Bidirectional Learning for Domain Adaptation of Semantic Segmentation

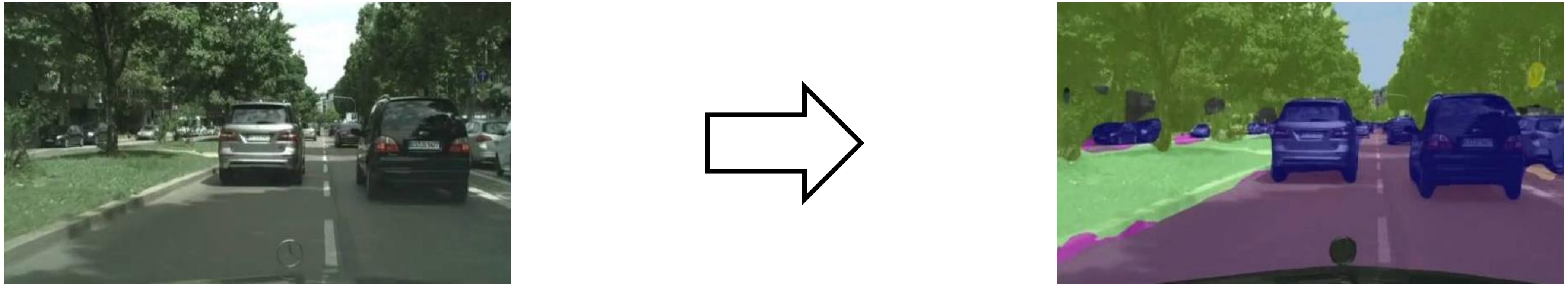
CVPR 2019

Yunsheng Li*, Lu Yuan`, Nuno Vasconcelos*

*UC San Deigo, `Microsoft

발표자 : 한양대학교 인공지능연구실 김병조

Problem – semantic segmentation



Problem – domain adaptation

Source

Target

Problem – domain adaptation of semantic segmentation

Source
a lot of dataset

Target
a few dataset

Problem – domain adaptation of semantic segmentation

Source
a lot of dataset

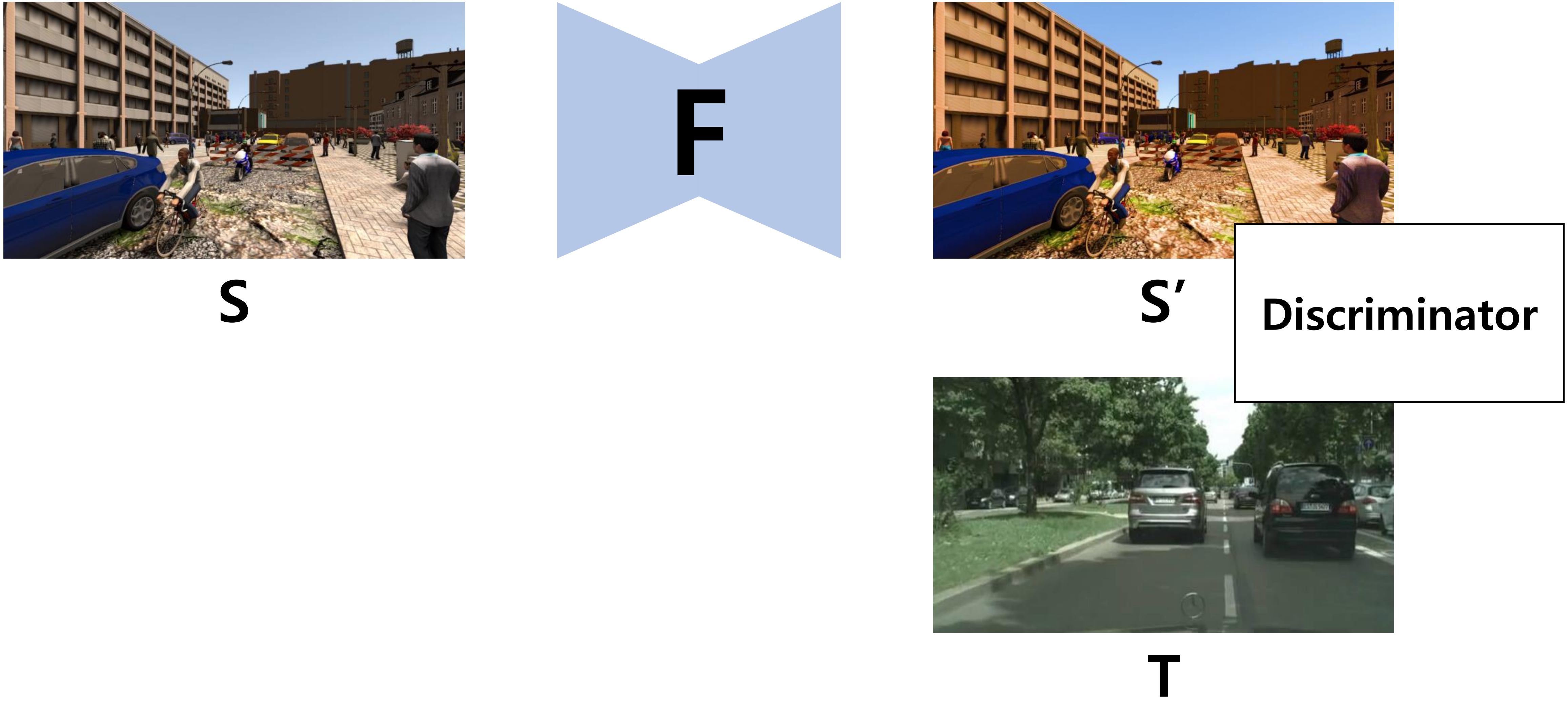
Target
a few dataset

Previous Work

Two Steps

1. Image-to-Image translation : Reduce domain gap
2. Learn Semantic Segmentation Model

Previous Work



Previous Work

S'

Previous Work

T

Previous Work

Two Steps

1. Image-to-Image translation : Reduce domain gap
