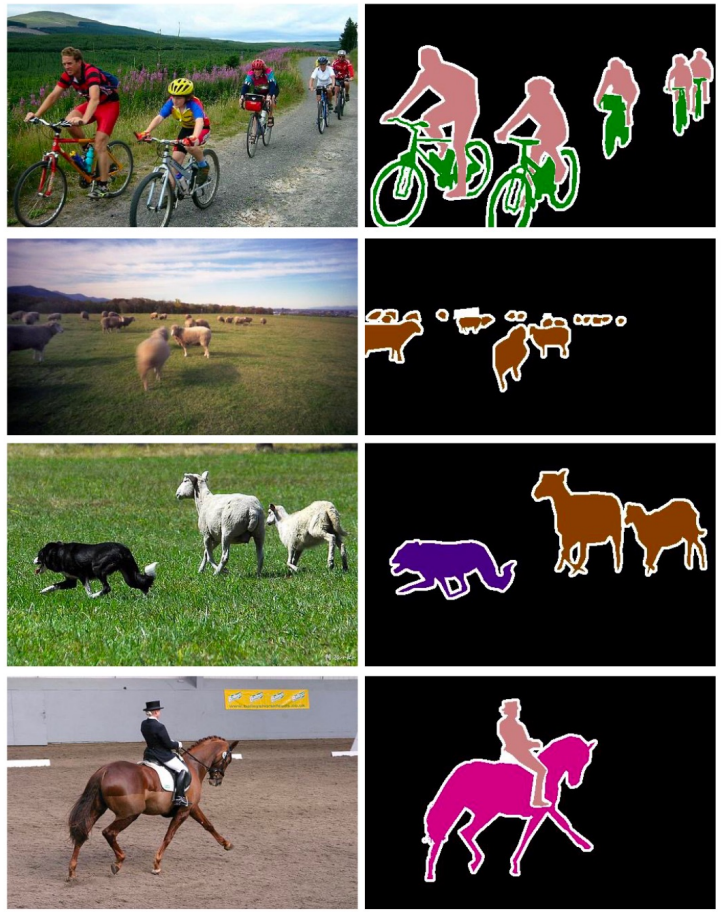
Semi-Supervised Semantic Segmentation

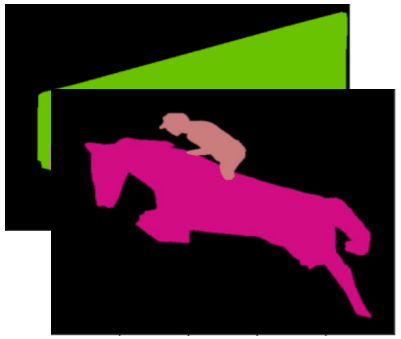
background

- Data

Fully-Supervised
labeled images



Semi-Supervised



labeled images

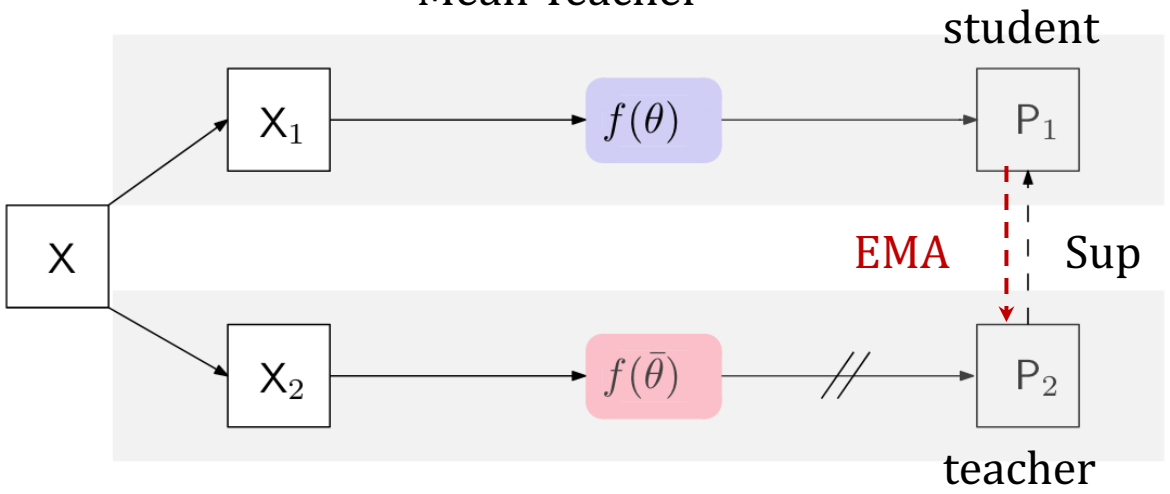
+

unlabeled images

Semi-Supervised Segmentation

- Mainstream Methods

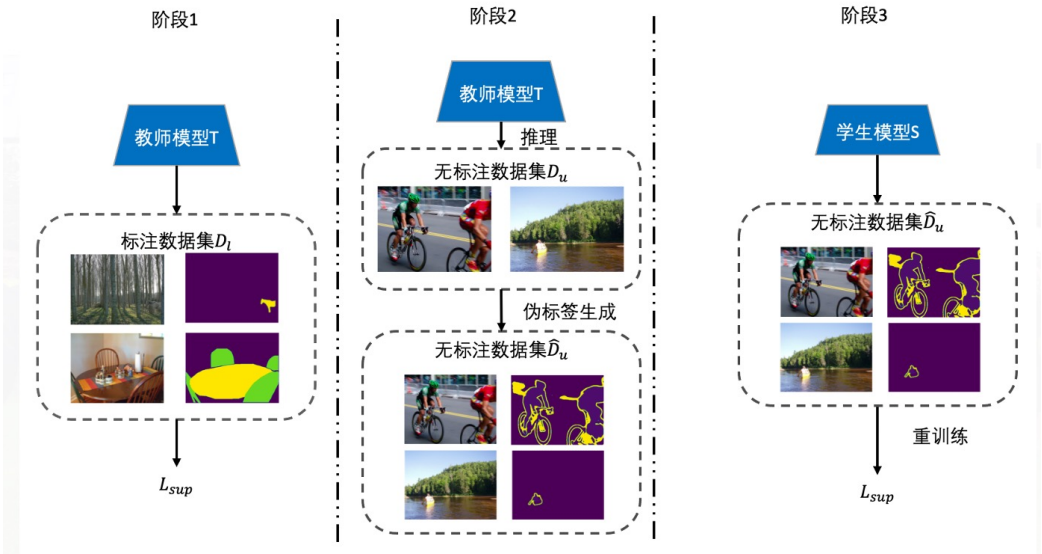
Mean-Teacher



features:

