2023 동계 세미나

Domain adaptation by contrastive learning

Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

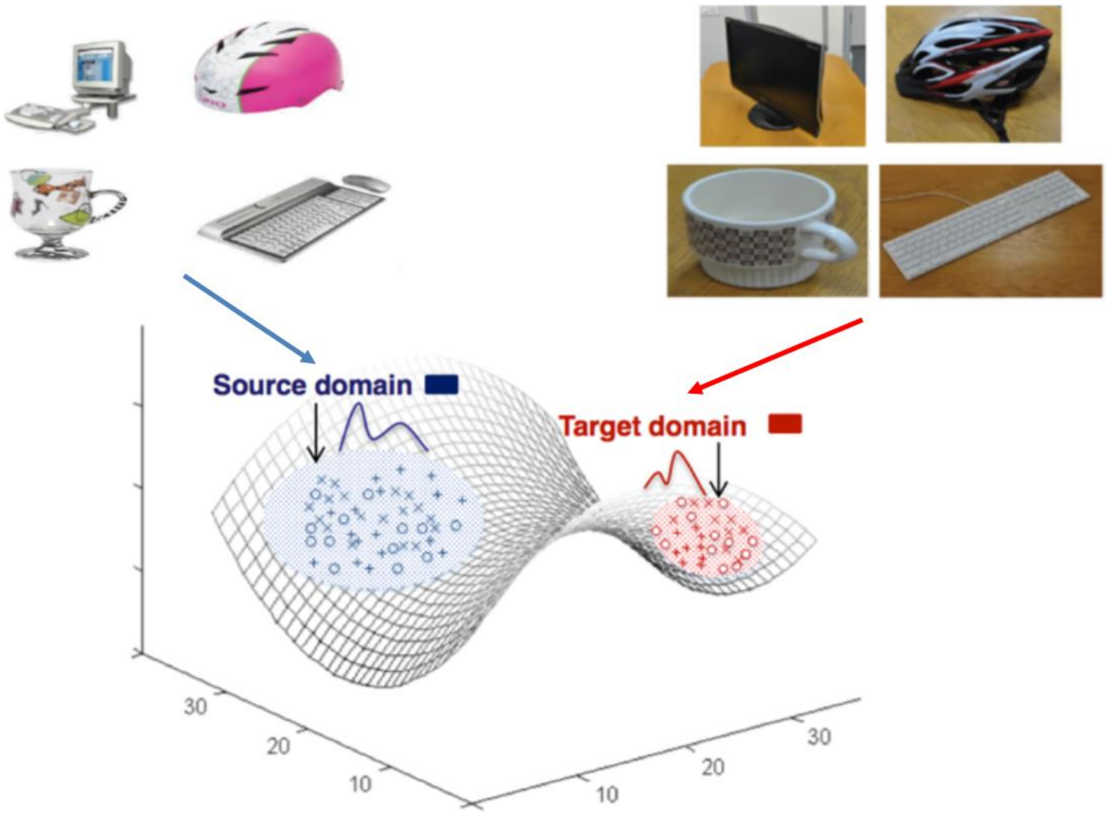
Presented by
전창렬

Outline

- Background
  - Domain adaptation
  - Contrastive learning
- Domain adaptation by contrastive learning
  - Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation (ECCV 2022)
  - Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation (ECCV 2022)
- Conclusion