2. Learn Semantic Segmentation Model

Problem

If F is fails, Nothing can be done..

Once F is learnt, it is fixed. -> No Feedback from M.

One-trial learning for M seems to just learn limited transferable knowledge.

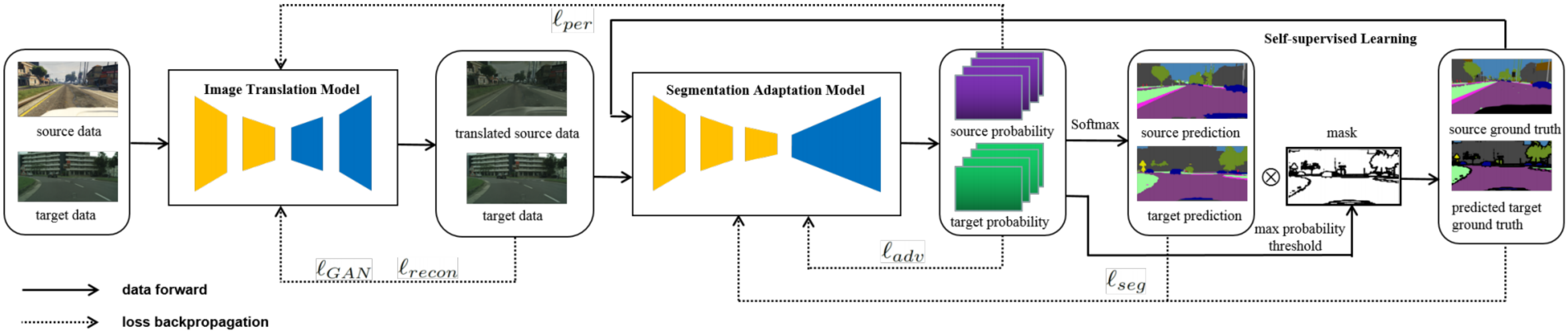
Method

Keyword(Novelty) in this paper

Bidirectional learning with perceptual loss

self-supervised learning

Method



ℓ_{per}
Self-supervised Learning
source data
target data
Image Translation Model
translated source data
target data
Segmentation Adaptation Model
source probability
target probability
Softmax
source prediction
target prediction
mask
max probability threshold
source ground truth
predicted target ground truth
ℓ_{GAN}
ℓ_{recon}
ℓ_{adv}
ℓ_{seg}
data forward
loss backpropagation

