol{x}_i)_{\mathbf{w}} - \lambda \sum_{k \in \mathcal{K}} \tilde{\boldsymbol{y}}^o \log[f(\boldsymbol{x}_k)_{\mathbf{w}} \setminus \tilde{y}_k],$$

$$\mathcal{L}_{Anchor}^{SimT} = \sum_{c=1}^{C+n} \sum_{k=1}^{C} \left\| \boldsymbol{T}_{ck} - p(\tilde{y} = k \mid \boldsymbol{x}^c; \mathbf{w}^{fixed}) \right\|^2$$

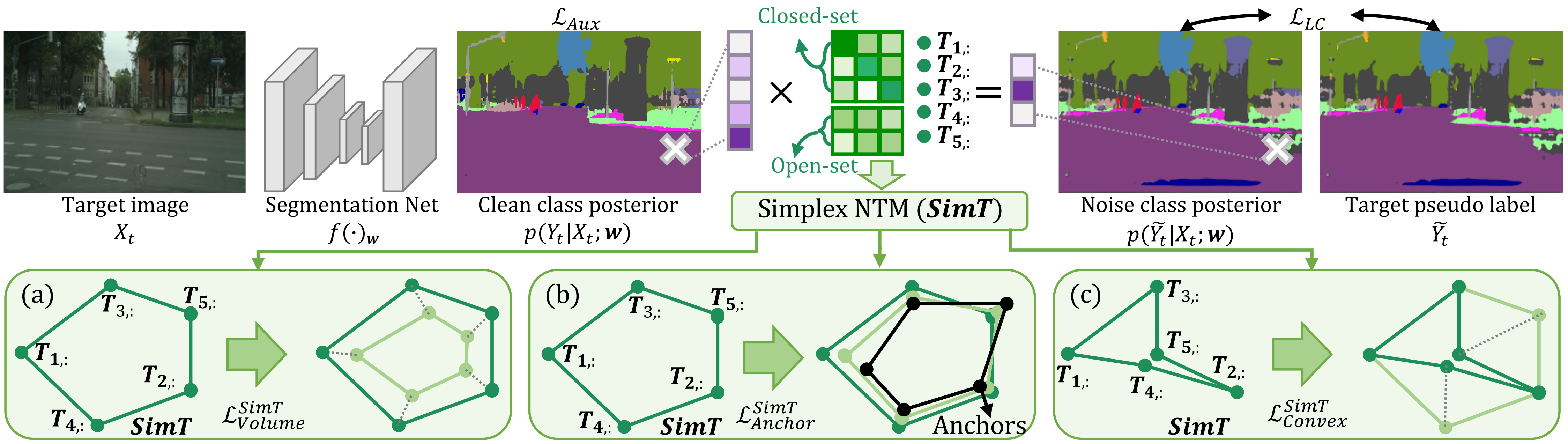


Method

- **Three proposed regularizations for estimating SimT:**
  - Convex guarantee:

$$\boldsymbol{u} \in [0,1]^{(C+n)\times(C+n)} \quad \boldsymbol{u}_{j,k=j} = -1 \quad \sum_k \boldsymbol{u}_{j,k\neq j} = 1$$

$$\mathcal{L}_{Convex}^{SimT} = -\left\|\boldsymbol{u}\boldsymbol{T}\right\|^2, \text{ where } \boldsymbol{u} = \operatorname*{argmin}_{\boldsymbol{u}} \left\|\boldsymbol{u}\boldsymbol{T}\right\|^2$$



Experiments

- **Scenario of GTA5→Cityscapes**
  - **GTA5**: 24,966 images captured from a virtual video game, including 19 shared classes
  - **CityScapes**: a real-world dataset collected in driving scenarios, including 15 open-set classes