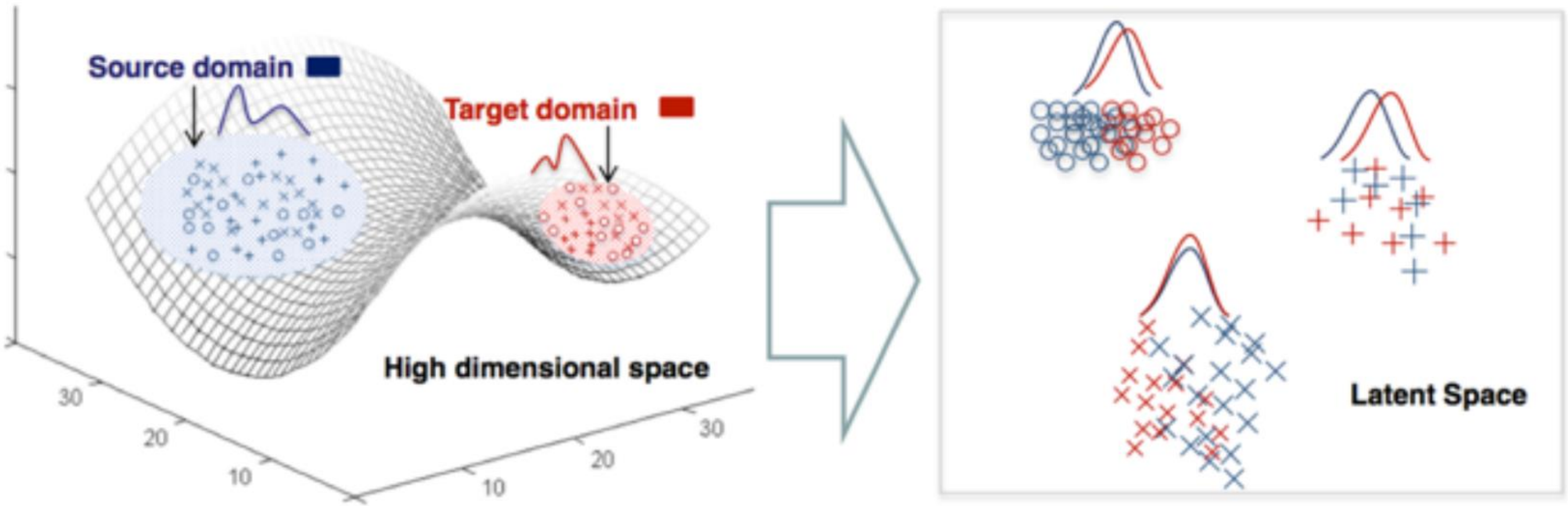
Background

- Domain Adaptation
  - Domain gap



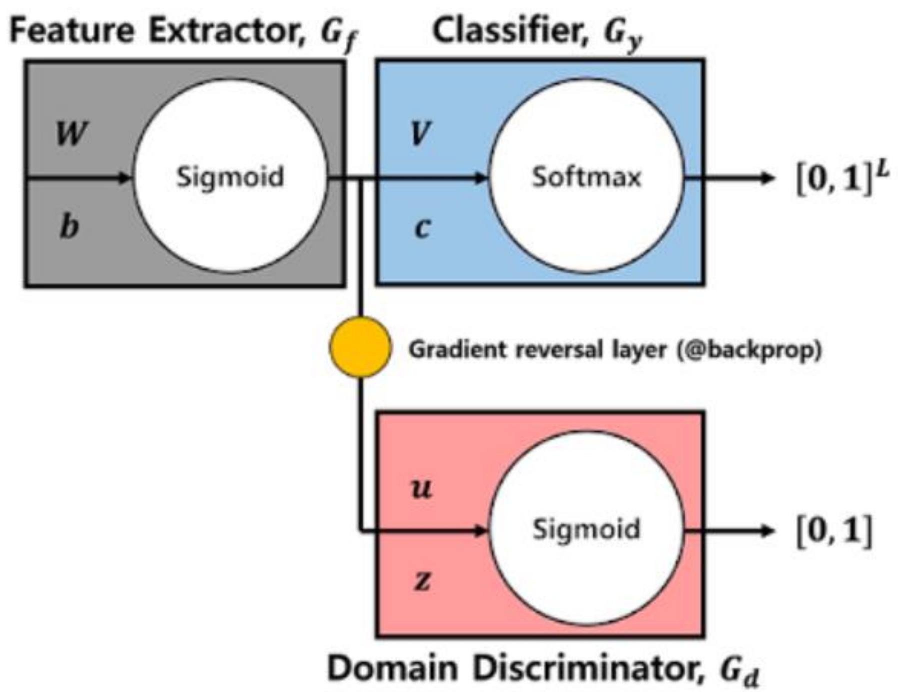
Background

- Domain Adaptation
  - Domain adaptation



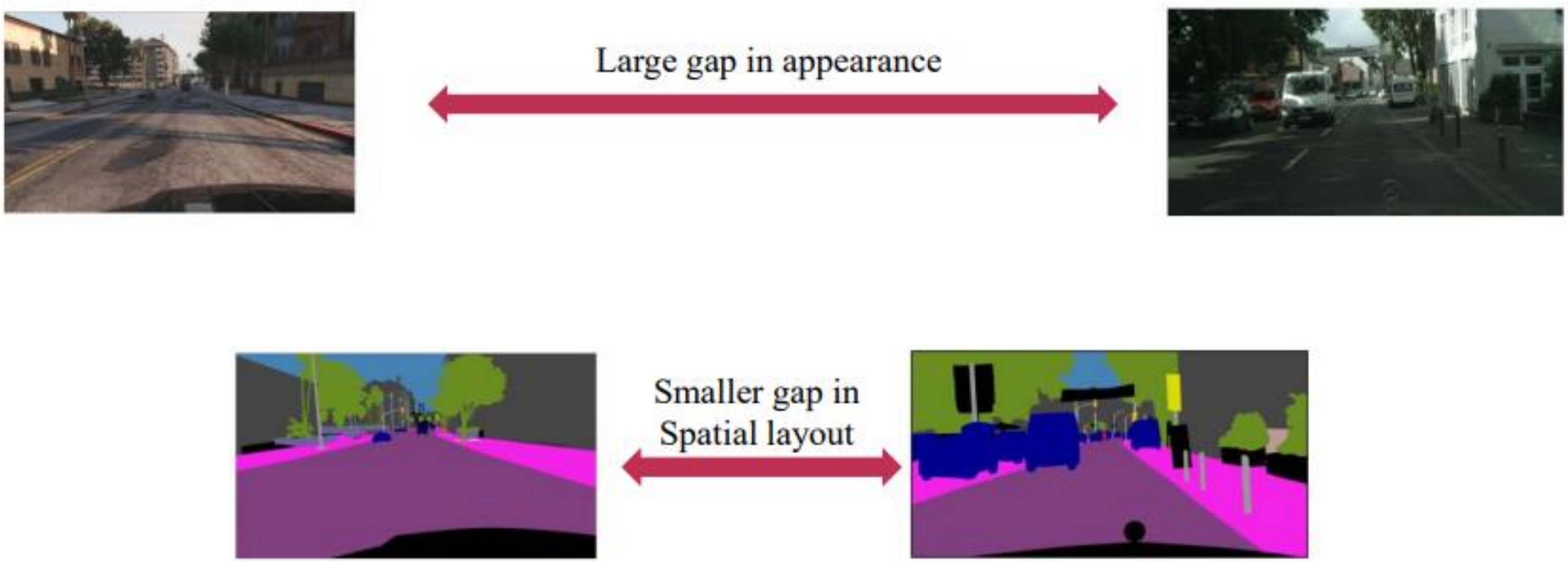
Background

- Domain Adaptation
  - Adversarial training
    - Target task(classification)의 역할은 잘 하도록 유지
    - Sample의 feature representation이 source domain에서 왔는지 target domain에서 왔는지를 구별 못하게 domain discriminator를 약화하는 방향으로 학습



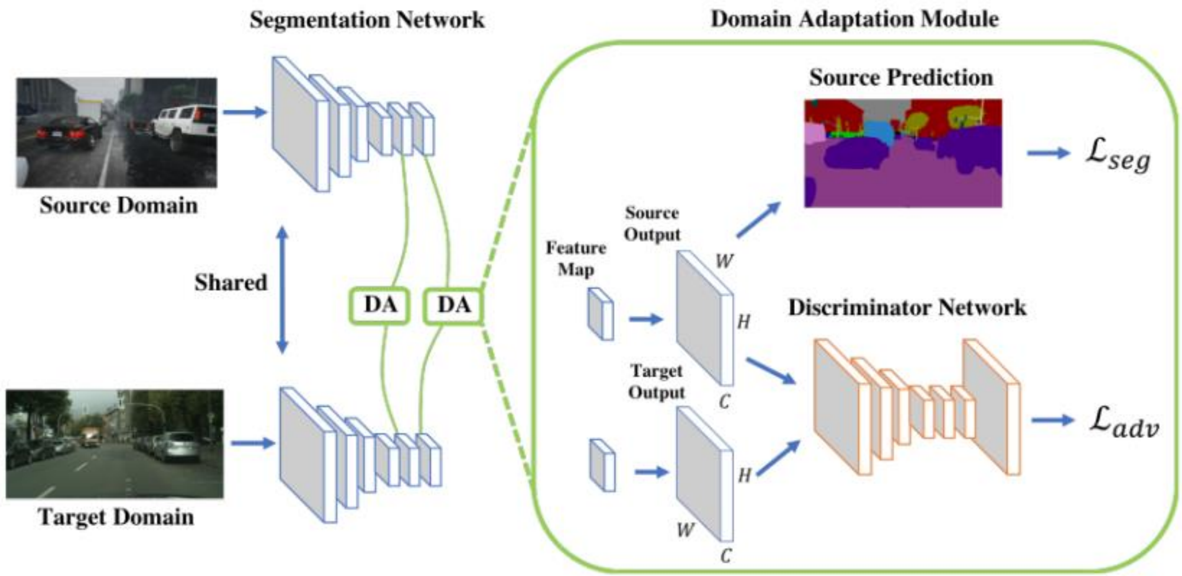
Background

- Learning to Adapt Structured Output Space for Semantic Segmentation (AdaptSegNet)
  - Motivation
    - Raw image들과 달리 같은 network를 통과한 segmentation output간에는 domain gap이 더 줄어들 것



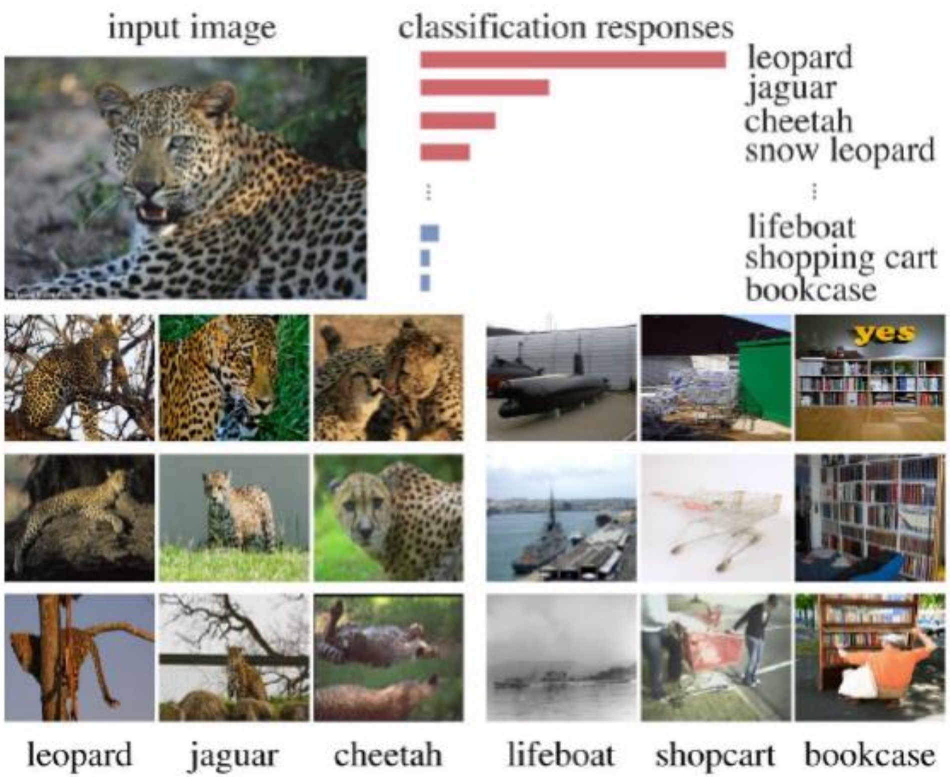
Background

- Learning to Adapt Structured Output Space for Semantic Segmentation (AdaptSegNet)
  - 학습과정
    - Source domain image를 사용해 segmentation network를 학습
    - 학습된 segmentation network에 target 이미지를 넣어 target prediction을 얻음
    - 동일한 network를 통과하여 얻어진 각각의 output에 대하여 adversarial loss를 계산



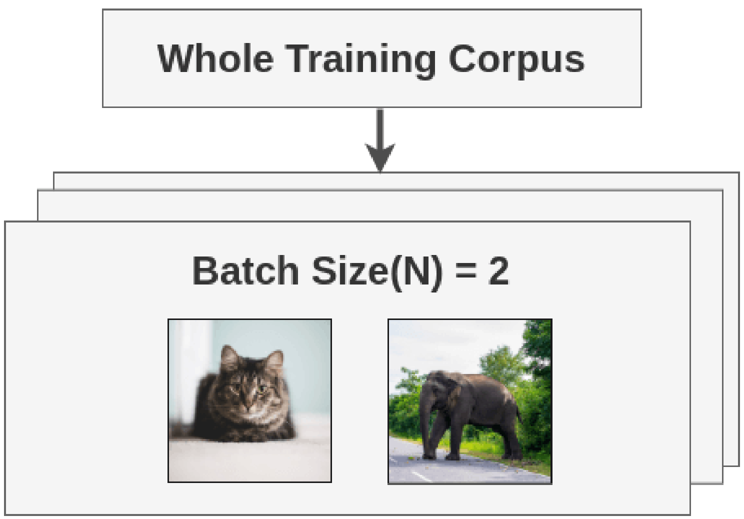
Background

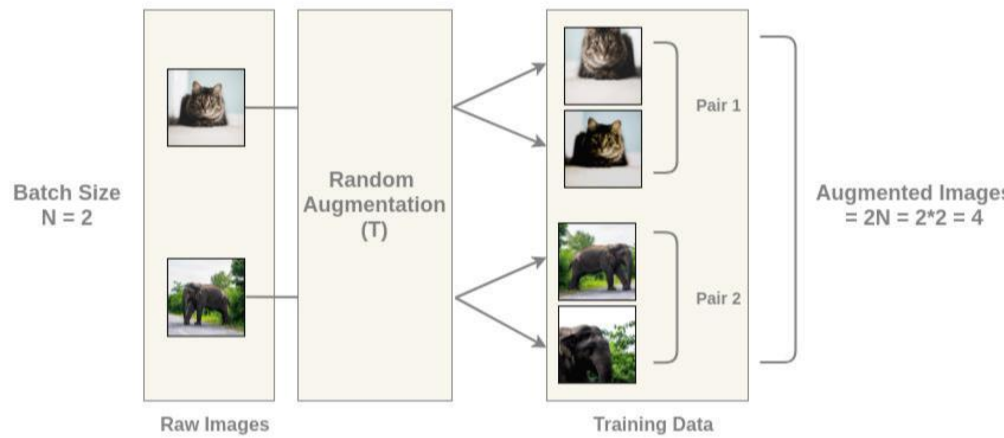
- Contrastive learning
  - Motivation
    - 잘 추출된 feature 값들은 instance간의 유사도 정보를 가지고 있을 것이라는 가정