Method

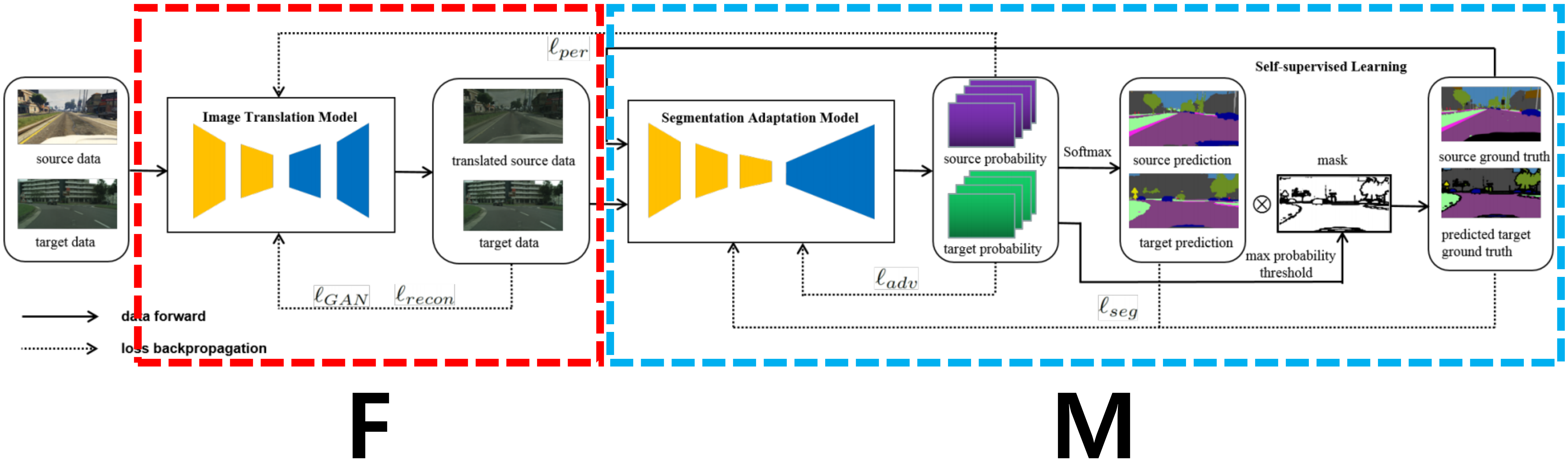
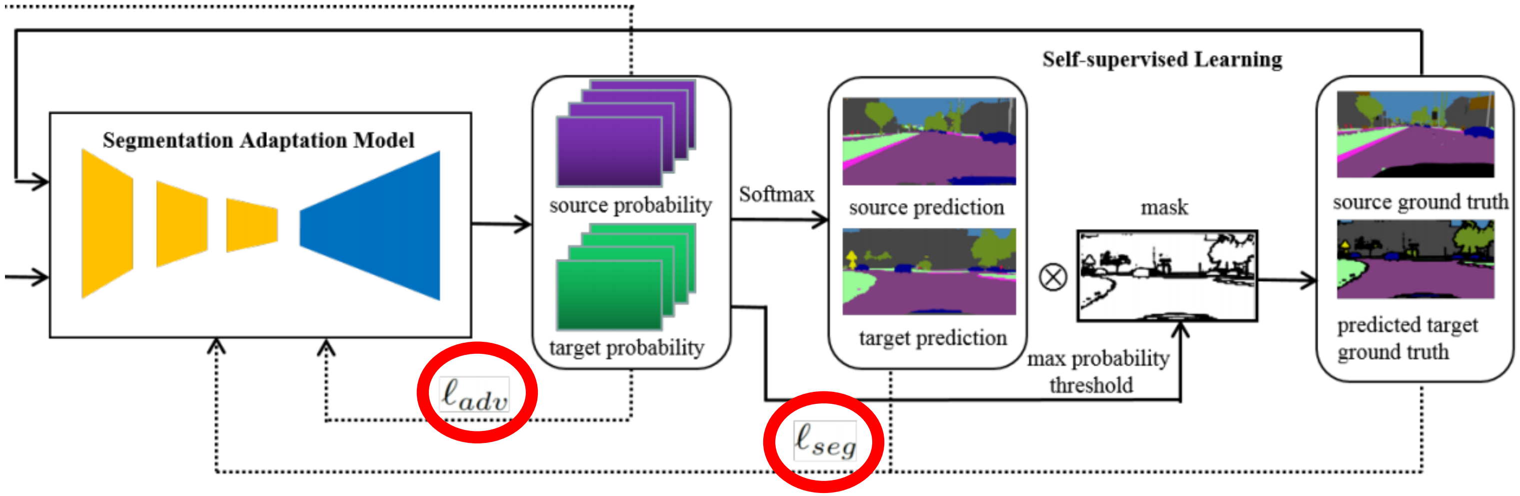


Image Translation Model
Segmentation Adaptation Model
Self-supervised Learning
source data
target data
translated source data
target data
source probability
target probability
Softmax
source prediction
target prediction
mask
max probability threshold
source ground truth
predicted target ground truth
ℓ_{per}
ℓ_{GAN}
ℓ_{recon}
ℓ_{adv}
ℓ_{seg}
data forward
loss backpropagation
F
M

Method

$$\ell_{\mathbf{M}} = \lambda_{adv}\ell_{adv}(\mathbf{M}(\mathcal{S}'), \mathbf{M}(\mathcal{T})) + \ell_{seg}(\mathbf{M}(\mathcal{S}'), Y_{\mathcal{S}})$$



Method

$$\ell_{\mathbf{M}} = \boxed{\lambda_{adv}\ell_{adv}(\mathbf{M}(\mathcal{S}'), \mathbf{M}(\mathcal{T}))} + \ell_{seg}(\mathbf{M}(\mathcal{S}'), Y_{\mathcal{S}})$$

target image

$$\ell_{adv}(\mathbf{M}(\mathcal{S}'), \mathbf{M}(\mathcal{T})) = \mathbb{E}_{I_{\mathcal{T}} \sim \mathcal{T}}[D_{\mathbf{M}}(\mathbf{M}(I_{\mathcal{T}}))] + \mathbb{E}_{I_{\mathcal{S}} \sim \mathcal{S}}[1 - D_{\mathbf{M}}(\mathbf{M}(I'_{\mathcal{S}}))]$$

translated source image
(given by F)