    Training set: 2,975 unlabeled images Test set: 500 images

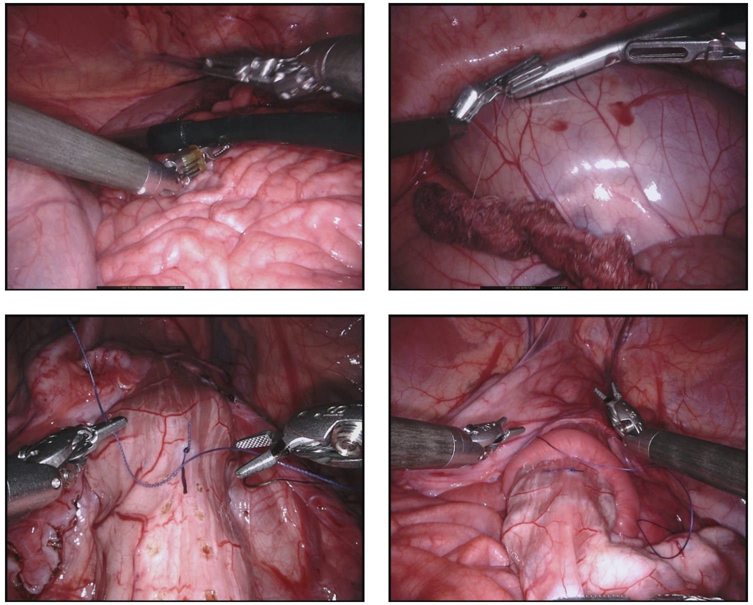
GTA5

CityScapes

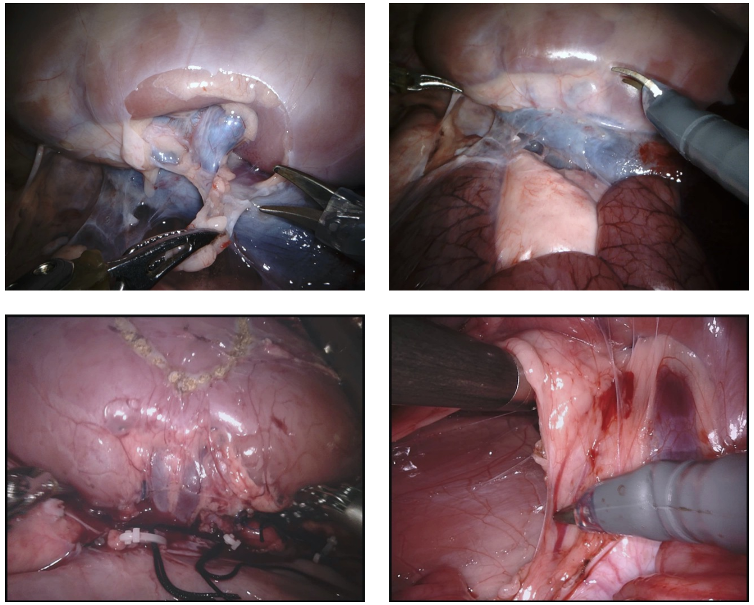
Experiments

- **Scenario of Endovis17→Endovis18**
  - **Endovis17**: 1800 images include 3 shared classes `scissor', `needle driver', `forceps'
  - **Endovis18**: 1800 images include 3 open-set classes `ultrasound probe', `suction instrument', `clip applier'

    Training set: 1639 unlabeled images   Test set: 596 validation images



Endovis17



Endovis18

Experiments

Department of Electrical Engineering
香港城市大學
City University of Hong Kong
THE UNIVERSITY OF SYDNEY

➢ Quantitative Results

- Results of UDA and SFDA on adapting GTA5 to Cityscapes
- Results of UDA and SFDA on adapting Endovis17 to Endovis18

Table 1. Results of UDA and SFDA on adapting GTA5 to Cityscapes.