Background

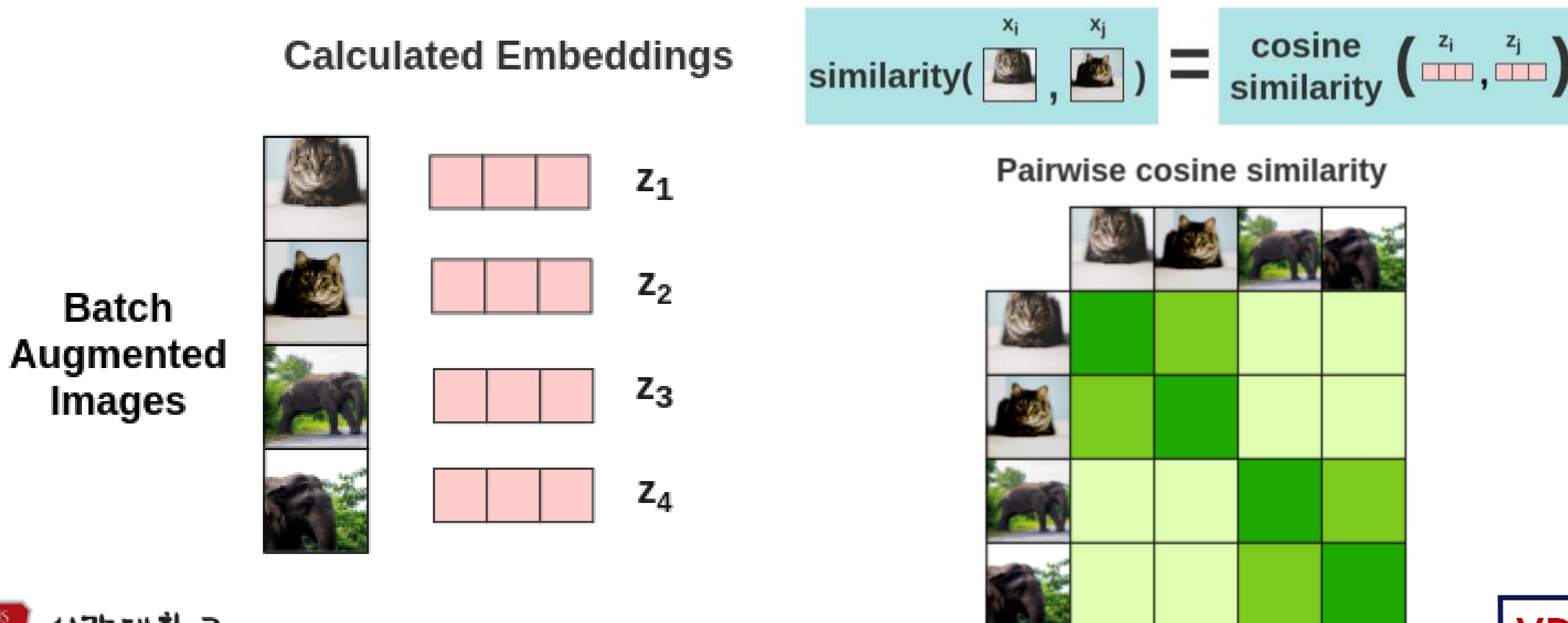
- Contrastive learning(SimCLR)
  - Training 과정
    - Label이 없는 전체 whole training corpus에서 크기가 N(아래 예시에서 2)의 batch를 생성
    - Data augmentation을 적용하여 batch의 각 이미지에 대해 2개의 이미지 쌍을 얻음





Background

- Contrastive learning(SimCLR)
  - Training 과정
    - 각각의 이미지가 네트워크를 통과하여 feature embedding z를 획득
    - Feature embedding z 간의 similarity를 계산
      - 같은 이미지에서 나온 이미지 간 similarity가 높게 나타남

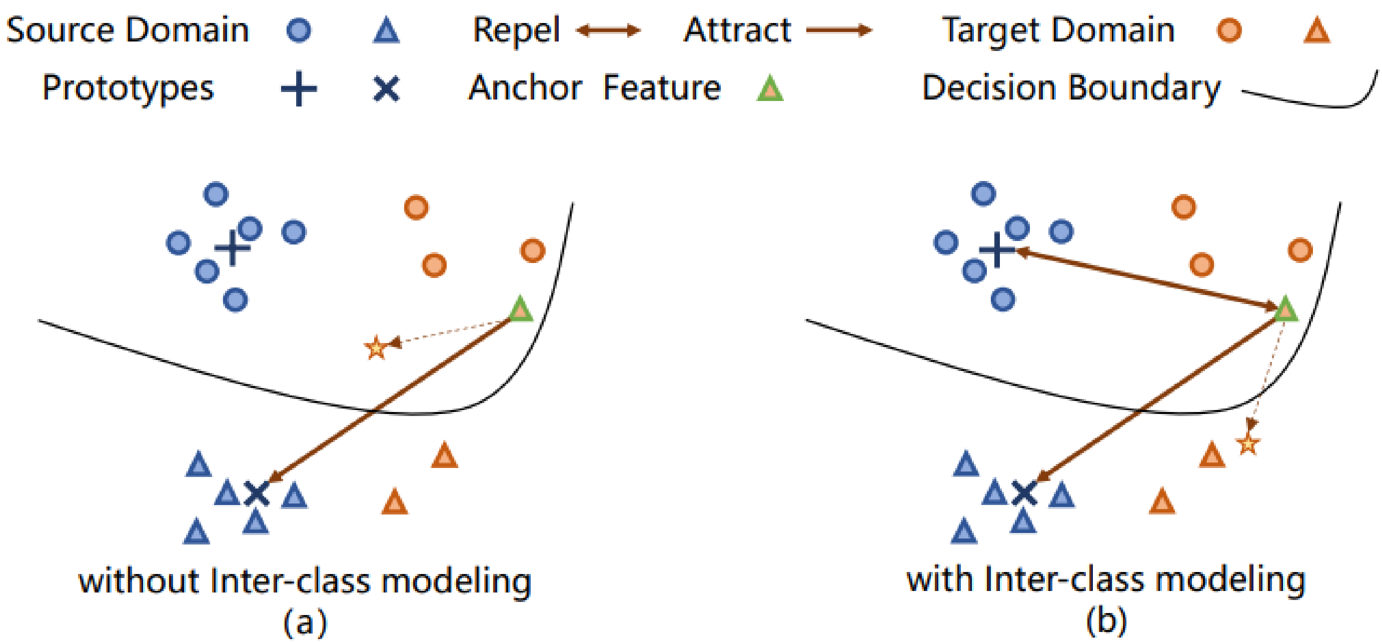


Background

- Contrastive learning(SimCLR)
  - Training 과정
    - Similarity를 loss로 가짐으로써 positive pair 간에는 similarity가 크게, negative pair 간에는 similarity가 작아지는 방향으로 학습이 진행