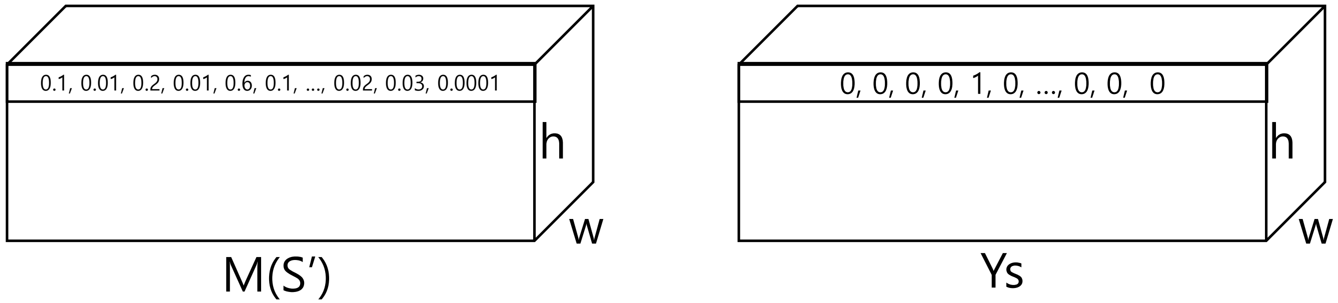
Method

$$\ell_{\mathbf{M}} = \lambda_{adv}\ell_{adv}(\mathbf{M}(\mathcal{S}'), \mathbf{M}(\mathcal{T})) + \boxed{\ell_{seg}(\mathbf{M}(\mathcal{S}'), Y_{\mathcal{S}})}$$

$$\ell_{seg}(\mathbf{M}(\mathcal{S}'), Y_{\mathcal{S}}) = -\frac{1}{HW}\sum_{H,W}\sum_{c=1}^{C}\mathbb{1}_{[c=y_{\mathcal{S}}^{hw}]}\log P_{\mathcal{S}}^{hwc}$$

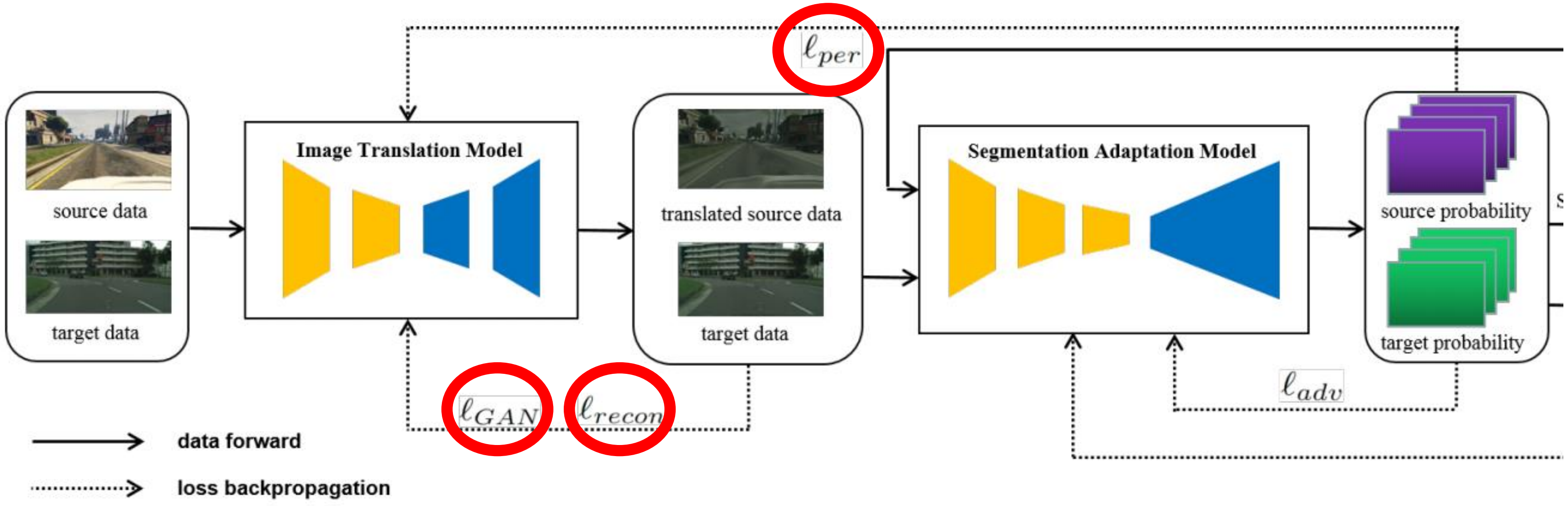
$$\mathbf{M}(I'_{\mathcal{S}})$$

cross-entropy loss



Method

$$\ell_{\mathbf{F}} = \lambda_{GAN}[\ell_{GAN}(\mathcal{S}', \mathcal{T}) + \ell_{GAN}(\mathcal{S}, \mathcal{T}')] + \lambda_{recon}[\ell_{recon}(\mathcal{S}, \mathbf{F}^{-1}(\mathcal{S}')) + \ell_{recon}(\mathcal{T}, \mathbf{F}(\mathcal{T}'))] + \ell_{per}(\mathbf{M}(\mathcal{S}), \mathbf{M}(\mathcal{S}')) + \ell_{per}(\mathbf{M}(\mathcal{T}), \mathbf{M}(\mathcal{T}')),$$