Methods	road	sidewalk	building	wall	fence	pole	t-light	t-sign	veg.	terrain	sky	person	rider	car	truck	bus	train	motor	bike	mIoU
UDA																				
AdaptSegNet (18') [34]	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
LTIR (20') [14]	92.9	55.0	85.3	34.2	31.1	34.9	40.7	34.0	85.2	40.1	87.1	61.0	31.1	82.5	32.3	42.9	0.3	36.4	46.1	50.2
MLSL (20') [13]	89.0	45.2	78.2	22.9	27.3	37.4	46.1	43.8	82.9	18.6	61.2	60.4	26.7	85.4	35.9	44.9	**36.4**	37.2	49.3	49.0
IntraDA (20') [25]	90.6	37.1	82.6	30.1	19.1	29.5	32.4	20.6	85.7	40.5	79.7	58.7	31.1	86.3	31.5	48.3	0.0	30.2	35.8	46.3
RPLL (21') [46]	90.4	31.2	85.1	36.9	25.6	37.5	48.8	48.5	85.3	34.8	81.1	64.4	36.8	86.3	34.9	52.2	1.7	29.0	44.6	50.3
MetaCorrection (21') [11]	92.8	58.1	86.2	39.7	33.1	36.3	42.0	38.6	85.5	37.8	87.6	62.8	31.7	84.8	35.7	50.3	2.0	36.8	48.0	52.1
DPL-Dual (21') [5]	92.8	54.4	86.2	41.6	32.7	36.4	49.0	34.0	85.8	41.3	86.0	63.2	34.2	87.2	39.3	44.5	18.7	42.6	43.1	53.3
UncerDA (21') [37]	90.5	38.7	86.5	41.1	32.9	40.5	48.2	42.1	86.5	36.8	84.2	64.5	38.1	87.2	34.8	50.4	0.2	41.8	54.6	52.6
DSP (21') [9]	92.4	48.0	87.4	33.4	35.1	36.4	41.6	46.0	87.7	43.2	**89.8**	66.6	32.1	89.9	57.0	56.1	0.0	44.1	57.8	55.0
ProDA (21') [42]	87.8	56.0	79.7	**46.3**	**44.8**	**45.6**	**53.5**	53.5	**88.6**	45.2	82.1	70.7	39.2	88.8	45.5	59.4	1.0	48.9	56.4	57.5
BAPA-Net (21') [22]	**94.4**	**61.0**	88.0	26.8	39.9	38.3	46.1	55.3	87.8	46.1	89.4	68.8	40.0	90.2	**60.4**	59.0	0.0	45.1	54.2	57.4
Ours (SimT)	94.2	60.0	**88.5**	30.3	39.7	41.2	47.8	**60.8**	**88.6**	**47.3**	89.3	**71.5**	**45.0**	**90.7**	54.2	**60.2**	0.0	**51.8**	**58.4**	**58.9**
SFDA																				
Source only	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
Test-time BN (20') [24]	79.5	79.5	81.3	72.7	29.8	74.1	28.0	58.8	25.3	22.2	19.4	11.0	22.6	17.0	24.5	19.9	34.3	2.4	14.7	37.7
SHOT (20') [18]	88.7	34.9	82.1	27.6	22.5	30.9	31.4	24.7	83.6	37.7	76.8	58.1	24.9	83.9	34.1	39.5	6.0	26.8	24.6	44.1
TENT (21') [35]	87.3	79.8	83.8	**85.0**	39.0	77.7	21.2	57.9	34.7	19.6	24.3	4.5	16.6	20.8	24.9	17.8	25.1	2.0	16.6	38.9
SFDA (21') [23]	84.2	82.7	82.4	80.0	39.2	**85.3**	25.9	58.7	30.5	22.1	27.5	30.6	21.9	31.5	33.1	22.1	31.1	3.6	27.8	43.2
S4T (21') [28]	89.7	**84.4**	**86.8**	**85.0**	**46.7**	79.3	39.5	**61.2**	41.8	29.0	25.7	9.3	**36.8**	28.2	19.3	26.7	**45.1**	5.3	11.8	44.8
URMA (21') [30]	92.3	55.2	81.6	30.8	18.8	37.1	17.7	12.1	84.2	35.9	83.8	57.7	24.1	81.7	27.5	44.3	6.9	24.1	40.4	45.1
SFDASeg (21') [15]	91.7	53.4	86.1	37.6	32.1	37.4	38.2	35.6	86.7	**48.5**	**89.9**	62.6	34.3	87.2	**51.0**	50.8	4.2	42.7	53.9	53.4
Ours (SimT)	**92.3**	55.8	86.3	34.4	31.7	37.8	**39.9**	41.4	**87.1**	47.8	88.5	**64.7**	36.3	**87.3**	41.7	**55.2**	0.0	**47.4**	**57.6**	**54.4**

Table 2. Results on adapting Endovis17 to Endovis18.

Methods	scissor	needle driver	forceps	mIoU
UDA				
I2I (19') [27]	60.78	15.34	52.55	42.89
FDA (20') [40]	68.25	18.29	50.29	45.61
S2RC (20') [32]	64.48	22.79	49.34	45.54
IntraDA (20') [25]	68.07	25.88	52.63	48.86
ASANet (20') [47]	70.19	11.34	50.70	44.08
EI-MUNIT(21') [6]	62.73	29.83	48.32	46.96
RPLL (21') [46]	65.11	13.17	51.75	43.34
IGNet (21') [21]	70.65	37.60	52.97	53.74
Ours (SimT)	**76.24**	**39.83**	**58.94**	**58.34**
SFDA				
Source only	45.97	22.08	38.92	35.66
SHOT (20') [18]	68.44	22.90	47.95	46.43
SFDA (21') [23]	66.05	24.90	45.79	45.58
SFDASeg (21') [15]	67.90	35.31	52.46	51.89
Ours (SimT)	**75.56**	**36.46**	**53.66**	**55.23**

Experiments

- **Quantitative Results**
  - Ablation study
  - SimT outperforms existing ST-based methods
  - Robustness to various types of noise