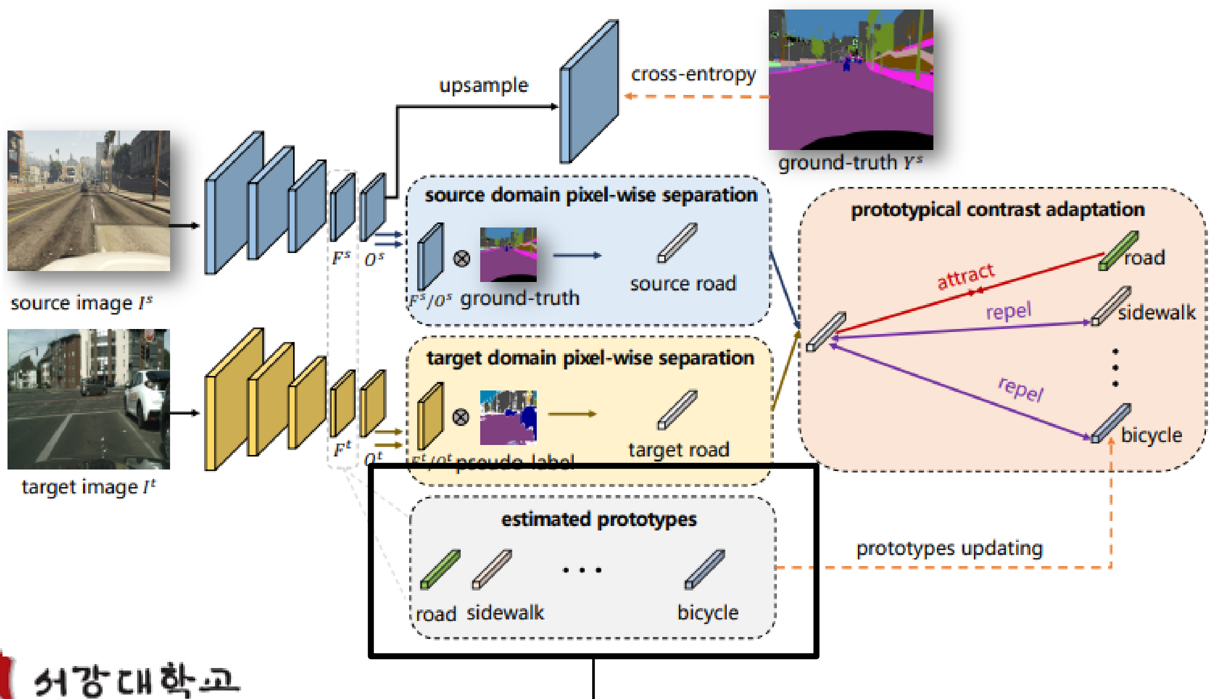
Domain adaptation by contrastive learning

- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Motivation
    - adversarial training을 통한 feature alignment는 target domain에서 class 별로 분리 되어야 한다는 요소를 고려하지 않는 adaptation 방법
    - Class 간의 정보를 고려하기위해 contrastive learning 방법을 도입



Domain adaptation by contrastive learning

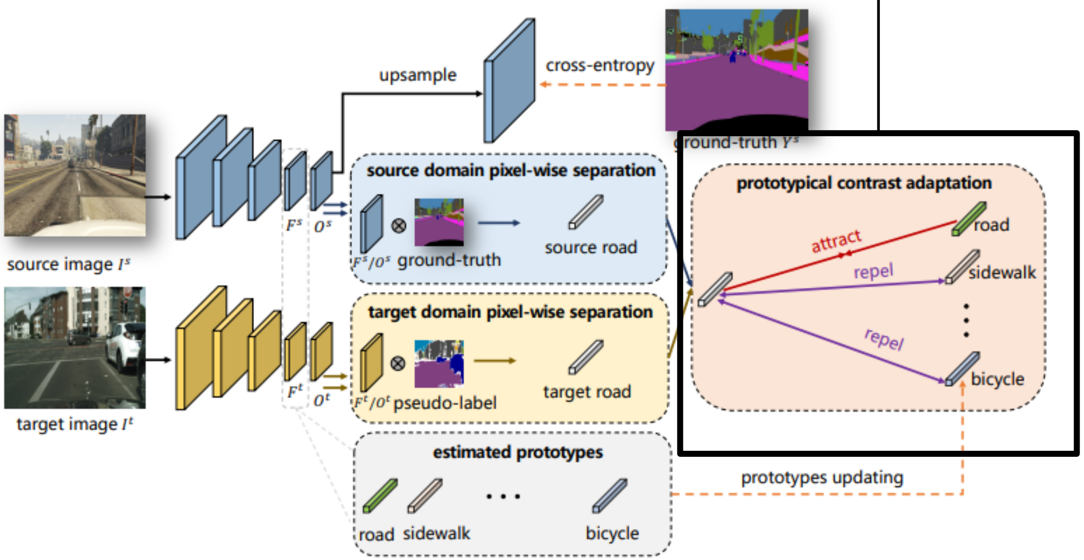
- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Prototypical contrast adaptation
    - Prototypes initialization
      - 모델을 source domain에서 학습을 진행 한 후 class-aware prototypes을 계산



$$\mathbf{p}_c^{feat} = \frac{\sum_{n=1}^{N_s}\sum_{i=1}^{H}\sum_{j=1}^{W} F^s_{n,i,j}\,\mathbb{1}[Y^s_{n,i,j}=c]}{\sum_{n=1}^{N_s}\sum_{i=1}^{H}\sum_{j=1}^{W}\mathbb{1}[Y^s_{n,i,j}=c]}$$

Domain adaptation by contrastive learning

- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Prototypical contrast adaptation
    - Contrast adaptation
      - Target domain feature들이 각각의 source domain에서 얻어진 prototype과 contrastive learning을 진행함



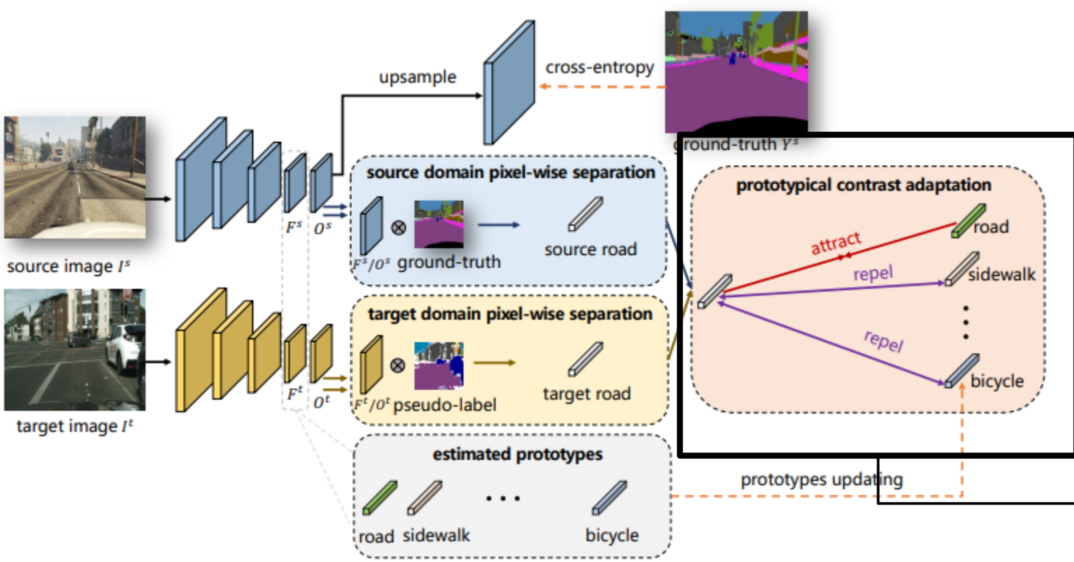
$$P^{t\to s}_{n,i,j,c} = \frac{\exp(\mathbf{p}^{feat}_c \cdot F^t_{n,i,j}/\tau)}{\sum_{c=1}^{C}\exp(\mathbf{p}^{feat}_c \cdot F^t_{n,i,j}/\tau)},$$

$$\mathcal{L}^{t\to s}_n = -\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{c=1}^{C}\tilde{y}^t_{n,i,j,c}\log P^{t\to s}_{n,i,j}$$

	C1	C2	C3	C4
Channel-wise contrastive loss	0.7	0.4	0.9	0.8
Pseudo label	0	0	0	1

Domain adaptation by contrastive learning

- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Prototypical contrast adaptation
    - Contrast adaptation
      - Source domain data에 대해서도 이와 같은 loss를 적용하여 source domain 간의 intra class variation을 추가로 확보
      - 최종적으로는 t → $s, s \to s$ 두 방향으로 적용한 loss의 합을 통해 학습 진행