Method

$$\ell_{\mathbf{F}} = \boxed{\lambda_{GAN}[\ell_{GAN}(\mathcal{S}', \mathcal{T}) + \ell_{GAN}(\mathcal{S}, \mathcal{T}')]}$$
$$+ \lambda_{recon}[\ell_{recon}(\mathcal{S}, \mathbf{F}^{-1}(\mathcal{S}')) + \ell_{recon}(\mathcal{T}, \mathbf{F}(\mathcal{T}')]$$
$$+ \ell_{per}(\mathbf{M}(\mathcal{S}), \mathbf{M}(\mathcal{S}')) + \ell_{per}(\mathbf{M}(\mathcal{T}), \mathbf{M}(\mathcal{T}'),$$

$$\ell_{GAN}(\mathcal{S}', \mathcal{T}) = \mathbb{E}_{I_{\mathcal{T}} \sim \mathcal{T}}[D_{\mathbf{F}}(I_{\mathcal{T}})] + \mathbb{E}_{I_{\mathcal{S}} \sim \mathcal{S}}[1 - D_{\mathbf{F}}((I'_{\mathcal{S}}))]$$

S′ = F(S) F() : S domain에서 T domain으로

T′ = $\mathbf{F}^{-1}$(T) $\mathbf{F}^{-1}$() : T domain에서 S domain으로

Method

$$\ell_{\mathbf{F}} = \lambda_{GAN}[\ell_{GAN}(\mathcal{S}', \mathcal{T}) + \ell_{GAN}(\mathcal{S}, \mathcal{T}')]$$
$$+ \boxed{\lambda_{recon}[\ell_{recon}(\mathcal{S}, \mathbf{F}^{-1}(\mathcal{S}')) + \ell_{recon}(\mathcal{T}, \mathbf{F}(\mathcal{T}'))]}$$
$$+ \ell_{per}(\mathbf{M}(\mathcal{S}), \mathbf{M}(\mathcal{S}')) + \ell_{per}(\mathbf{M}(\mathcal{T}), \mathbf{M}(\mathcal{T}')),$$

$$\ell_{recon}(\mathcal{S}, \mathbf{F}^{-1}(\mathcal{S}')) = \mathbb{E}_{I_{\mathcal{S}} \sim \mathcal{S}}[||\mathbf{F}^{-1}((I'_{\mathcal{S}})) - I_{\mathcal{S}}||_1]$$

$\mathbf{F}^{-1}$(F(S))

S′ = F(S) F() : S domain에서 T domain으로

T′ = $\mathbf{F}^{-1}$(T) $\mathbf{F}^{-1}$() : T domain에서 S domain으로

Method

$$\ell_{\mathbf{F}} = \lambda_{GAN}[\ell_{GAN}(\mathcal{S}', \mathcal{T}) + \ell_{GAN}(\mathcal{S}, \mathcal{T}')] + \lambda_{recon}[\ell_{recon}(\mathcal{S}, \mathbf{F}^{-1}(\mathcal{S}')) + \ell_{recon}(\mathcal{T}, \mathbf{F}(\mathcal{T}')] + \boxed{\ell_{per}(\mathbf{M}(\mathcal{S}), \mathbf{M}(\mathcal{S}')) + \ell_{per}(\mathbf{M}(\mathcal{T}), \mathbf{M}(\mathcal{T}'),}$$

$$\ell_{per}(\mathbf{M}(\mathcal{S}), \mathbf{M}(\mathcal{S}')) = \lambda_{per}\mathbb{E}_{I_{\mathcal{S}}\sim\mathcal{S}}||\mathbf{M}(I_{\mathcal{S}}) - \mathbf{M}((I'_{\mathcal{S}}))||_1 + \lambda_{per\_recon}\mathbb{E}_{I_{\mathcal{S}}\sim\mathcal{S}}[||\mathbf{M}(\mathbf{F}^{-1}((I'_{\mathcal{S}}))) - \mathbf{M}(I_{\mathcal{S}})||_1]$$

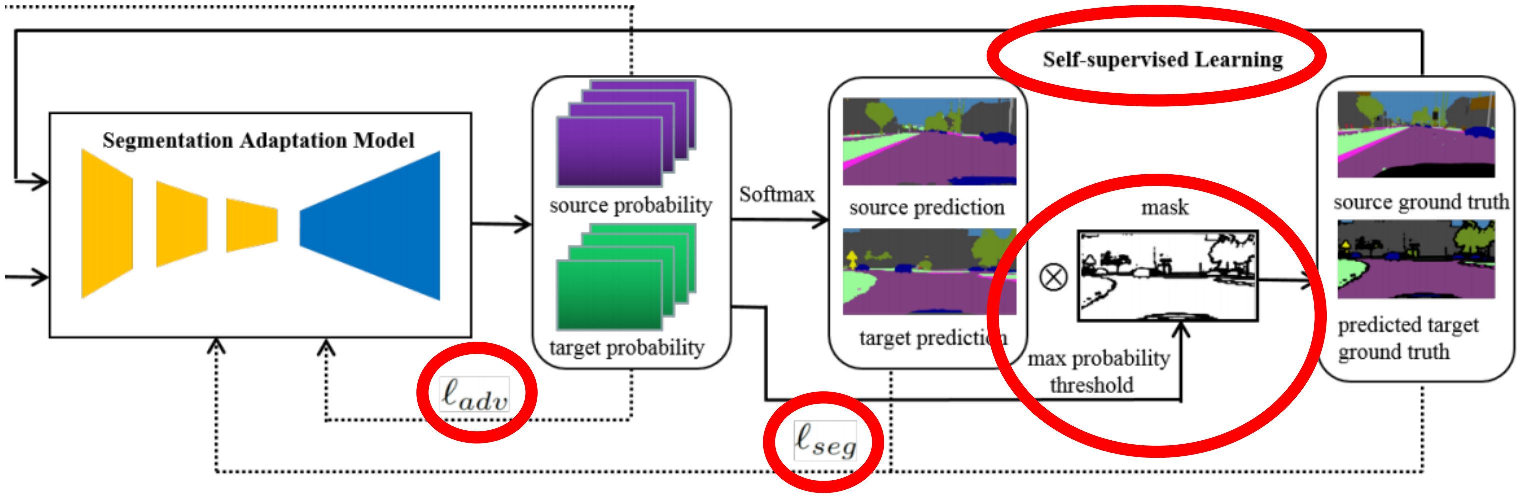
to keep the semantic consistency

between S and S'

between S, $\mathbf{F}^{-1}$(T)

Method

$$\ell_{\mathbf{M}} = \lambda_{adv}\ell_{adv}(\mathbf{M}(\mathcal{S}'), \mathbf{M}(\mathcal{T})) + \ell_{seg}(\mathbf{M}(\mathcal{S}'), Y_{\mathcal{S}}) + \ell_{seg}(\mathbf{M}(\mathcal{T}_{ssl}), \widehat{Y}_{\mathcal{T}})$$



Self-Supervised Learning (SSL)

- labeled data로 모델 학습
- unlabeled data 예측
- 확률 값이 높은 데이터들을 pseudo-labeled data로 사용
- 다시 학습
- …

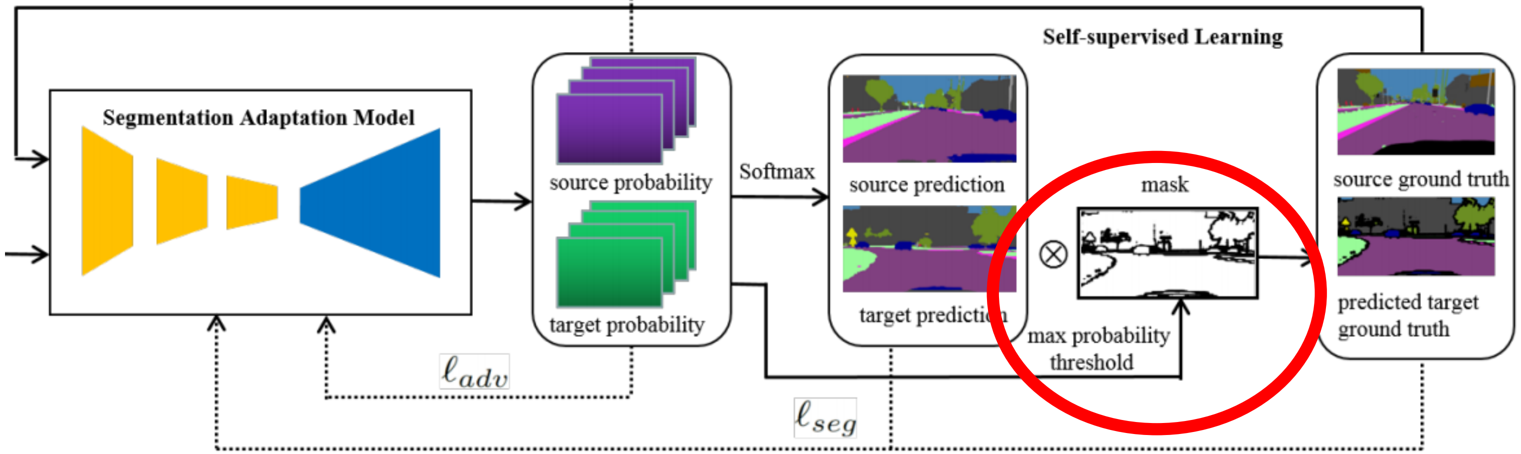
Method

$$\ell_{\mathbf{M}} = \lambda_{adv}\ell_{adv}(\mathbf{M}(\mathcal{S}'), \mathbf{M}(\mathcal{T})) + \ell_{seg}(\mathbf{M}(\mathcal{S}'), Y_{\mathcal{S}}) + \boxed{\ell_{seg}(\mathbf{M}(\mathcal{T}_{ssl}), \widehat{Y}_{\mathcal{T}})}$$

$$m_{\mathcal{T}} = \mathbb{1}_{[\mathrm{argmax}\,\mathbf{M}(I_{\mathcal{T}}) > \mathrm{threshold}]}$$

$$\ell_{seg}(\mathbf{M}(\mathcal{T}_{ssl}), \widehat{Y}_{\mathcal{T}}) = -\frac{1}{HW}\sum_{H,W} m_{\mathcal{T}}^{hw} \sum_{c=1}^{C} \mathbb{1}_{[c=y_{\mathcal{T}}^{hw}]} \log P_{\mathcal{T}}^{hwc}$$

to help promote the segmentation adaptation model M



Training Algorithm

Algorithm 1 Training process of our network

Input: $(\mathcal{S}, Y_{\mathcal{S}}), (\mathcal{T}, \mathcal{T}_{ssl} = \emptyset)$, $\mathbf{M^{(0)}}$
Output: $\mathrm{M}_N^{(K)}(\mathbf{F}^{(K)})$
  for $k \leftarrow 1$ to K **do** (Bidirectional Learning)
    train $\mathbf{F}^{(k)}$ with Equation 2
    train $\mathrm{M}_0^{(k)}$ with Equation 1
    for $i \leftarrow 1$ to N **do** (SSL)
      update $\mathcal{T}_{ssl}$ with $\mathrm{M}_{i-1}^{(k)}$
      train $\mathrm{M}_i^{(k)}$ again with Equation 3
    end for
  end for

outer loop

$$\mathbf{M}_i^{(k)}(\mathbf{F}^{(k)})$$

inner loop (SSL)

Bidirectional Learning without SSL

GTA5 $\rightarrow$ Cityscapes	
model	mIoU
$\mathbf{M}^{(0)}$	33.6
$\mathbf{M}^{(1)}$	40.9
$\mathbf{M}^{(0)}(\mathbf{F}^{(1)})$	41.1
$\mathbf{M}_0^{(1)}(\mathbf{F}^{(1)})$	42.7
$\mathbf{M}_0^{(2)}(\mathbf{F}^{(2)})$	43.3

Bidirectional Learning with SSL

Table 3: Influence of threshold

GTA5 → Cityscapes		
model	threshold	mIoU
$\mathbf{M}_1^{(1)}(\mathbf{F}^{(1)})$	0.95	45.7
$\mathbf{M}_1^{(1)}(\mathbf{F}^{(1)})$	0.9	46.8
$\mathbf{M}_1^{(1)}(\mathbf{F}^{(1)})$	0.8	46.4
$\mathbf{M}_1^{(1)}(\mathbf{F}^{(1)})$	0.7	45.9
$\mathbf{M}_1^{(1)}(\mathbf{F}^{(1)})$	–	44.9

Table 4: Influence of N

GTA5 → Cityscapes		
model	pixel ratio	mIoU
$\mathbf{M}_0^{(1)}$	66%	40.9
$\mathbf{M}_0^{(1)}(\mathbf{F}^{(1)})$	69%	42.7
$\mathbf{M}_1^{(1)}(\mathbf{F}^{(1)})$	79%	46.8
$\mathbf{M}_2^{(1)}(\mathbf{F}^{(1)})$	81%	47.2
$\mathbf{M}_3^{(1)}(\mathbf{F}^{(1)})$	81%	47.1

Bidirectional Learning with SSL

GTA5 → Cityscapes																					
		road	sidewalk	building	wall	fence	pole	t-light	t-sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorbike	bicycle	mIoU
	$\mathbf{M}^{(0)}$	69.0	12.7	69.5	9.9	19.5	22.8	31.7	15.3	73.9	11.3	67.2	54.7	23.9	53.4	29.7	4.6	11.6	26.1	32.5	33.6
$k = 1$	$\mathbf{M}_0^{(1)}(\mathbf{F}^{(1)})$	89.1	42.0	82.0	24.3	15.1	27.4	35.7	24.6	81.1	32.4	78.0	57.6	28.7	76.0	26.5	36.0	4.0	25.7	24.9	42.7
	$\mathbf{M}_1^{(1)}(\mathbf{F}^{(1)})$	91.2	47.8	84.0	34.8	28.9	31.7	37.7	36.0	84.0	40.4	76.6	57.9	25.3	80.4	31.2	41.7	2.8	27.2	32.4	46.8
	$\mathbf{M}_2^{(1)}(\mathbf{F}^{(1)})$	91.4	47.9	84.2	32.4	26.0	31.8	37.3	33.0	83.3	39.2	79.2	57.7	25.6	81.3	36.3	39.7	2.6	31.3	33.5	47.2
$k = 2$	$\mathbf{M}_0^{(2)}(\mathbf{F}^{(2)})$	88.2	41.3	83.2	28.8	21.9	31.7	35.2	28.2	83.0	26.2	83.2	57.6	27.0	77.1	27.5	34.6	2.5	28.3	36.1	44.3
	$\mathbf{M}_1^{(2)}(\mathbf{F}^{(2)})$	91.2	46.1	83.9	31.6	20.6	29.9	36.4	31.9	85.0	39.7	84.7	57.5	29.6	83.1	38.8	46.9	2.5	27.5	38.2	47.6
	$\mathbf{M}_2^{(2)}(\mathbf{F}^{(2)})$	91.0	44.7	84.2	34.6	27.6	30.2	36.0	36.0	85.0	43.6	83.0	58.6	31.6	83.3	35.3	49.7	3.3	28.8	35.6	48.5