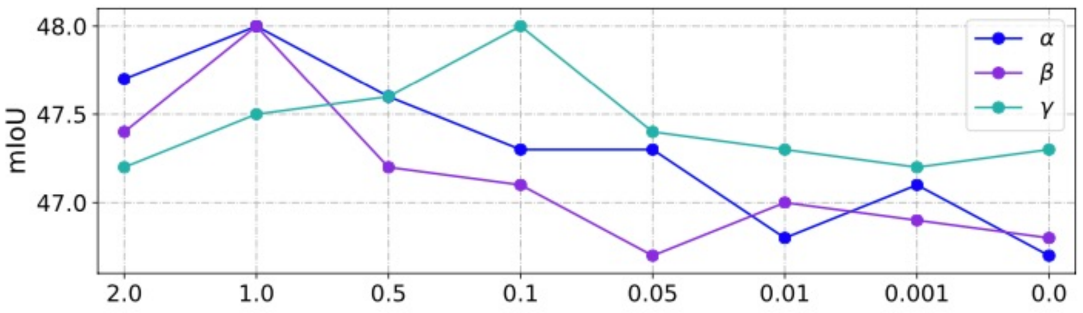


Figure 3. Ablation study for volume regularization, anchor guidance and convex guarantee of SimT.

Table 3. Ablation study. 'Pseudo Label' denotes we employ pseudo labels generated by the corresponding UDA model.

Method	Pseudo Label	GTA5 → Cityscapes	Δ
AdaptSegNet (CVPR18') [34]	—	42.4	–
Self-training (MRENT, ICCV 19' [49])	AdaptSegNet	45.1	2.7
Self-training (Threshold, ECCV18' [50])	AdaptSegNet	44.4	2.0
Self-training (Ucertainty, IJCV21' [46])	AdaptSegNet	46.1	3.7
Self-training (MetaCorrection, CVPR21' [11])	AdaptSegNet	47.3	4.9
Ours (w/o $\mathcal{L}_{Volume}^{SimT}$)	AdaptSegNet	46.8	4.4
Ours (w/o $\mathcal{L}_{Anchor}^{SimT}$)	AdaptSegNet	46.9	4.5
Ours (w/o $\mathcal{L}_{Aux}$)	AdaptSegNet	46.7	4.3
Ours (w/o $\mathcal{L}_{Convex}^{SimT}$)	AdaptSegNet	47.3	4.9
Ours	AdaptSegNet	**48.0**	**5.6**
LTIR (CVPR20') [14]	—	50.2	–
Self-training (MetaCorrection, CVPR21' [11])	LTIR	52.1	1.9
Ours	LTIR	**52.3**	**2.1**
DSP (AAAI21'') [9]	—	55.0	–
Self-training (MetaCorrection, CVPR21' [11])	DSP	56.4	1.4
Ours	DSP	**57.3**	**2.3**
BAPA-Net (ICCV21') [22]	—	57.4	–
Self-training (MetaCorrection, CVPR21' [11])	BAPA-Net	57.7	0.3
Ours	BAPA-Net	**58.9**	**1.5**

Experiments

- **Quantitative Results**
  - Ablation study
  - SimT outperforms existing ST-based methods
  - Robustness to various types of noise

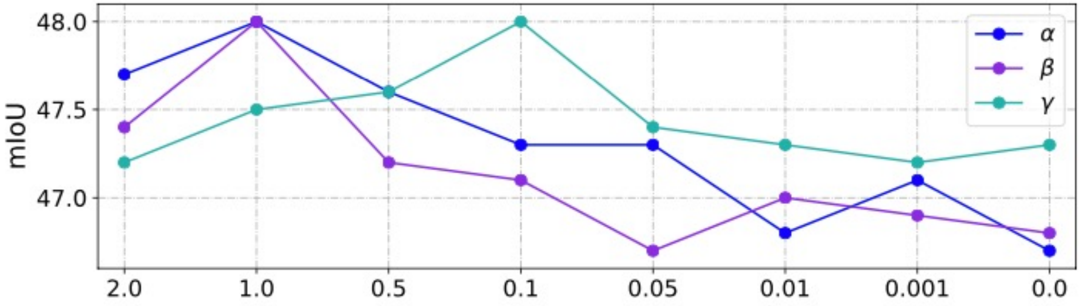


Figure 3. Ablation study for volume regularization, anchor guidance and convex guarantee of SimT.

Table 3. Ablation study. 'Pseudo Label' denotes we employ pseudo labels generated by the corresponding UDA model.