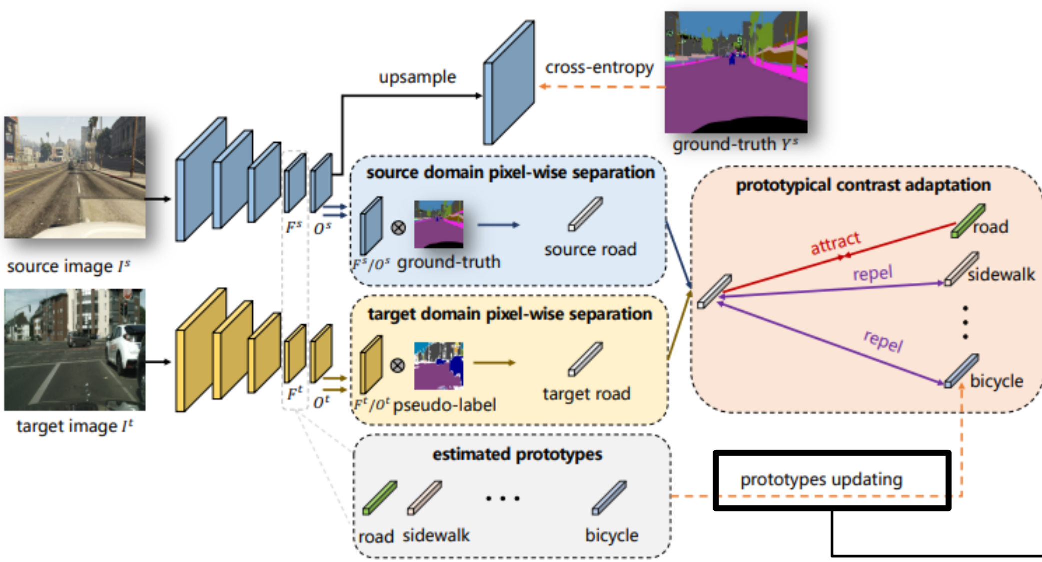


$$\mathcal{L}_n^{s\to s} = -\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{c=1}^{C} y^s_{n,i,j,c}\log P^{s\to s}_{n,i,j,c},$$

$$\mathcal{L}_{\text{ContraFeat}} = \sum_{n=1}^{N_t}\mathcal{L}_n^{t\to s} + \sum_{n=1}^{N_s}\mathcal{L}_n^{s\to s}$$

Domain adaptation by contrastive learning

- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Prototypes Updating
    - Strict statistical mean을 활용하여 새로 들어오는 데이터에 대하여 지속적으로 prototype update를 진행
    - Target domain의 정보를 활용하기 위해 target domain에서 prototype을 계산 한 후 prototype 간 convex combination으로 update를 진행

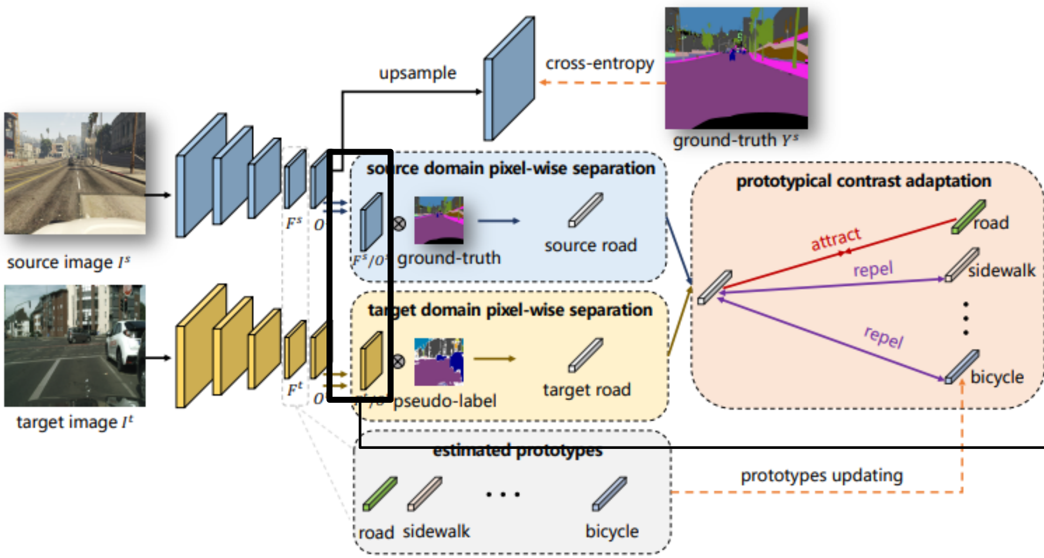


$$\mathbf{p}_c^{feat} \leftarrow \frac{\mathbf{p}_c^{feat} n_c^{feat} + \tilde{\mathbf{p}}_c^{feat} \tilde{n}_c^{feat}}{n_c^{feat} + \tilde{n}_c^{feat}},$$

$$\mathbf{p}_c^{feat} \leftarrow m\mathbf{p}_c^{feat^s} + (1-m)\mathbf{p}_c^{feat^t},$$

Domain adaptation by contrastive learning

- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Label space adaptation
    - 앞선 과정을feature level에서 적용을 진행하였으며, 이에 추가로 본 논문에서는 label space에서 해당 과정을 같이 진행



$$\mathcal{L}_{\text{Contra}} = \mathcal{L}_{\text{ContraFeat}} + \mathcal{L}_{\text{ContraOut}}.$$

Domain adaptation by contrastive learning

- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Class-wise adaptive Pseudo-Label Thresholds
    - Source와 target domain간 유사한 class들만 confidence가 높게 나타남
      - 이러한 상황에서 모든 class에 일관적인 threshold를 적용시 target domain에서 특정 class들만 학습에 사용되는 문제가 발생
      - class 별로 다른 threshold를 적용함으로써 해당 문제를 해결
        - ✓ Class별 pixel들의 confidence를 높은 순으로 정렬 후 동일한 비율의 confidence를 취함으로써 class별로 다른 threshold를 얻을 수 있음

$$\mathbf{l}_c = \sum_{n=1}^{N_t} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbb{1}[\widetilde{Y}^t_{n,i,j} = c],$$

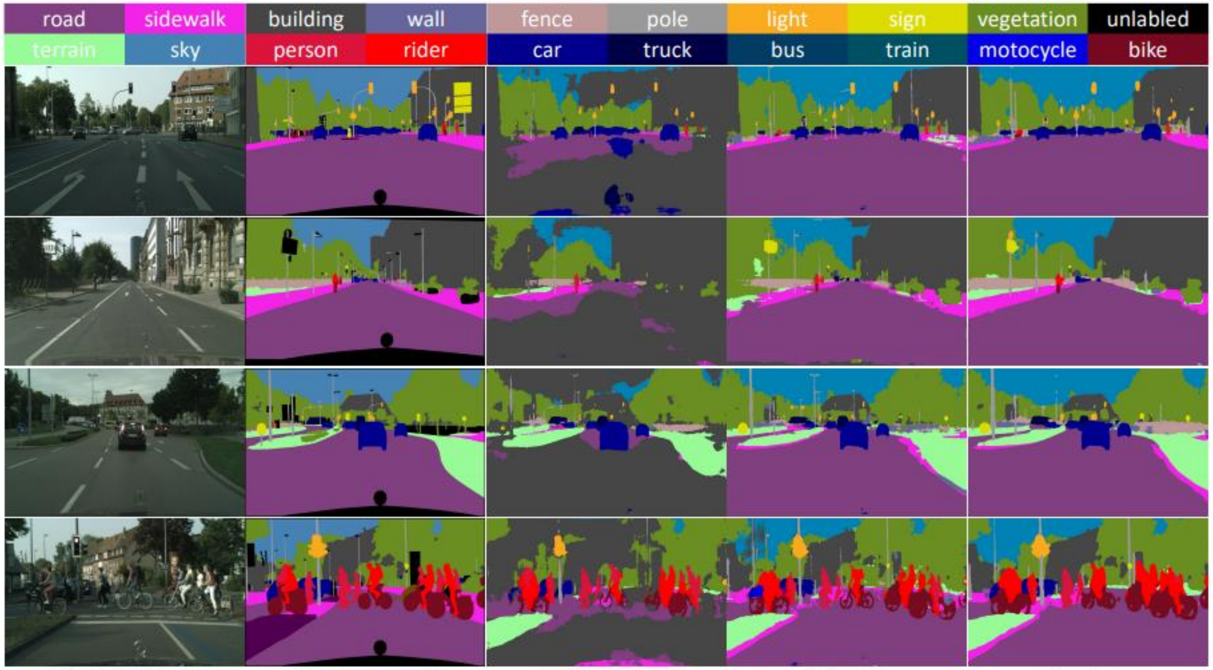
threshold

도로	0.99	0.97	0.95	0.90	0.87	0.84
표지판	0.6	0.55	0.54	0.50	0.45	0.40
자동차	0.87	0.85	0.81	0.77	0.72	0.68