Table 5: Comparison results from GTA5 to Cityscapes

GTA5 → Cityscapes																					
Oracle	Method	road	sidewalk	building	wall	fence	pole	t-light	t-sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorbike	bicycle	mIoU
ResNet101[11] 65.1	Cycada[12]	86.7	35.6	80.1	19.8	17.5	**38.0**	**39.9**	**41.5**	82.7	27.9	73.6	**64.9**	19	65.0	12.0	28.6	4.5	31.1	**42.0**	42.7
	AdaptSegNet[33]	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
	DCAN[36]	85.0	30.8	81.3	25.8	21.2	22.2	25.4	26.6	83.4	36.7	76.2	58.9	24.9	80.7	29.5	42.9	2.50	26.9	11.6	41.7
	CLAN[19]	87.0	27.1	79.6	27.3	23.3	28.3	35.5	24.2	83.6	27.4	74.2	58.6	28.0	76.2	33.1	36.7	6.7	**31.9**	31.4	43.2
	Ours	**91.0**	**44.7**	**84.2**	**34.6**	**27.6**	30.2	36.0	36.0	**85.0**	**43.6**	**83.0**	58.6	**31.6**	**83.3**	**35.3**	**49.7**	3.3	28.8	35.6	**48.5**
VGG16[32] 60.3	Curriculum[37]	74.9	22.0	71.7	6.0	11.9	8.4	16.3	11.1	75.7	13.3	66.5	38.0	9.3	55.2	18.8	18.9	0.0	16.8	16.6	28.9
	CBST[39]	66.7	26.8	73.7	14.8	9.5	**28.3**	25.9	10.1	75.5	15.7	51.6	47.2	6.2	71.9	3.7	2.2	**5.4**	18.9	**32.4**	30.9
	Cycada[12]	85.2	37.2	76.5	21.8	15.0	23.8	22.9	**21.5**	80.5	31.3	60.7	50.5	9.0	76.9	17.1	28.2	4.5	9.8	0	35.4
	DCAN[36]	82.3	26.7	77.4	23.7	**20.5**	20.4	**30.3**	15.9	80.9	25.4	69.5	52.6	11.1	79.6	24.9	21.2	1.30	17.0	6.70	36.2
	CLAN[19]	88.0	30.6	79.2	23.4	**20.5**	26.1	23.0	14.8	81.6	34.5	72.0	45.8	7.9	80.5	**26.6**	**29.9**	0.0	10.7	0.0	36.6
	Ours	**89.2**	**40.9**	**81.2**	**29.1**	19.2	14.2	29.0	19.6	**83.7**	**35.9**	**80.7**	**54.7**	**23.3**	**82.7**	25.8	28.0	2.3	**25.7**	19.9	**41.3**

Table 6: Comparison results from SYNTHIA to Cityscapes

SYNTHIA → Cityscapes																		
Oracle	Method	road	sidewalk	building	wall	fence	pole	t-light	t-sign	vegetation	sky	person	rider	car	bus	motorbike	bicycle	mIoU
ResNet101[11] 71.7	AdaptSegNet[33]	79.2	37.2	78.8	-	-	-	9.9	10.5	78.2	80.5	53.5	19.6	67.0	29.5	21.6	31.3	45.9
	CLAN[19]	81.3	37.0	80.1	-	-	-	**16.1**	**13.7**	78.2	**81.5**	53.4	21.2	73.0	32.9	22.6	30.7	47.8
	Ours	**86.0**	**46.7**	**80.3**	-	-	-	14.1	11.6	**79.2**	81.3	**54.1**	**27.9**	**73.7**	**42.2**	**25.7**	**45.3**	**51.4**
VGG16[32] 59.5	FCN wild[13]	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
	Curriculum[37]	65.2	26.1	**74.9**	0.1	0.5	10.7	3.5	3.0	76.1	70.6	47.1	8.2	43.2	20.7	0.7	13.1	29.0
	CBST[39]	69.6	28.7	69.5	**12.1**	0.1	**25.4**	**11.9**	13.6	**82.0**	**81.9**	49.1	14.5	66.0	6.6	3.7	32.4	35.4
	DCAN[36]	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
	Ours	**72.0**	30.3	74.5	0.1	0.3	24.6	10.2	**25.2**	80.5	80.0	**54.7**	**23.2**	**72.7**	**24.0**	7.5	**44.9**	**39.0**

QnA