Method	Pseudo Label	GTA5 → Cityscapes	Δ
AdaptSegNet (CVPR18') [34]	—	42.4	–
Self-training (MRENT, ICCV 19' [49])	AdaptSegNet	45.1	2.7
Self-training (Threshold, ECCV18' [50])	AdaptSegNet	44.4	2.0
Self-training (Ucertainty, IJCV21' [46])	AdaptSegNet	46.1	3.7
Self-training (MetaCorrection, CVPR21' [11])	AdaptSegNet	47.3	4.9
Ours (w/o $\mathcal{L}_{Volume}^{SimT}$)	AdaptSegNet	46.8	4.4
Ours (w/o $\mathcal{L}_{Anchor}^{SimT}$)	AdaptSegNet	46.9	4.5
Ours (w/o $\mathcal{L}_{Aux}$)	AdaptSegNet	46.7	4.3
Ours (w/o $\mathcal{L}_{Convex}^{SimT}$)	AdaptSegNet	47.3	4.9
Ours	AdaptSegNet	**48.0**	**5.6**
LTIR (CVPR20') [14]	—	50.2	–
Self-training (MetaCorrection, CVPR21' [11])	LTIR	52.1	1.9
Ours	LTIR	**52.3**	**2.1**
DSP (AAAI21'') [9]	—	55.0	–
Self-training (MetaCorrection, CVPR21' [11])	DSP	56.4	1.4
Ours	DSP	**57.3**	**2.3**
BAPA-Net (ICCV21') [22]	—	57.4	–
Self-training (MetaCorrection, CVPR21' [11])	BAPA-Net	57.7	0.3
Ours	BAPA-Net	**58.9**	**1.5**

Experiments

- **Quantitative Results**
  - Ablation study
  - SimT outperforms existing ST-based methods
  - Robustness to various types of noise

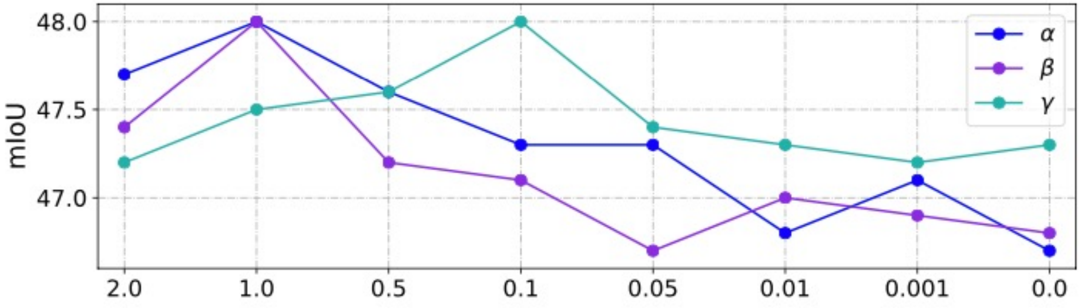


Figure 3. Ablation study for volume regularization, anchor guidance and convex guarantee of SimT.

Table 3. Ablation study. 'Pseudo Label' denotes we employ pseudo labels generated by the corresponding UDA model.

Method	Pseudo Label	GTA5 → Cityscapes	Δ
AdaptSegNet (CVPR18') [34]	—	42.4	–
Self-training (MRENT, ICCV 19' [49])	AdaptSegNet	45.1	2.7
Self-training (Threshold, ECCV18' [50])	AdaptSegNet	44.4	2.0
Self-training (Ucertainty, IJCV21' [46])	AdaptSegNet	46.1	3.7
Self-training (MetaCorrection, CVPR21' [11])	AdaptSegNet	47.3	4.9
Ours (w/o $\mathcal{L}^{SimT}_{Volume}$)	AdaptSegNet	46.8	4.4
Ours (w/o $\mathcal{L}^{SimT}_{Anchor}$)	AdaptSegNet	46.9	4.5
Ours (w/o $\mathcal{L}_{Aux}$)	AdaptSegNet	46.7	4.3
Ours (w/o $\mathcal{L}^{SimT}_{Convex}$)	AdaptSegNet	47.3	4.9
Ours	AdaptSegNet	**48.0**	**5.6**
LTIR (CVPR20') [14]	—	50.2	–
Self-training (MetaCorrection, CVPR21' [11])	LTIR	52.1	1.9
Ours	LTIR	**52.3**	**2.1**
DSP (AAAI21") [9]	—	55.0	–
Self-training (MetaCorrection, CVPR21' [11])	DSP	56.4	1.4
Ours	DSP	**57.3**	**2.3**
BAPA-Net (ICCV21') [22]	—	57.4	–
Self-training (MetaCorrection, CVPR21' [11])	BAPA-Net	57.7	0.3
Ours	BAPA-Net	**58.9**	**1.5**