Domain adaptation by contrastive learning

- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Experiments

Source Only	$s \to s$	$t \to s$	mIoU
✓			37.3
✓	✓		44.9
✓		✓	46.8
✓	✓	✓	48.8



(a) Target image (b) Ground truth (c) Source Only (d) FADA (e) Ours

Domain adaptation by contrastive learning

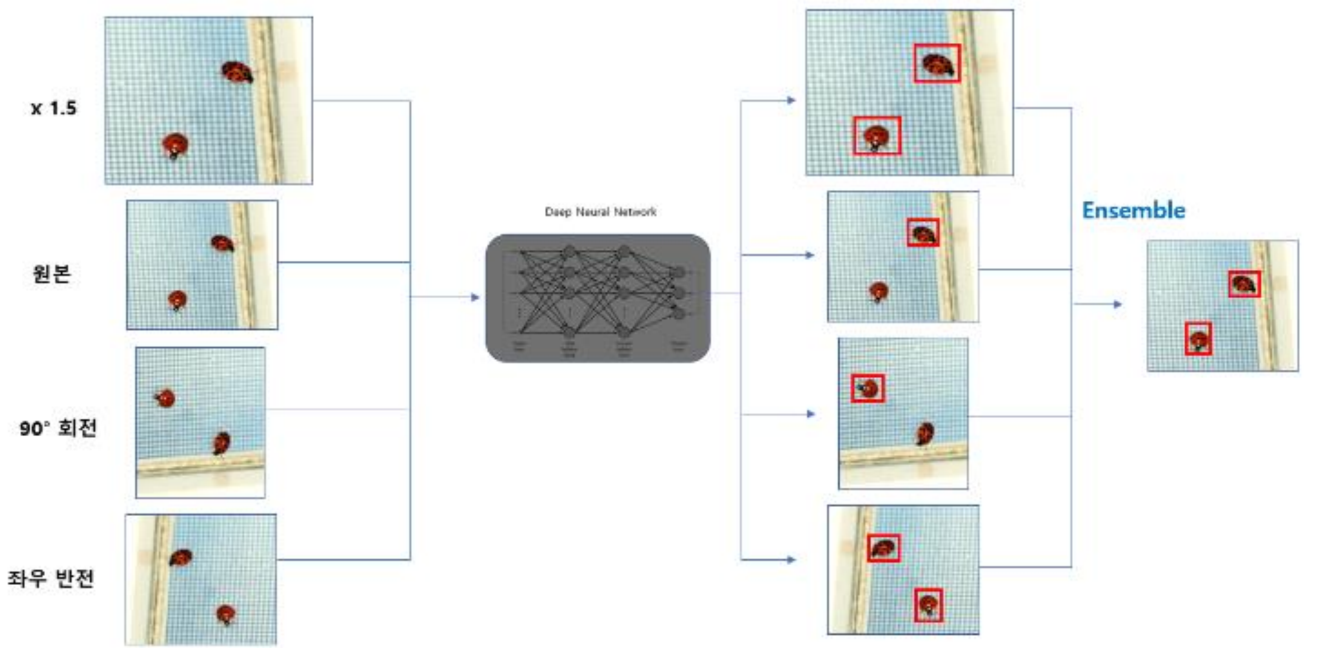
- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Experiments

?

Source Only	F	O	Ada-ST	MST	mIoU
✓					37.3
✓	✓				47.9
✓		✓			48.4
✓	✓	✓			48.8
✓			✓		43.9
✓	✓	✓	✓		55.1
✓	✓	✓	✓	✓	**56.3**

Domain adaptation by contrastive learning

- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Multi-scale testing
    - 여러 scale에서 testing을 진행 후 testing 결과를 ensemble을 통해 성능을 높이는 단순하면서 효과적인 기법



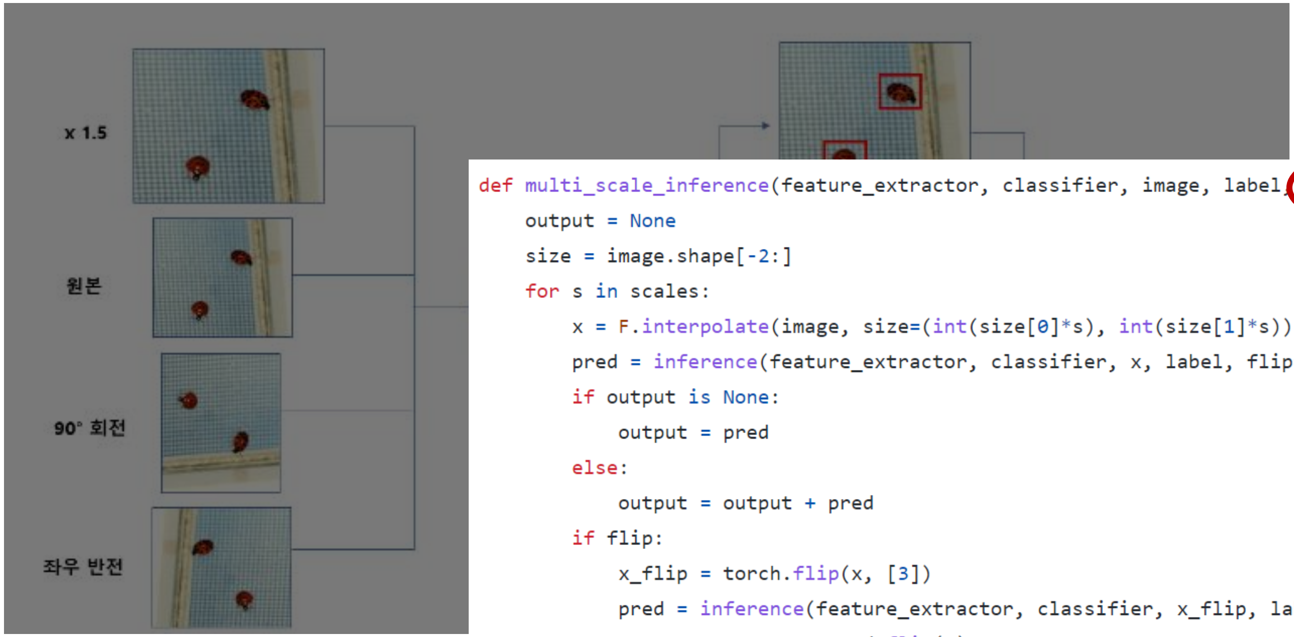

```
0,1.3], flip=True):
r', align_corners=True)
)
            output = output + pred.flip(3)
    if flip:
        return output/len(scales)/2
    return output/len(scales)
```

서강대학교 SOGANG UNIVERSITY

VDS LAB

Domain adaptation by contrastive learning

- Prototypical Contrast Adaptation for Domain Adaptive Semantic Segmentation
  - Multi-scale testing
    - 여러 scale에서 testing을 진행 후 testing 결과를 ensemble을 통해 성능을 높이는 단순하면서 효과적인 기법



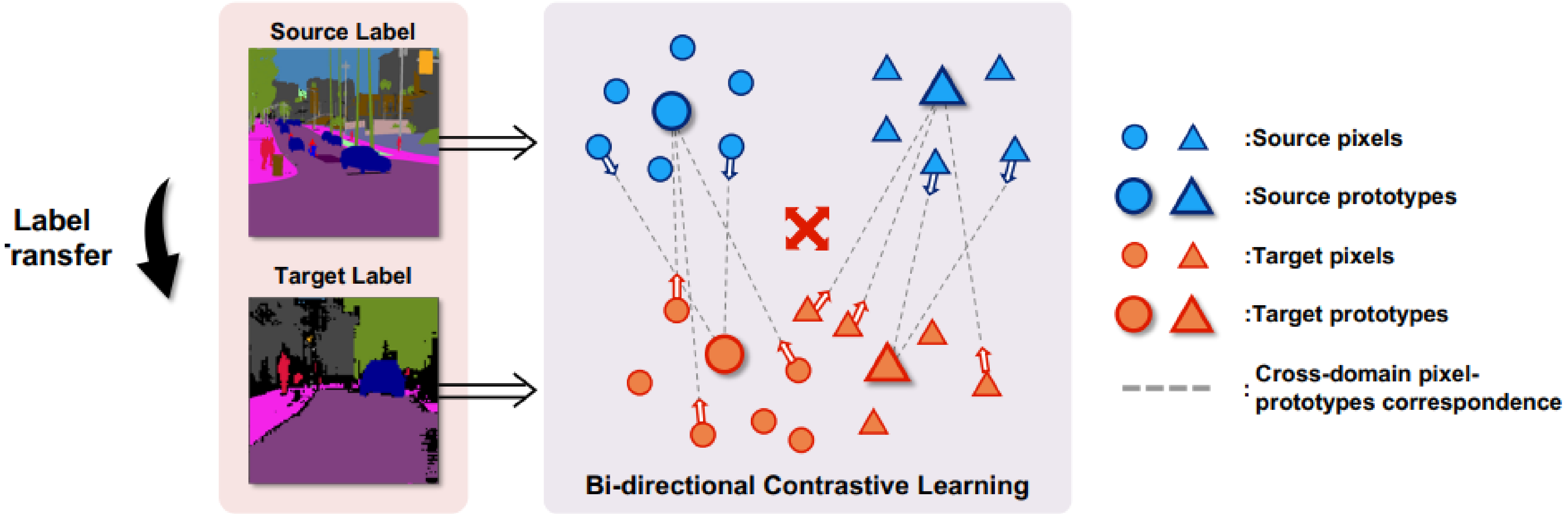

```
def multi_scale_inference(feature_extractor, classifier, image, label, scales=[0.7,1.0,1.3], flip=True):
    output = None
    size = image.shape[-2:]
    for s in scales:
        x = F.interpolate(image, size=(int(size[0]*s), int(size[1]*s)), mode='bilinear', align_corners=True)
        pred = inference(feature_extractor, classifier, x, label, flip=False)
        if output is None:
            output = pred
        else:
            output = output + pred
        if flip:
            x_flip = torch.flip(x, [3])
            pred = inference(feature_extractor, classifier, x_flip, label, flip=False)
            output = output + pred.flip(3)
    if flip:
        return output/len(scales)/2
    return output/len(scales)
```

서강대학교 SOGANG UNIVERSITY

VDS LAB

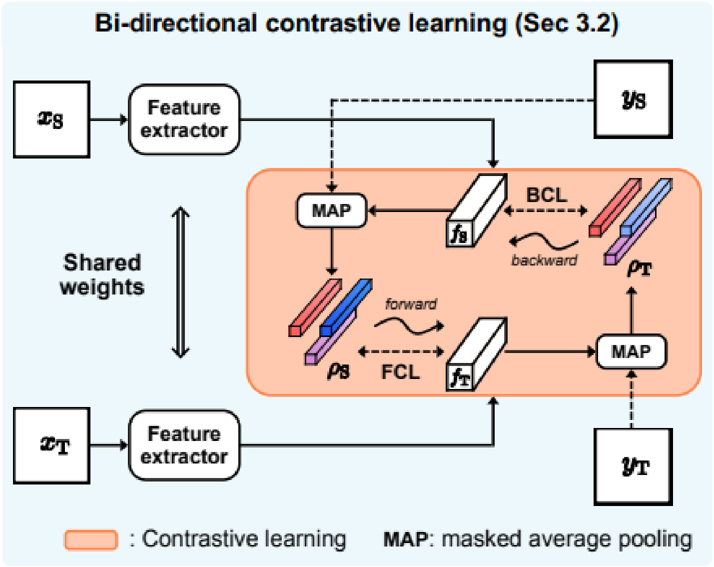
Domain adaptation by contrastive learning

- Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation
  - Motivation
    - 기존의 confidence 기반 pseudo labeling을 거치면 target domain에서 label이 sparse 하다는 특징이 있음
      - 초기 모델을 fitting 하는데에 있어 사용되는 sample이 적어지므로 부정확한 예측이 됨
    - Contrastive learning을 활용 시 compact한 feature 구성을 지니면서 discriminative 능력을 가짐



Domain adaptation by contrastive learning

- Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation
  - Bi-directional contrastive learning
    - FCL: source의 class별 prototype에 대한 target feature의 contrastive learning 진행
    - BCL: target의 class별 prototype에 대한 source feature의 contrastive learning 진행
    - FCL을 통해 target feature가 source의 특성을, BCL을 통해 source feature가 target의 특성을 공유하는 방식으로 학습이 진행되면서 gap을 줄이게 됨



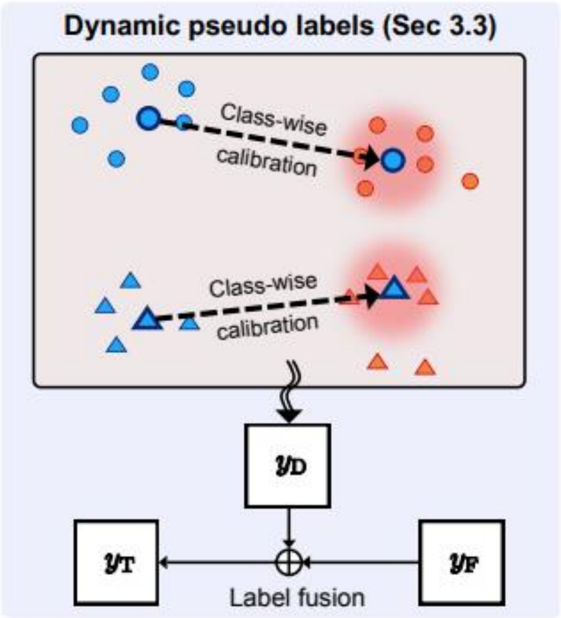
$$\rho_S(c) = \frac{\sum_p f_S(p) y_S(p,c)}{\sum_p y_S(p,c)}, \rho_T(c) = \frac{\sum_p f_T(p) y_T(p,c)}{\sum_p y_T(p,c)}$$

$$\mathcal{L}_{FC} = -\sum_c \sum_p y_T(p,c) \log \frac{\exp\left(s(f_T(p), \rho_S(c))/\tau\right)}{\sum_c \exp\left(s(f_T(p), \rho_S(c))/\tau\right)}$$

$$\mathcal{L}_{BC} = -\sum_c \sum_p y_S(p,c) \log \frac{\exp\left(s(f_S(p), \rho_T(c))/\tau\right)}{\sum_c \exp\left(s(f_S(p), \rho_T(c))/\tau\right)}$$

Domain adaptation by contrastive learning

- Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation
  - Dynamic pseudo labeling
    - Pseudo label update: 기존 prototype을 새로 들어온 데이터의 prototype과 EMA 업데이트를 통해 계속 업데이트
    - Class-wise domain bias: Source, target domain의 class별 prototype 간 거리
    - Calibrated prototypes: 해당 학습 iteration에 들어온 input image들의 prototype에 class-wise domain bias를 더함



$$\mu_S(c) \leftarrow \lambda\mu_S(c) + (1-\lambda)\rho_S(c),$$
$$\mu_T(c) \leftarrow \lambda\mu_T(c) + (1-\lambda)\rho_T(c),$$

$$\xi(c) = \mu_T(c) - \mu_S(c),$$

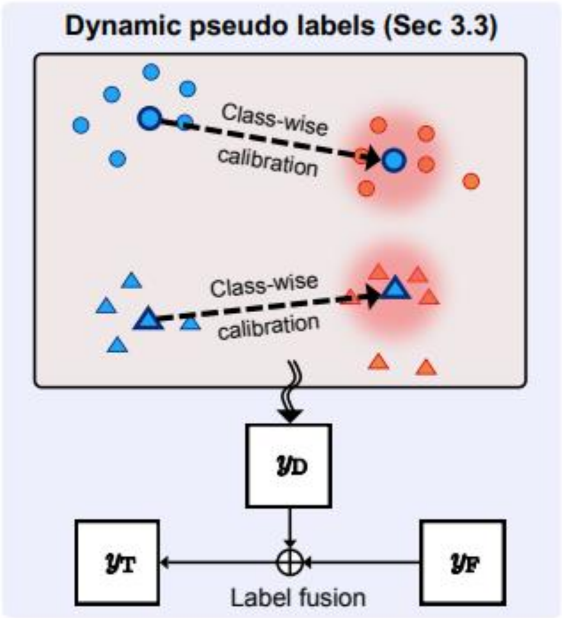
$$\rho_{S\to T}(c) = \rho_S(c) + \xi(c).$$

Sogang University 서강대학교

VDS LAB

Domain adaptation by contrastive learning

- Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation
  - Dynamic pseudo labeling
    - 기존의 source domain에 overfitting된 classifier를 통해 얻어진 prediction을 활용하는 label과 달리 dynamic pseudo label은 calibrate된 source prototype과 target feature의 similarity를 통해 labeling이 진행됨
    - 이를 통해 domain간의 correspondence가 더 향상된 labeling이 가능하므로 domain adaptation에 더 최적화

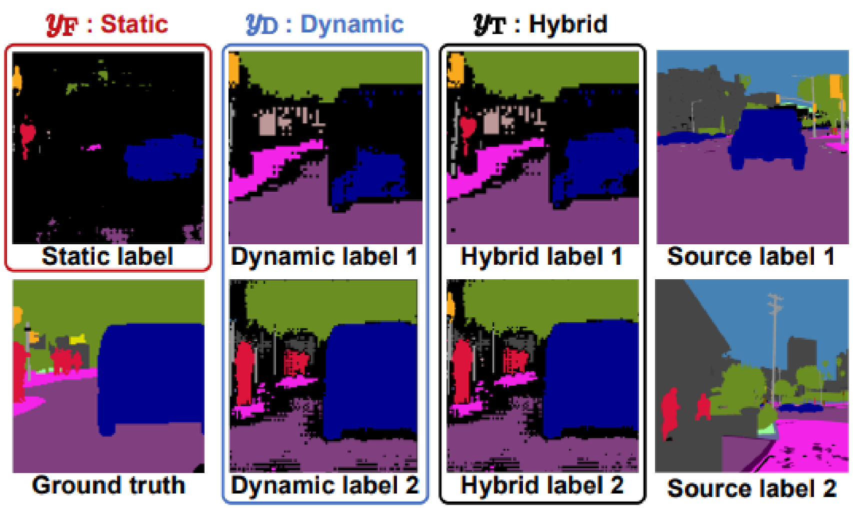


$$y_D(p, c) = \begin{cases} 1, & \text{if } s(f_T(p)), \rho_{S\rightarrow T}(c)) > \mathcal{T} \text{and } c = c' \\ 0, & \text{otherwise} \end{cases}$$

$$c' = \underset{c}{\operatorname{argmax}}(s(f_T(p)), \rho_{S\rightarrow T}(c))).$$

Domain adaptation by contrastive learning

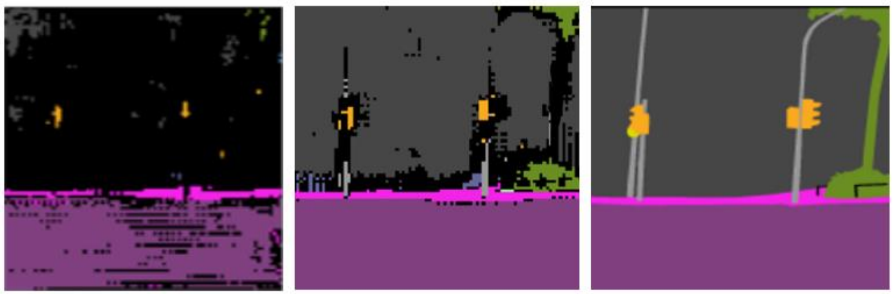
- Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation
  - Hybrid pseudo labels
    - Dynamic label은 기존 pseudo label에 비해 domain간 correspondence가 높으면서 dense한 label
    - 반면 기존 pseudo label은 dynamic label에 비해 더 reliable한 label
    - 두 label의 이점을 모두 취하기 위해 hybrid pseudo label을 취함



$$y_T(p,c) = \begin{cases} y_D(p,c), & \text{if } y_D(p,c) = 1 \\ y_F(p,c), & \text{if } y_D(p,c') = 0 \text{ for } c' \in \mathcal{C}, \text{ and } y_F(p,c) = 1 \\ 0, & \text{otherwise} \end{cases}$$

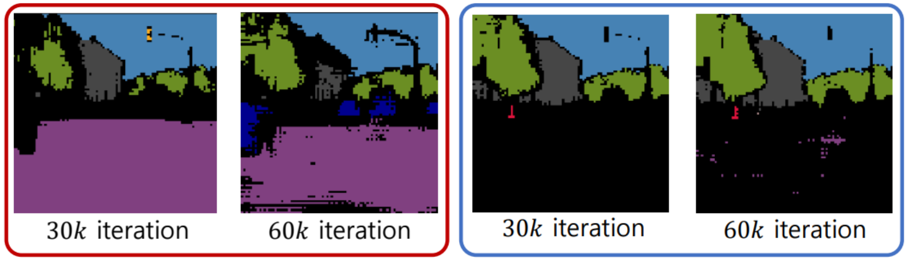
Domain adaptation by contrastive learning

- Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation
  - Experiments



(a) w/o cal. (b) w/ cal. (c) GT labels.

Pseudo labels	Density(%)	Accuracy(%)
Static [58]	20.1	98.5
Dyn. (w/o cal.)	22.2	98.6
Dyn. (w/ cal.)	34.3	98.6
Hybrid	42.3	98.8



(a) Using ρ_{S}. (b) Using μ_{S}.

$$y_{\mathrm{D}}(p,c) = \begin{cases} 1, & \text{if } s(f_{\mathrm{T}}(p)), \rho_{\mathrm{S}\to\mathrm{T}}(c)) > \mathcal{T} \text{and } c = c' \\ 0, & \text{otherwise} \end{cases}$$

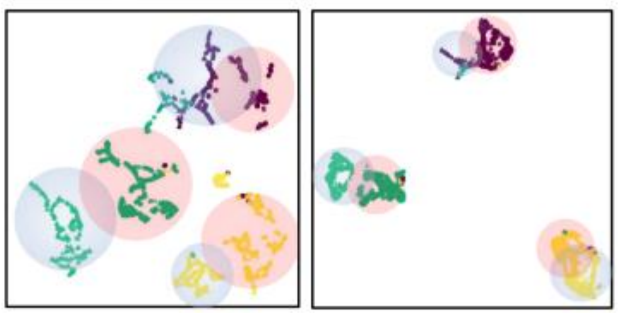
서강대학교 Sogang University

VDS LAB

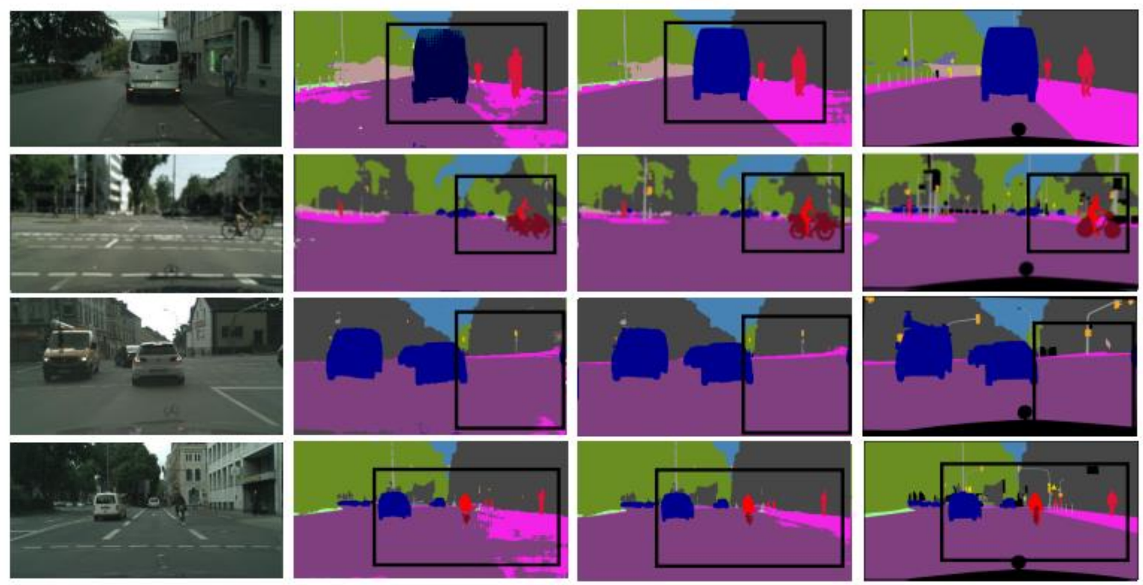
Domain adaptation by contrastive learning

- Bi-directional Contrastive Learning for Domain Adaptive Semantic Segmentation
  - Experiments

$\mathcal{L}_{\text{base}}$	$\mathcal{L}_{FC}$	$\mathcal{L}_{BC}$	$+y_{\text{D}}$ (w/o cal.)	$+y_{\text{D}}$ (w/ cal.)	Source dataset GTA5	Source dataset SYNTHIA
✓					49.5	45.1
✓	✓				51.2	48.8
✓	✓	✓			53.5	51.3
✓	✓	✓	✓		55.3	53.5
✓	✓	✓		✓	57.1	55.6



(a) Baseline. (b) Ours.



(a) Target images. (b) Our baseline. (c) Our model. (d) GT labels.

Conclusion

1	**SemiSegContrast** (DeepLab v3+ with ResNet-50 backbone, MSCOCO pretrained)	64.9%	×	Semi-Supervised Semantic Segmentation with Pixel-Level Contrastive Learning from a Class-wise Memory Bank			2021	ResNet
2	**SegSDE** (MTL decoder with ResNet101, ImageNet pretrained, unlabeled image sequences)	62.09%	✓	Three Ways to Improve Semantic Segmentation with Self-Supervised Depth Estimation			2020	
3	**ReCo** (DeepLab v3+ with ResNet-101 backbone, ImageNet pretrained)	60.28%	×	Bootstrapping Semantic Segmentation with Regional Contrast			2021	ResNet
4	**SemiSegContrast** (DeepLab v2 with ResNet-101 backbone, MSCOCO pretrained)	59.4%	×	Semi-Supervised Semantic Segmentation with Pixel-Level Contrastive Learning from a Class-wise Memory Bank			2021	ResNet
5	**GIST and RIST** (DeepLabv2 with ResNet101, MSCOCO pre-trained)	58.70%	×	The GIST and RIST of Iterative Self-Training for Semi-Supervised Segmentation			2021	

Rank	Model	mIoU↑	Training Data	Paper	Code	Result	Year	Tags
1	**MIC**	75.9	×	MIC: Masked Image Consistency for Context-Enhanced Domain Adaptation			2022	Transformer
2	**HRDA + PiPa**	75.6	×	PiPa: Pixel- and Patch-wise Self-supervised Learning for Domain Adaptative Semantic Segmentation			2022	Transformer
3	**CLUDA+HRDA**	74.4	×	CLUDA : Contrastive Learning in Unsupervised Domain Adaptation for Semantic Segmentation			2022	Transformer
4	**HRDA**	73.8	×	HRDA: Context-Aware High-Resolution Domain-Adaptive Semantic Segmentation			2022	Transformer
5	**DAFormer + PiPa**	71.7	×	PiPa: Pixel- and Patch-wise Self-supervised Learning for Domain Adaptative Semantic Segmentation			2022	Transformer
6	**SePiCo**	70.3	×	SePiCo: Semantic-Guided Pixel Contrast for Domain Adaptive Semantic Segmentation			2022	Transformer

서강대학교 SOGANG UNIVERSITY

VDS LAB