1. Joint training.
2. Exponential Moving Average to update teacher.

Self Training



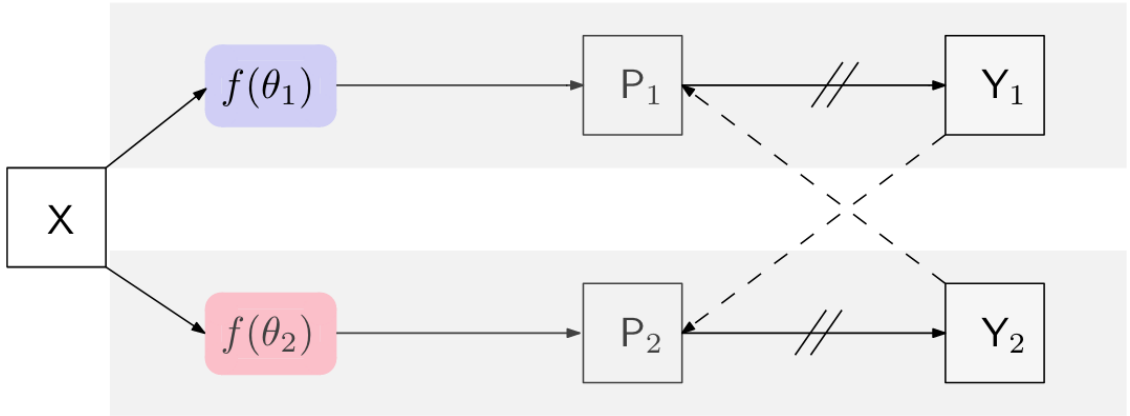
features:

1. serial training
2. multi-stage training procedure
3. Intermediate results

Semi-Supervised Segmentation

- Mean-Teacher

Classic CPS:



Features:

1. Cross Pseudo-label supervision.
2. use CutMix SDA to further boost performance.

representative works:

Mean-Teacher [NeurIPS 2017]

CCT [CVPR 2020]
Consistency method

CutMix-Seg [BMCV 2020]
CutMix SDA in SSSS

GCT [ECCV 2020]

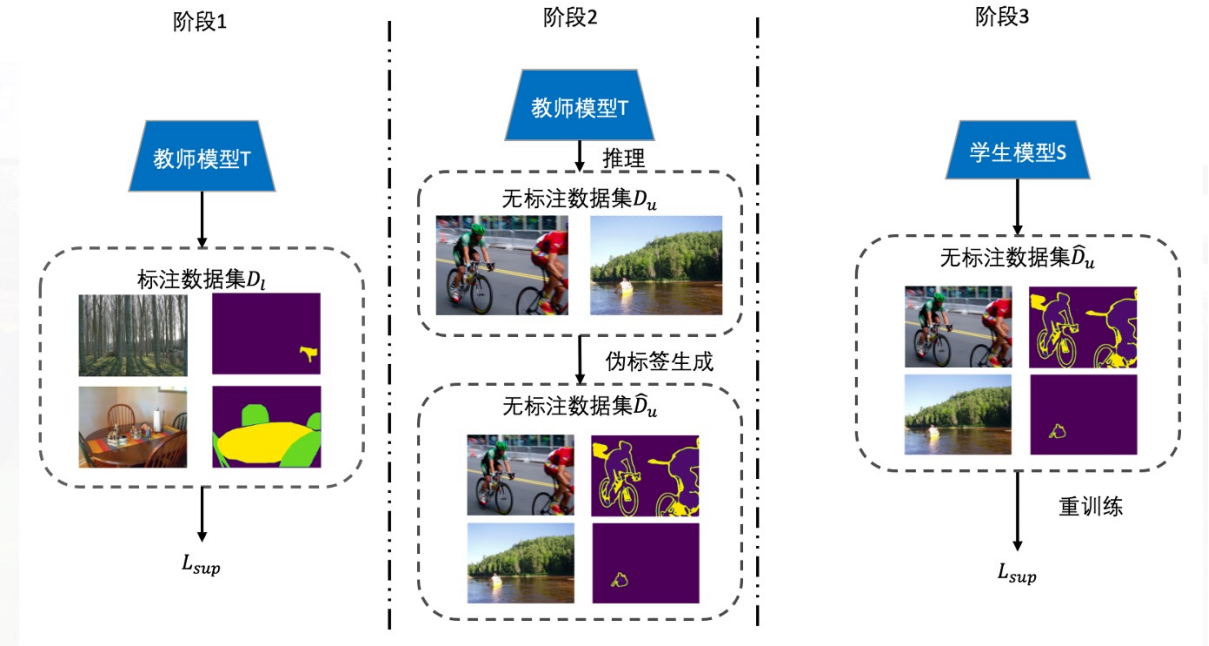
CPS [CVPR 2021]

AEL [NeurIPS 2021]

U2PL [CVPR 2022]

Semi-Supervised Segmentation

- Self-Training



representative works:

Naïve Student [ECCV 2020]

Noisy Student [CVPR 2020]

DSBN [ICCV 2021]

ST++ [CVPR 2022]

research status:

1. few attention
2. Limited improvement
3. Weak performance

Compared with MT kinds.

Semi-Supervised Semantic Segmentation Using Unreliable Pseudo-Labels

Yuchao Wang[1*] Haochen Wang[1*] Yujun Shen[2] Jingjing Fei[3]
Wei Li[3] Guoqiang Jin[3] Liwei Wu[3] Rui Zhao[3] Xinyi Le[1]

[1]Shanghai Jiao Tong University [2]The Chinese University of Hong Kong [3]SenseTime Research

U2PL

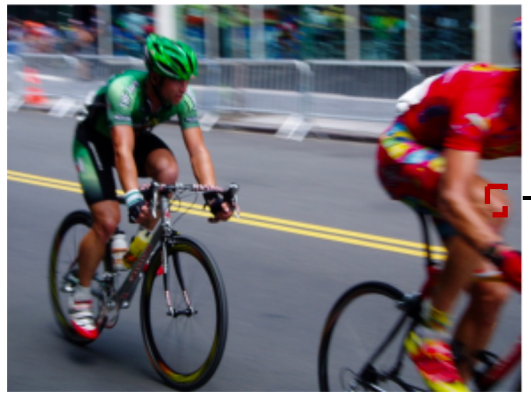
- motivation

Previously works select the highly confident predictions as the pseudo ground-truth.

Weakness: most pixels may be left unused due to unreliability.

-> make sufficient use of unlabeled images

Intuition



Person	0.60
Bicycle	0.18
Motorbike	0.14
Car	0.03
Sofa	0.03
Airplane	0.02

1. An unreliable prediction may get confused among the top classes, however, it should be confident about the pixel not belonging to the remaining classes.
2. Such a pixel can be convincingly treated as a negative sample to those most unlikely categories. -> Contrastive Learning

U2PL

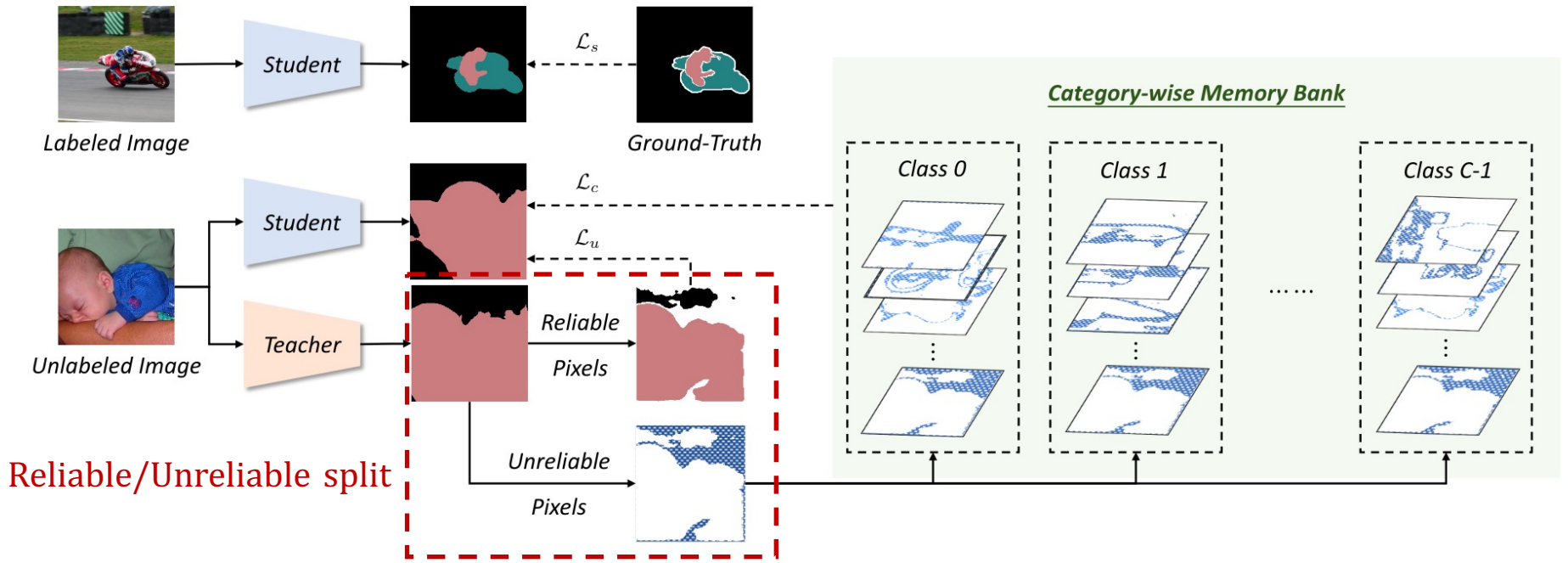
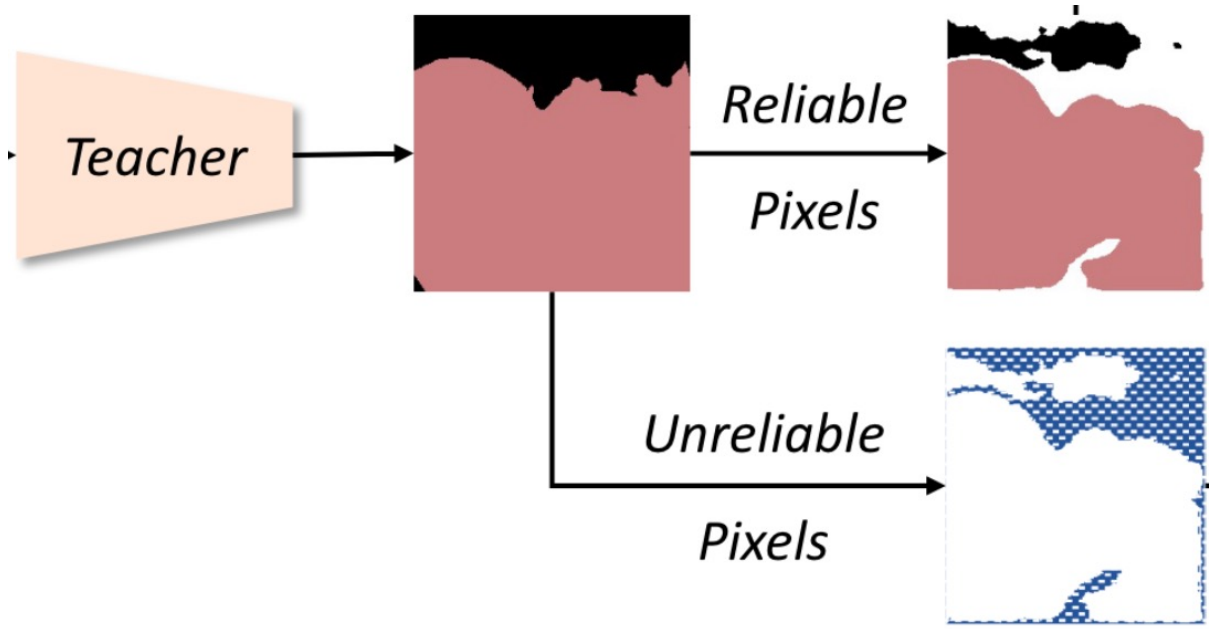


Figure 3. **An overview of our proposed U^{2}PL method.** U^{2}PL contains a student network and a teacher network, where the teacher is momentum-updated with the student. Labeled data is directly fed into the student network for supervised training. Given an unlabeled image, we first use the teacher model to make a prediction, and then separate the pixels into reliable ones and unreliable ones based on their entropy. Such a process is formulated as Eq. (6). The reliable predictions are directly used as the pseudo-labels to advise the student, while each unreliable prediction is pushed into a category-wise memory bank. Pixels in each memory bank are regarded as the negative samples to the corresponding class, which is formulated as Eq. (4).

U2PL



DPA -- Dynamic Partition Adjustment

Reliable criterion quantify

$$\mathcal{H}(\mathbf{p}_{ij}) = -\sum_{c=0}^{C-1} p_{ij}(c) \log p_{ij}(c),$$

$p_{ij} \in R^C$, softmax probabilities generated by the segmentation head of the teacher model for the *i*th unlabeled image at pixel *j*

$$\hat{y}_{ij}^{u} = \begin{cases} \arg\max_c p_{ij}(c), & \text{if } \mathcal{H}(\mathbf{p}_{ij}) < \gamma_t, \\ \text{ignore}, & \text{otherwise}, \end{cases}$$

γ_t=`np.percentile(H.flatten(),100*(1-`α_t`))`

$$\alpha_t = \alpha_0 \cdot \left(1 - \frac{t}{\text{total epoch}}\right),$$

Anchor, Positive and negative

PRT -- Probability Rank Threshold

Anchor pixels (representation)
for class c

$$\mathcal{A}_c^l = \{\mathbf{z}_{ij} \mid y_{ij} = c, p_{ij}(c) > \delta_p\},$$

$$\mathcal{A}_c^u = \{\mathbf{z}_{ij} \mid \hat{y}_{ij} = c, p_{ij}(c) > \delta_p\}.$$

$$\mathcal{A}_c = \mathcal{A}_c^l \cup \mathcal{A}_c^u$$

δ_p = 0.3

Positive samples

$$\mathbf{z}_c^+ = \frac{1}{|\mathcal{A}_c|} \sum_{\mathbf{z}_c \in \mathcal{A}_c} \mathbf{z}_c$$

Negative samples

$$n_{ij}(c) = \begin{cases} n_{ij}^l(c), & \text{if image } i \text{ is labeled,} \\ n_{ij}^u(c), & \text{otherwise,} \end{cases}$$

Labeled images:
(a) not belongs to class c;
(b) similar to class c.

$$\mathcal{O}_{ij} = \texttt{argsort}(\mathbf{p}_{ij})$$

$$n_{ij}^l(c) = \mathbb{1}\left[y_{ij} \neq c\right] \cdot \mathbb{1}\left[0 \leq \mathcal{O}_{ij}(c) < r_l\right],$$

Unlabeled images:
(a) be unreliable;
(b) probably not belongs to class c;
(c) not belongs to most unlikely classes.

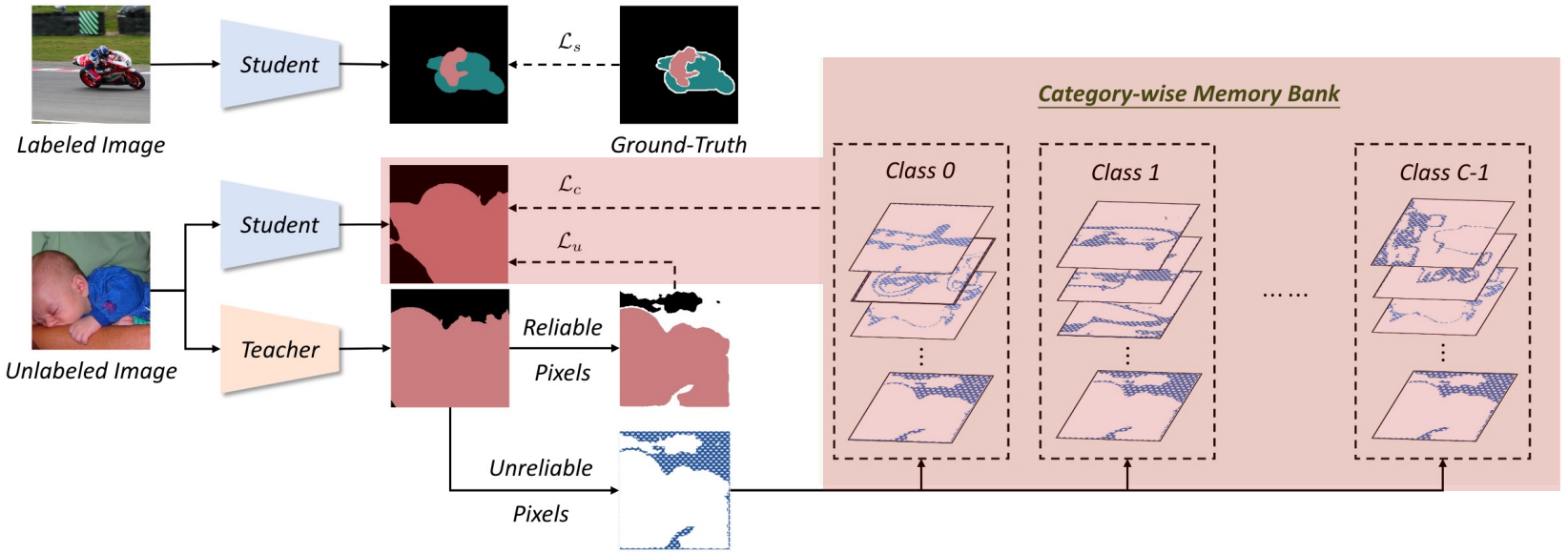
$$n_{ij}^u(c) = \mathbb{1}\left[\mathcal{H}(\mathbf{p}_{ij}) > \gamma_t\right] \cdot \mathbb{1}\left[r_l \leq \mathcal{O}_{ij}(c) < r_h\right]$$

$$\mathcal{N}_c = \{\mathbf{z}_{ij} \mid n_{ij}(c) = 1\}$$

U2PL

Algorithm 1: Using Unreliable Pseudo-Labels

Initialize $\mathcal{L} \leftarrow 0$;
Sample labeled images $\mathcal{B}_l$ and unlabeled images $\mathcal{B}_u$;
for $\mathbf{x}_i \in \mathcal{B}_l \cup \mathcal{B}_u$ **do**
  Get probabilities: $\mathbf{p}_i \leftarrow f \circ h(\mathbf{x}_i; \theta_t)$;
  Get representations: $\mathbf{z}_i \leftarrow g \circ h(\mathbf{x}_i; \theta_s)$;
  for $c \leftarrow 0$ **to** $C-1$ **do**
    Get anchors $\mathcal{A}_c$ based on Eq. (11);
    Sample M anchors: $\mathcal{B}_A \leftarrow \texttt{sample}\ (\mathcal{A}_c)$;
    Get negatives $\mathcal{N}_c$ based on Eq. (16);
    Push $\mathcal{N}_c$ into memory bank $\mathcal{Q}_c$;
    Pop oldest ones out of $\mathcal{Q}_c$ if necessary;
    Sample N negatives: $\mathcal{B}_N \leftarrow \texttt{sample}\ (\mathcal{Q}_c)$;
    Get $\mathbf{z}^+$ based on Eq. (12);
    $\mathcal{L} \leftarrow \mathcal{L} + \ell(\mathcal{B}_A, \mathcal{B}_N, \mathbf{z}^+)$ based on Eq. (4);
  end
end
Output: contrastive loss $\mathcal{L}_c \leftarrow \frac{1}{|\mathcal{B}| \times C}\mathcal{L}$



InfoNCE loss

$$\mathcal{L}_c = -\frac{1}{C \times M} \sum_{c=0}^{C-1} \sum_{i=1}^{M} \log \left[\frac{e^{\langle \mathbf{z}_{ci}, \mathbf{z}_{ci}^+ \rangle / \tau}}{e^{\langle \mathbf{z}_{ci}, \mathbf{z}_{ci}^+ \rangle / \tau} + \sum_{j=1}^{N} e^{\langle \mathbf{z}_{ci}, \mathbf{z}_{cij}^- \rangle / \tau}} \right],$$

Experiments

Table 1. Comparison with state-of-the-art methods on *classic* **PASCAL VOC 2012** `val` set under different partition protocols. The labeled images are selected from the original VOC `train` set, which consists of $1,464$ samples in total. The fractions denote the percentage of labeled data used for training, followed by the actual number of images. All the images from SBD [18] are regarded as unlabeled data. "SupOnly" stands for supervised training without using any unlabeled data. † means we reproduce the approach.

Method	1/16 (92)	1/8 (183)	1/4 (366)	1/2 (732)	Full (1464)
SupOnly	45.77	54.92	65.88	71.69	72.50
MT† [38]	51.72	58.93	63.86	69.51	70.96
CutMix† [15]	52.16	63.47	69.46	73.73	76.54
PseudoSeg [50]	57.60	65.50	69.14	72.41	73.23
PC^2Seg [48]	57.00	66.28	69.78	73.05	74.15
U^2PL (w/ CutMix)	**67.98** (+15.82)	**69.15** (+5.68)	**73.66** (+4.20)	**76.16** (+2.43)	**79.49** (+2.95)

Experiments

Results on Pascal VOC

Table 2. Comparison with state-of-the-art methods on *blender* **PASCAL VOC 2012** `val` set under different partition protocols. All labeled images are selected from the augmented VOC `train` set, which consists of $10,582$ samples in total. "SupOnly" stands for supervised training without using any unlabeled data. † means we reproduce the approach.

Method	1/16 (662)	1/8 (1323)	1/4 (2646)	1/2 (5291)
SupOnly	67.87	71.55	75.80	77.13
MT† [38]	70.51	71.53	73.02	76.58
CutMix† [15]	71.66	75.51	77.33	78.21
CCT [33]	71.86	73.68	76.51	77.40
GCT [22]	70.90	73.29	76.66	77.98
CPS [9]	74.48	76.44	77.68	78.64
AEL [21]	77.20	77.57	78.06	80.29
U^{2}PL (w/ CutMix)	**77.21** (+5.55)	**79.01** (+3.50)	**79.30** (+1.97)	**80.50** (+2.29)

Results on Cityscpaes

Table 3. Comparison with state-of-the-art methods on **Cityscapes** `val` set under different partition protocols. All labeled images are selected from the Cityscapes `train` set, which consists of $2,975$ samples in total. "SupOnly" stands for supervised training without using any unlabeled data. † means we reproduce the approach.

Method	1/16 (186)	1/8 (372)	1/4 (744)	1/2 (1488)
SupOnly	65.74	72.53	74.43	77.83
MT† [38]	69.03	72.06	74.20	78.15
CutMix† [15]	67.06	71.83	76.36	78.25
CCT [33]	69.32	74.12	75.99	78.10
GCT [22]	66.75	72.66	76.11	78.34
CPS† [9]	69.78	74.31	74.58	76.81
AEL† [21]	74.45	75.55	77.48	79.01
U^{2}PL (w/ CutMix)	***70.30*** (+3.24)	***74.37*** (+2.54)	***76.47*** (+0.11)	***79.05*** (+0.80)
U^{2}PL (w/ AEL)	**74.90** (+0.45)	**76.48** (+0.93)	**78.51** (+1.03)	**79.12** (+0.11)
Ours (w/o CutMix Aug.)	70.50	75.71	77.41	80.08
Ours (w/ CutMix Aug.)	**74.72**	**77.62**	**79.21**	**80.21**
AEL (Ours)	**75.83**	**77.90**	**79.01**	**80.28**

Ablation Study

Table 6. **Ablation study on the effectiveness of various components in our U^{2}PL**, including unsupervised loss $\mathcal{L}_u$, contrastive loss $\mathcal{L}_c$, category-wise memory bank $\mathcal{Q}_c$, Dynamic Partition Adjustment (DPA), Probability Rank Threshold (PRT), and high entropy filtering (Unreliable).

$\mathcal{L}_c$	$\mathcal{Q}_c$	DPA	PRT	Unreliable	1/4 (2646)
					73.02
✓					77.08
✓	✓		✓	✓	78.49
✓	✓	✓		✓	79.07
✓	✓	✓	✓		77.57
✓	✓	✓	✓	✓	**79.30**

Ablation Study

Table 4. **Ablation study on using pseudo pixels with different reliability**, which is measured by the entropy of pixel-wise prediction (see Sec. 3.3). "Unreliable" denotes selecting negative candidates from pixels with top 20% highest entropy scores. "Reliable" denotes the bottom 20% counterpart. "All" denotes sampling regardless of entropy.

	Unreliable	Reliable	All
1/8 (1323)	**79.01**	77.30	77.40
1/4 (2646)	**79.30**	77.35	77.57

Table 5. **Ablation study on the probability rank threshold**, which is described in Sec. 3.3.

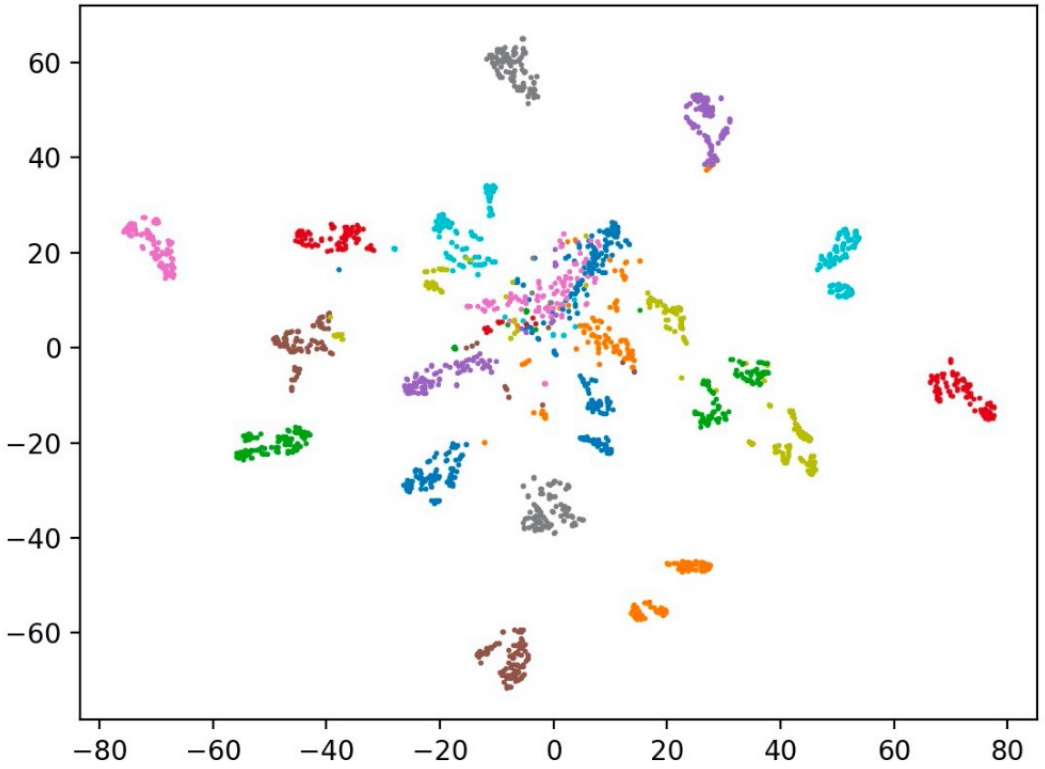
r_l	r_h	1/8 (1323)	1/4 (2646)
1	3	78.57	79.03
1	20	78.64	79.07
3	10	78.27	78.91
3	20	**79.01**	**79.30**
10	20	78.62	78.94

Table 7. **Ablation study on α_0 in Eq. (7)**, which controls the initial proportion between reliable and unreliable pixels.

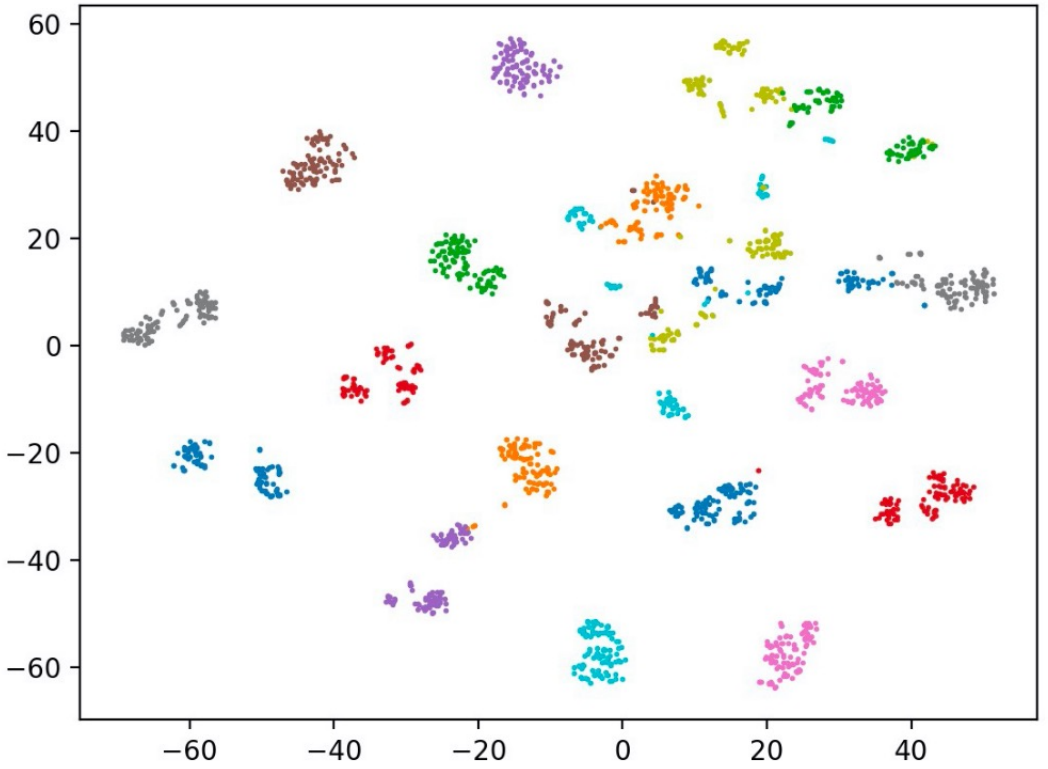
α_0	40%	30%	20%	10%
1/8 (1323)	76.77	77.34	**79.01**	77.80
1/4 (2646)	76.92	77.38	**79.30**	77.95

Analysis

Feature space visualization



(a) Supervised Only



(b) U^{2}PL

ST++: Make Self-training Work Better for Semi-supervised Semantic Segmentation

Lihe Yang[1] Wei Zhuo[3] Lei Qi[4,1] Yinghuan Shi[1,2*] Yang Gao[1]

[1]State Key Laboratory for Novel Software Technology, Nanjing University

[2]National Institute of Healthcare Data Science, Nanjing University

[3]Tencent [4]Southeast University

ST++

- motivation

1. construct a strong baseline of self-training via injecting strong data augmentations (SDA) on unlabeled images.

    alleviate overfitting noisy labels,

    decouple similar predictions between the teacher and student

2. [Novelty] selective re-training via prioritizing reliable unlabeled images.

    incorrect pseudo labels are prone to accumulate and degrade the performance.

ST++

Algorithm 1: ST Pseudocode

Input: Labeled training set $\mathcal{D}^l = \{(x_i, y_i)\}_{i=1}^M$,
Unlabeled training set $\mathcal{D}^u = \{u_i\}_{i=1}^N$,
Weak/strong augmentations $\mathcal{A}^w$/$\mathcal{A}^s$,
Teacher/student model T/S

Output: Fully trained student model S

Train T on $\mathcal{D}^l$ with cross-entropy loss $\mathcal{L}_{ce}$
Obtain pseudo labeled $\hat{\mathcal{D}}^u = \{(u_i, T(u_i))\}_{i=1}^N$
Over-sample $\mathcal{D}^l$ to around the size of $\hat{\mathcal{D}}^u$
for *minibatch* $\{(x_k, y_k)\}_{k=1}^B \subset (\mathcal{D}^l \cup \hat{\mathcal{D}}^u)$ **do**
  for $k \in \{1, \dots, B\}$ **do**
    if $x_k \in \mathcal{D}^u$ **then**
      $x_k, y_k \leftarrow \mathcal{A}^s(\mathcal{A}^w((x_k, y_k))$
    else
      $x_k, y_k \leftarrow \mathcal{A}^w(x_k, y_k)$
    $\hat{y}_k = S(x_k)$
  Update S to minimize $\mathcal{L}_{ce}$ of $\{(\hat{y}_k, y_k)\}_{k=1}^B$
return S

Algorithm 2: ST++ Pseudocode

Input: Same as Algorithm 1

Output: Same as Algorithm 1

Train T on $\mathcal{D}^l$ and save K checkpoints $\{T_j\}_{j=1}^K$
for $u_i \in \mathcal{D}^u$ **do**
  for $T_j \in \{T_j\}_{j=1}^K$ **do**
    Pseudo mask $M_{ij} = T_j(u_i)$
  Compute s_i with Equation 4 and $\{M_{ij}\}_{j=1}^K$
Select R highest scored images to compose $\mathcal{D}^{u_1}$
$\mathcal{D}^{u_2} = \mathcal{D}^u - \mathcal{D}^{u_1}$
$\mathcal{D}^{u_1} = \{(u_k, T(u_k))\}_{u_k \in \mathcal{D}^{u_1}}$
Train S on $(\mathcal{D}^l \cup \mathcal{D}^{u_1})$ with ST re-training
$\mathcal{D}^{u_2} = \{(u_k, S(u_k))\}_{u_k \in \mathcal{D}^{u_2}}$
Re-initialize S
Train S on $(\mathcal{D}^l \cup \mathcal{D}^{u_1} \cup \mathcal{D}^{u_2})$ with ST re-training
return S

Select and Prioritize Reliable Images

Observation:
positive correlation between the segmentation performance and the evolving stability of produced pseudo masks during the supervised training phase.

How to evaluate stability of pseudo masks?

Unlabeled image u_i,
K checkpoints $\{T_j\}_{j=1}^{K}$ saved during training,
Predict pseudo masks $\{M_{ij}\}_{j=1}^{K}$

$$s_i = \sum_{j=1}^{K-1} \texttt{meanIOU}\,(M_{ij}, M_{iK})$$

Experiments

Results on Pascal VOC val set

Network	Method	1/16 (662)	1/8 (1323)	1/4 (2645)	Network	Method	1/16 (662)	1/8 (1323)	1/4 (2645)
PSPNet ResNet-50	SupOnly	63.8	67.2	69.6	DeepLabv3+ ResNet-50	SupOnly	64.8	68.3	70.5
	CCT [41]	62.2	68.8	71.2		ECS [37]	-	70.2	72.6
	DCC [31]	67.1	71.3	72.5		DCC [31]	70.1	72.4	74.0
	ST	***69.1***	***73.0***	***73.2***		**ST**	***71.6***	***73.3***	***75.0***
	ST++	**69.9**	**73.2**	**73.4**		**ST++**	**72.6**	**74.4**	**75.4**
DeepLabv2 ResNet-101	SupOnly	64.3	67.6	69.5	DeepLabv3+ ResNet-101	SupOnly	66.3	70.6	73.1
	AdvSSL [26]	62.6	68.4	69.9		S4GAN [38]	69.1	72.4	74.5
	S4L [57]	61.8	67.2	68.4		GCT [28]	67.2	72.5	75.1
	GCT [28]	65.2	70.6	71.5		DCC [31]	72.4	74.6	76.3
	ST	***68.6***	***71.6***	***72.5***		**ST**	***72.9***	***75.7***	***76.4***
	ST++	**69.3**	**72.0**	**72.8**		**ST++**	**74.5**	**76.3**	**76.6**

Experiments

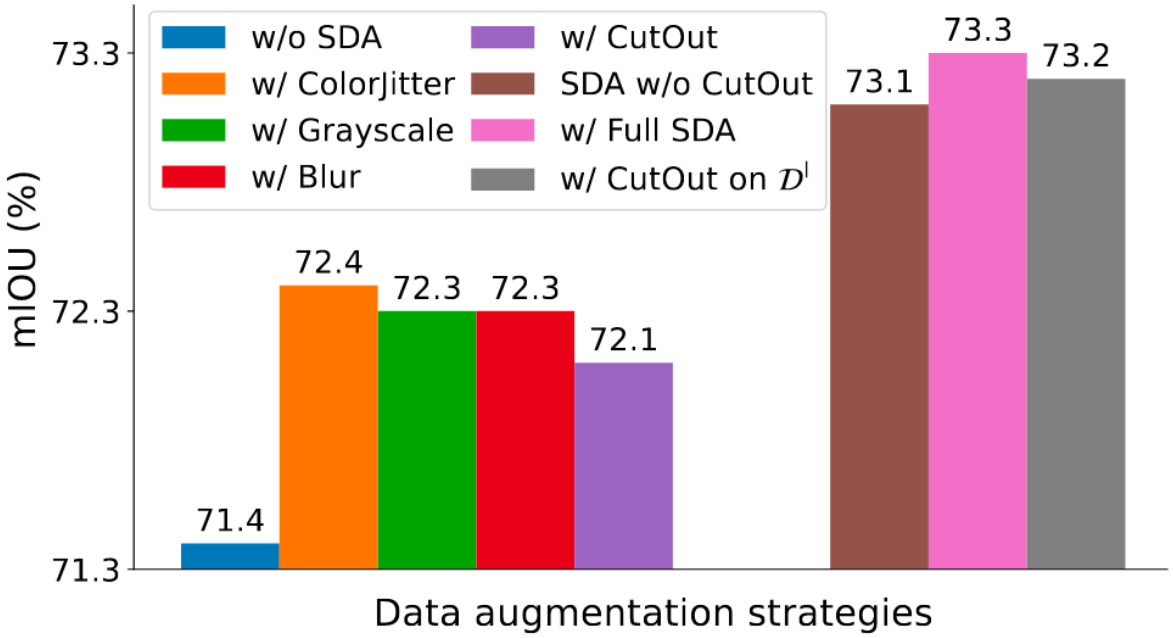
Results on cityscapes val set

Method	1/30 (100)	1/8 (372)	1/4 (744)
DeepLabv3+, ResNet-101			
DMT [17]	54.8	63.0	-
CutMix-Seg [18]	55.7	65.8	68.3
ClassMix [40]	-	61.4	63.6
PseudoSeg [66]	61.0	69.8	72.4
DeepLabv3+, ResNet-50			
SupOnly	55.1	65.8	68.4
DCC [31]	-	69.7	72.7
ST	***60.9***	***71.6***	***73.4***
ST++	**61.4**	**72.7**	**73.8**

Results on Pascal VOC val set

Method	# Labeled images (Total: 10582)				
	92	183	366	732	1464
SupOnly	50.7	59.1	65.0	70.6	74.1
GCT [28]	46.0	55.0	64.7	70.7	-
CutMix-Seg [18]	55.6	63.2	68.4	69.8	-
PseudoSeg [66]	57.6	65.5	69.1	72.4	73.2
CPS [12]	***64.1***	67.4	71.7	75.9	-
PC^2Seg [61]	57.0	66.3	69.8	73.1	74.2
ST	61.3	***68.2***	***73.5***	***76.3***	**78.9**
ST++	**65.2**	**71.0**	**74.6**	**77.3**	**79.1**
Fully-supervised setting (10582 images): 78.2					

Data Augmentation Strategies



Effectiveness of SDA

Apply SDA on labeled data	Apply SDA on unlabeled data	1/16 (662)	1/8 (1323)	1/4 (2645)
		70.9	71.4	73.5
✓	✓	71.0	73.0	74.3
	✓	**71.6**	**73.3**	**75.0**

Table 5. Effectiveness of full SDA. The first line without applying SDA is the plainest self-training [33]. The best results of only applying SDA on unlabeled data indicates that a more challenging optimization target for unlabeled data is vital to the success. And SDA on labeled data may destroy the clean data distribution.

Reliable Effectiveness

Reliable selection improve pseudo mask quality.

Need compare with retraining with all unlabeled data.

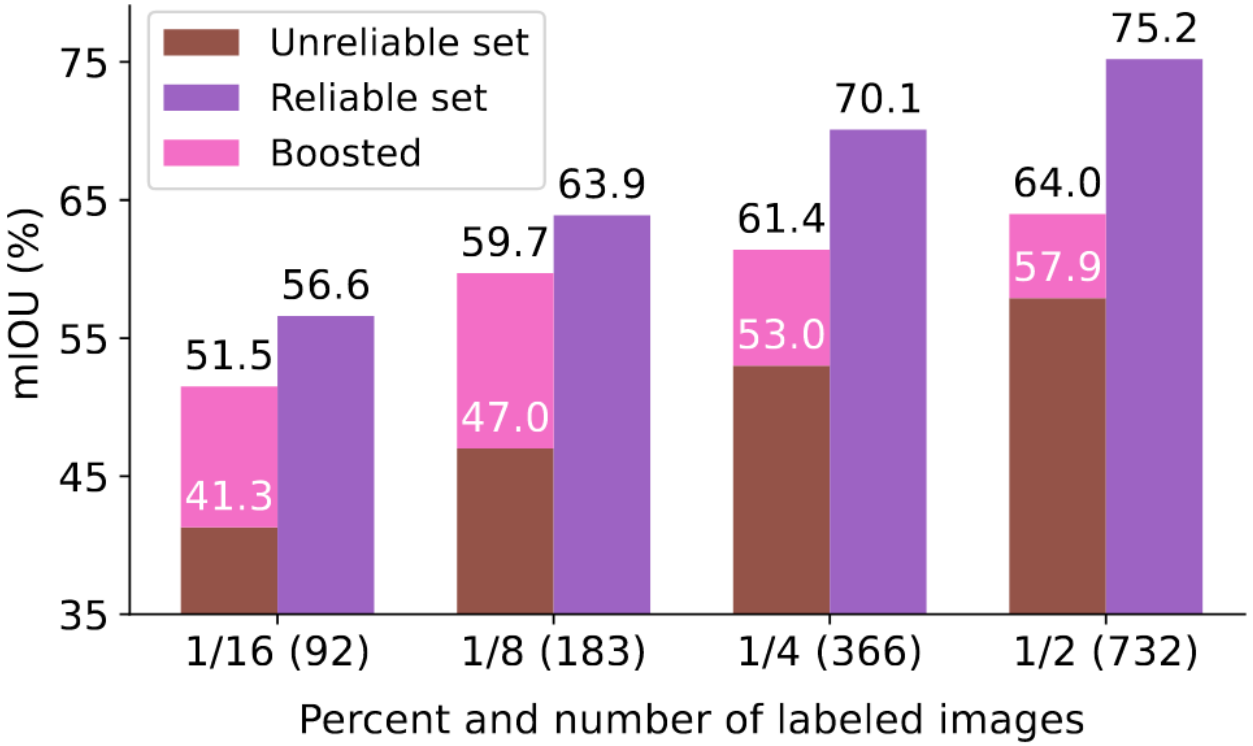


Figure 3. Pseudo mask quality of the reliable and unreliable images selected by ST++. The *Boosted* means the improved mIOU when re-labeling the unreliable images with the model trained on reliable images compared with only trained with labeled images.

Ablation Study

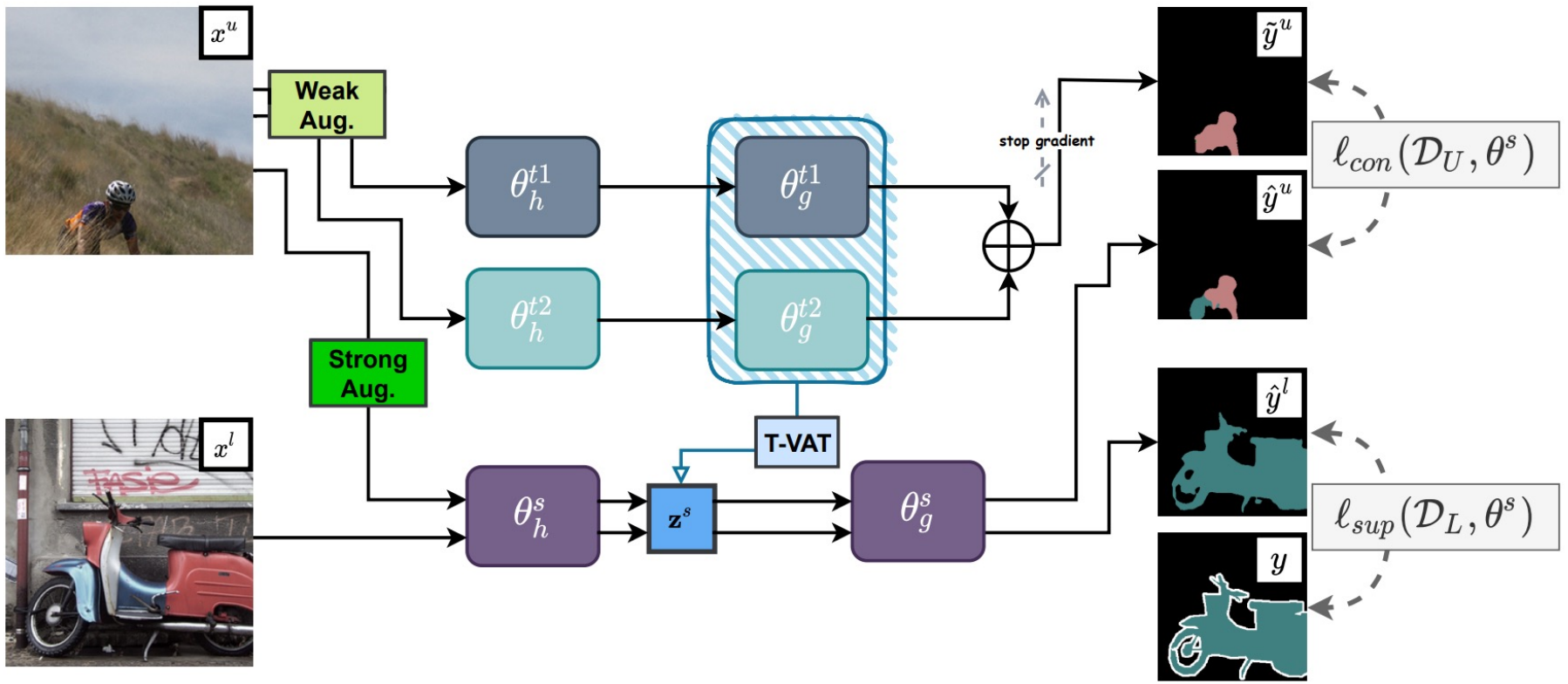
Ablation on different re-training methods

Method	1/16 (662)	1/8 (1323)	1/4 (2645)
One-stage re-training (our ST)	71.6	73.3	75.0
Random two-stage re-training	71.3	73.9	74.7
Selective re-training (our ST++)	**72.6**	**74.4**	**75.4**

Table 6. Effectiveness of the selective re-training in ST++. ST++ does not benefit from random two-stage re-training process, but the progressive reliable-to-unreliable selective re-training pipeline.

Perturbed and Strict Mean Teachers for Semi-supervised Semantic Segmentation

Yuyuan Liu[1] Yu Tian[1] Yuanhong Chen[1] Fengbei Liu[1]
Vasileios Belagiannis[2] Gustavo Carneiro[1]
[1] Australian Institute for Machine Learning, University of Adelaide
[2] Universität Ulm, Germany



UCC: Uncertainty guided Cross-head Co-training for Semi-Supervised Semantic Segmentation

Jiashuo Fan[1] Bin Gao[2] Huan Jin [2] Lihui Jiang [2*]

[1]Tsinghua-Berkeley Shenzhen Institute, Tsinghua University [2]Huawei Noah's Ark Lab