Experiments

- **Qualitative Results**
  - Results of UDA on adapting GTA5 to Cityscapes
  - Results of UDA on adapting Endovis17 to Endovis18

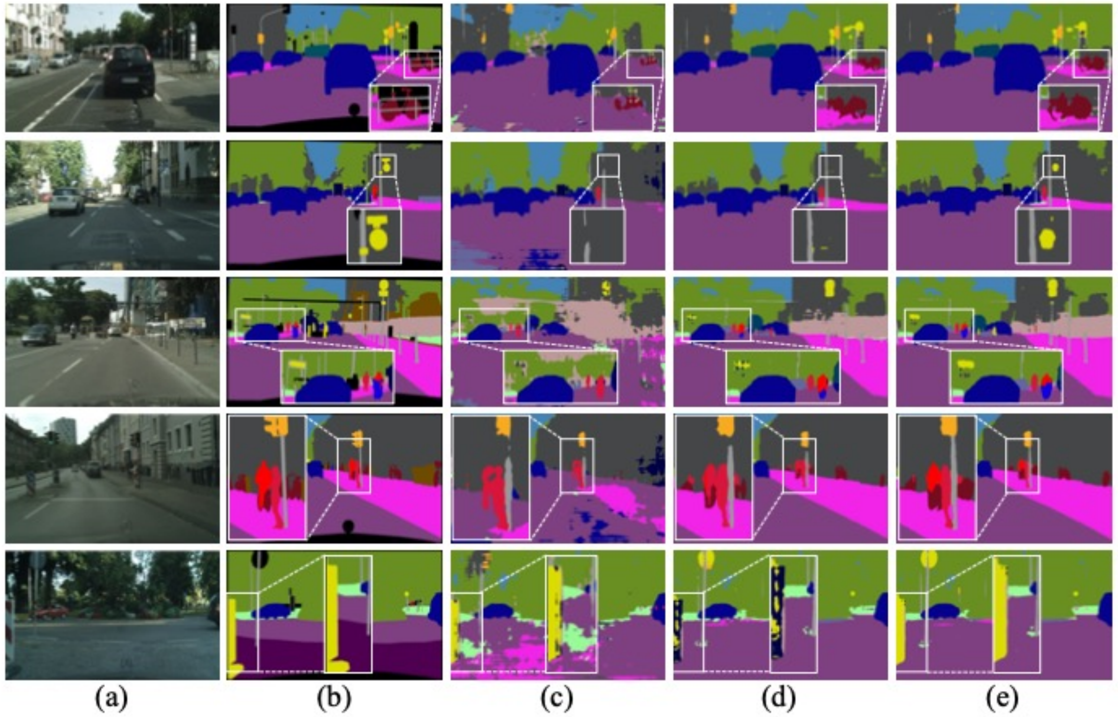


Figure 4. Qualitative results of UDA in GTA5→Cityscapes scenario. (a) Target image, (b) Ground truth, Predictions from (c) source only model, (d) BAPA-Net [22], (e) ours (SimT).

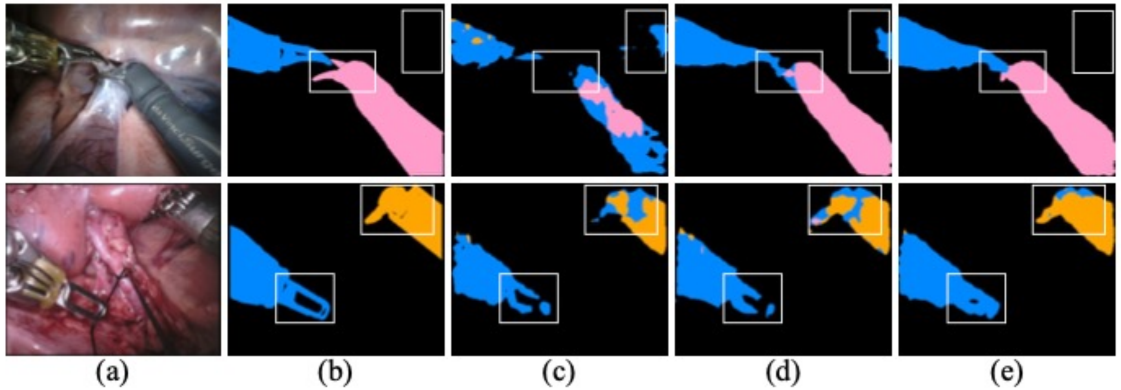


Figure 5. Qualitative results of UDA in Endovis17→Endovis18 scenario. (a) Target image, (b) Ground truth, Predictions from (c) source only model, (d) IGNet [21], (e) ours (SimT).

Experiments

- **Qualitative Results**
  - Results of SFDA on adapting GTA5 to Cityscapes
  - Results of SFDA on adapting Endovis17 to Endovis18

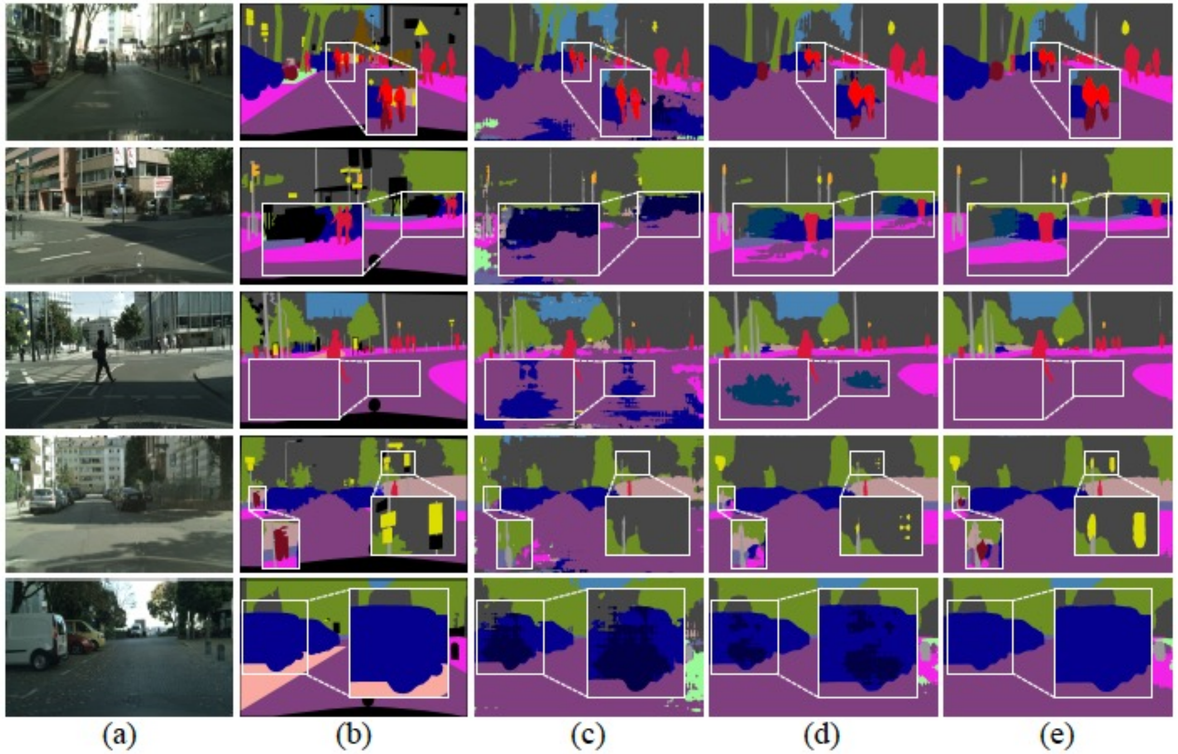


Figure 2. Qualitative results of SFDA in GTA5→Cityscapes scenario. (a) Target image, (b) Ground truth, Predictions from (c) source only model, (d) SFDASeg [6], (e) ours (SimT).

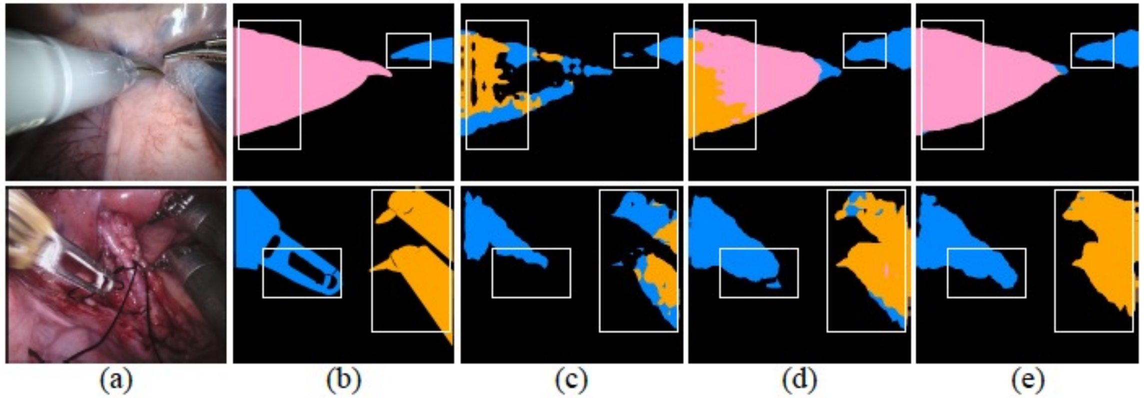


Figure 3. Qualitative results of SFDA in Endovis17→Endovis18 scenario. (a) Target image, (b) Ground truth, Predictions from (c) source only model, (d) SFDASeg [6], (e) ours (SimT).

Experiments

- **Qualitative Results**
  - Visualization of SimT
  - Visualization of detecting open-set class pixels

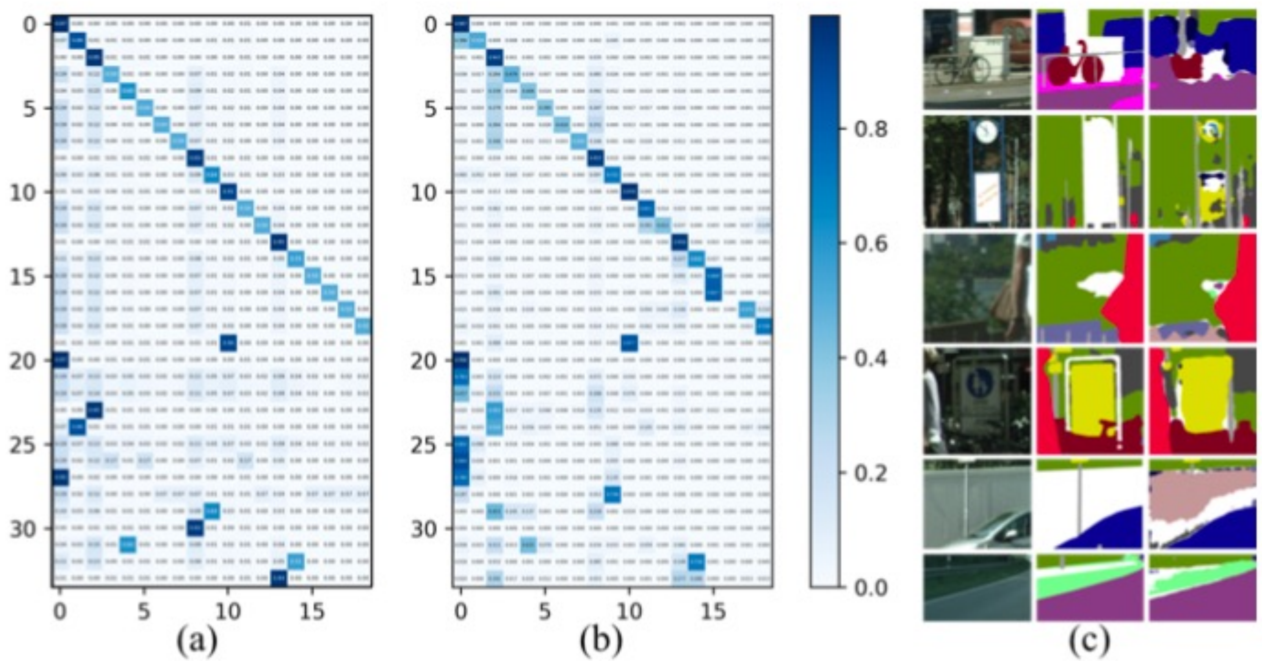


Figure 6. Visualization results. (a) The learned SimT, (b) confusion matrix of pseudo labels, (c) open-set class visualization.

Summarization

Department of Electrical Engineering
香港城市大學
City University of Hong Kong

THE UNIVERSITY OF SYDNEY

- We present a general DA method to robustly learn from noisy pseudo-labeled target data, including closed-set and open-set label noises.
- We model the closed- and open-set noise distributions of pseudo labels by SimT, and use geometry analysis to estimate it, which is further used to benefit loss correction for noisy target data learning.
- SimT is a plug-and-play technique that can be applied to a wide range of tasks, such as UDA and SFDA, to solve the pseudo label noise issue.
- **Future work:** introduce open-set class prior to enable the semantic separation inner open-set regions.

Thanks!

Emails: xqguo.ee@my.cityu.edu.hk

Code: