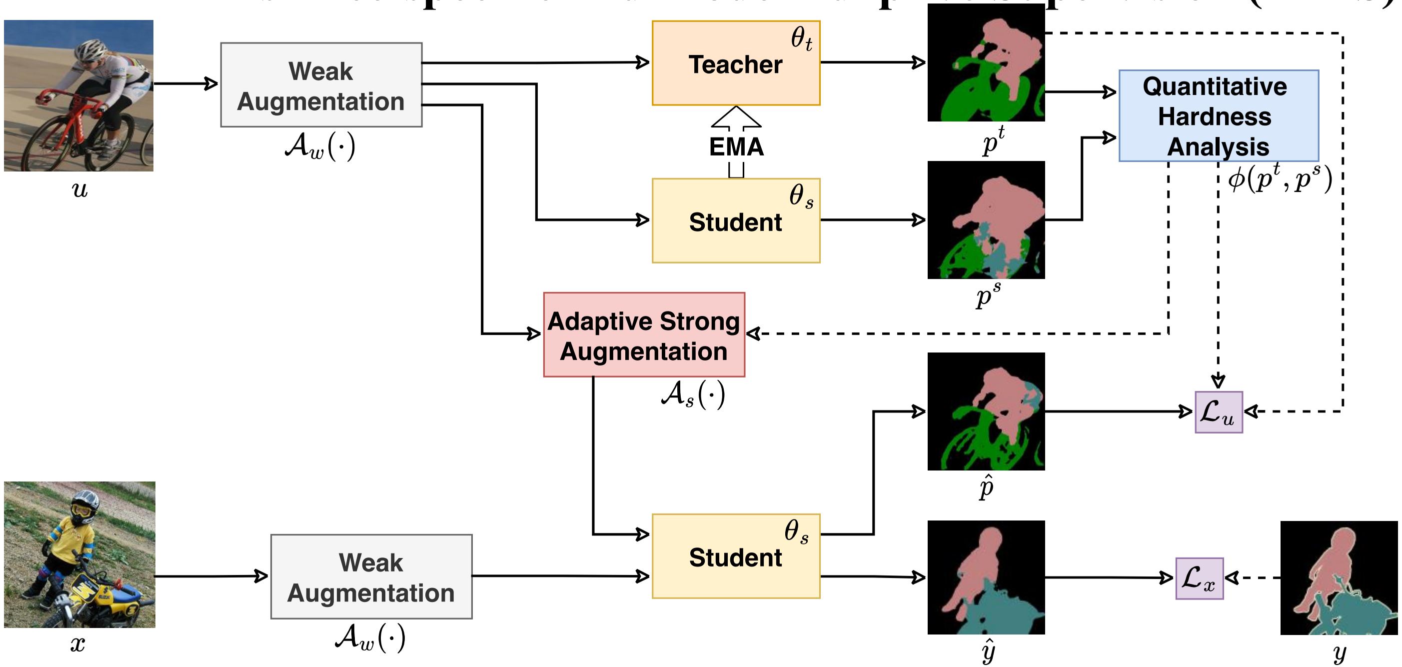
Instance-specific and Model-adaptive Supervision for Semi-supervised Semantic Segmentation

Zhen Zhao, Sifan Long, Jimin Pi, Luping Zhou, Jingdong Wang

THE UNIVERSITY OF SYDNEY

Brief Introduction

Instance-specific and Model-adaptive Supervision (iMAS)



- A standard teacher-student framework:
  - Quantitative hardness analysis: 1) unlabeled instance 2) model's training status
  - Hardness-based Model-adaptive supervision: 1) augmentations 2) unsupervised loss

Outline

- **Background and Motivation**
- **Proposed Method**
- **Experiments**
- **Conclusion**

Background

- **Why** Semi-supervised Semantic Segmentation (SSS)?
  - The success of supervised semantic segmentation depends closely on large datasets with **high-quality pixel-level** annotations.
  - Delicate and dense pixel-level labelling is **costly and time-consuming**, which becomes a significant bottleneck in practical applications with limited labelled data.

- **How** recent SSS work? **(leveraging the unlabeled data)**
  - **Pseudo-labeling:** Train on labeled data and then generate pseudo-labels on unlabeled data, iteratively adding high-confidence predicted unlabeled data to labeled set.
  - **Consistency regularization:** Apply data or model perturbations and enforce the prediction consistency between two differently-perturbed views for unlabeled data.

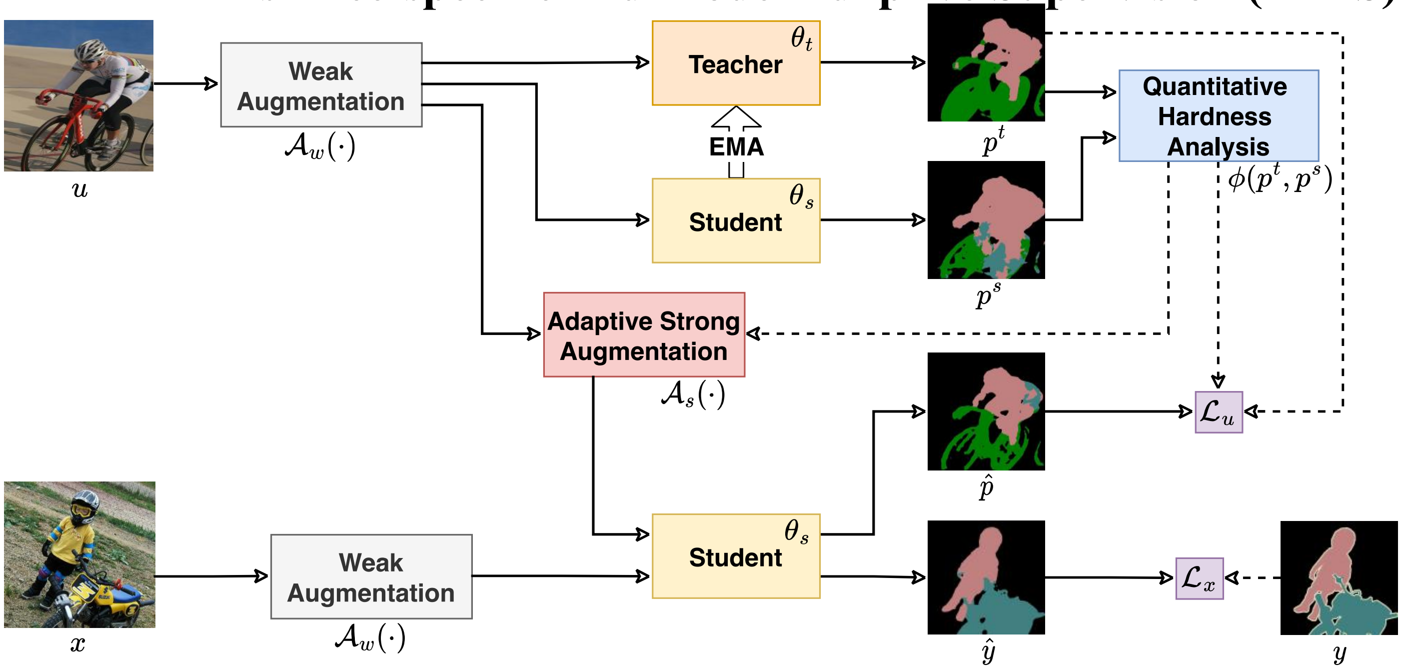
Motivation

- **Motivations:**
  - Weakness1: Despite their promising performance, **recent SSS studies** come at the cost of introducing extra network components or additional training procedures.
  - Weakness2: Most existing studies treat all unlabeled data equally and barely consider the differences and training difficulties among unlabeled instances.
  - We believe that differentiating unlabeled instances can promote instance-specific supervision to adapt to the model's evolution dynamically.

- **Our Goal: instance-specific and model-adaptive supervision (iMAS)**
  - all the operations on different unlabeled instances should
    - adapt to the training status of the model
    - be adjusted based on their learning difficulties

Method

Instance-specific and Model-adaptive Supervision (iMAS)



- A standard teacher-student framework:
  - Quantitative hardness analysis: 1) unlabeled instance 2) model's training status
  - Hardness-based Model-adaptive supervision: 1) augmentations 2) unsupervised loss

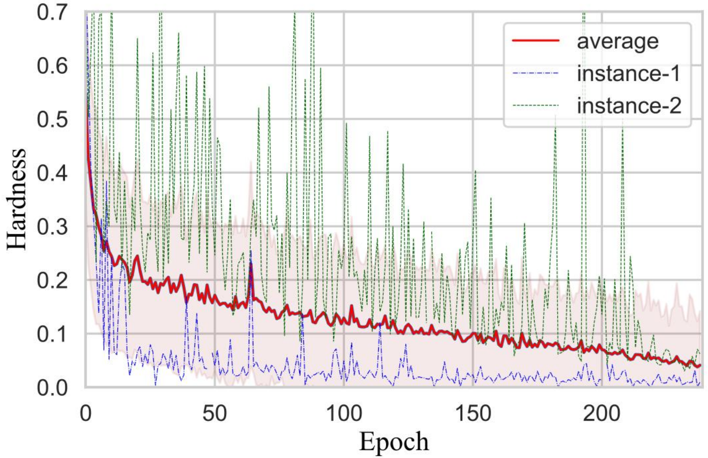
Method (cont.)

- **Quantitative hardness analysis**
  - A class-weighted teacher-student symmetric IoU.

$$p_i^s = f_{\theta_s}(\mathcal{A}_w(u_i)), \quad \rho_i^s = \frac{1}{H \times W} \sum_{j=1}^{H \times W} \mathbb{1}(\max(p_i^s(j)) \geq \tau)$$

$$p_i^t = f_{\theta_t}(\mathcal{A}_w(u_i)), \quad \rho_i^t = \frac{1}{H \times W} \sum_{j=1}^{H \times W} \mathbb{1}(\max(p_i^t(j)) \geq \tau),$$

$$\gamma_i = \phi(p_i^t, p_i^s) = 1 - [\frac{\rho_i^s}{2} \mathrm{wIOU}(p_i^s, p_i^t) + \frac{\rho_i^t}{2} \mathrm{wIOU}(p_i^t, p_i^s)]$$



Method (cont.)

- **Model-adaptive supervision**
  - Hardness-adaptive data perturbations:
    - Intensity-based augmentations

$$\mathcal{A}_s^I(u_i) \leftarrow \gamma_i \mathcal{A}_s^I(u_i) + (1-\gamma_i)\mathcal{A}_w(u_i) \quad (8)$$

    - Cut-mix-based augmentations

$$\left.\begin{aligned} \mathcal{A}_s^C(u_m) &\leftarrow M_m \odot u_n + (\mathbf{1} - M_m) \odot u_m \\ p_m^{t'} &\leftarrow M_m \odot p_n^t + (\mathbf{1} - M_m) \odot p^t, \\ \mathcal{A}_s^C(u_n) &\leftarrow M_n \odot u_m + (\mathbf{1} - M_n) \odot u_n \\ p_n^{t'} &\leftarrow M_n \odot p_m^t + (\mathbf{1} - M_n) \odot p_n^t \end{aligned}\right\}, \text{by a trigger probability of } \overline{\gamma} = \frac{1}{|\mathcal{B}_u|}\sum_{n=1}^{|\mathcal{B}_u|} \gamma_n \quad (9)$$

  - Hardness-adaptive unsupervised loss:

$$\begin{aligned} \mathcal{L}_u = \frac{1}{|\mathcal{B}_u|}\sum_{i=1}^{|\mathcal{B}_u|} \frac{1-\gamma_i}{2H \times W} \sum_{j=1}^{H \times W} [&\mathbb{1}(\max(p_i^t(j)) \geq \tau)\mathrm{H}(f_{\theta_s}(\mathcal{A}_s^I(u_i)), p_i^t(j)) + \\ &\mathbb{1}(\max(p_i^{t'}(j)) \geq \tau)\mathrm{H}(f_{\theta_s}(\mathcal{A}_s^C(u_i)), p_i^{t'}(j))]. \end{aligned} \quad (10)$$

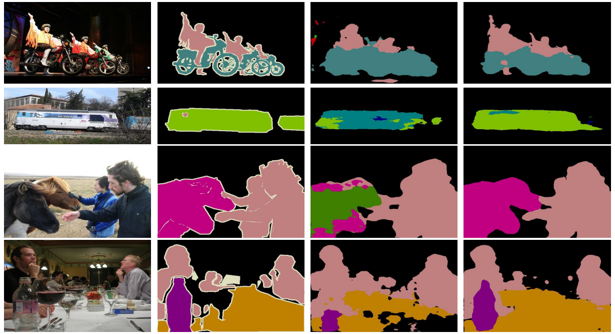
Experiments: Comparison with SOTAs

Method	ResNet-50			ResNet-101		
	1/16 (662)	1/8 (1323)	1/4 (2646)	1/16 (662)	1/8 (1323)	1/4 (2646)
Supervised*	63.8	69.0	72.5	67.4	72.1	74.7
MT [39]	66.8	70.8	73.2	70.6	73.2	76.6
CCT [33]	65.2	70.9	73.4	68.0	73.0	76.2
CutMix-Seg [12]	68.9	70.7	72.5	72.6	72.7	74.3
GCT [22]	64.1	70.5	73.5	69.8	73.3	75.3
CAC [24]	70.1	72.4	74.0	72.4	74.6	76.3
CPS [7]	72.0	73.7	74.9	74.5	76.4	77.7
PSMT† [26]	72.8	75.7	76.4	75.5	78.2	78.7
ELN [23]	70.5	73.2	74.6	72.5	75.1	76.6
ST++ [45]	72.6	74.4	75.4	74.5	76.3	76.6
iMAS (ours)	74.8	76.5	77.0	76.5	77.9	78.1
U^2PL‡ [41]	72.0	75.2	76.2	74.4	77.6	78.7
iMAS (ours)‡	**75.9**	**76.7**	**77.1**	**77.2**	**78.4**	**79.3**

Table 2. Comparison with SOTA methods on **PASCAL VOC 2012** `val` set under different partition protocols. Labeled images are sampled from the *blender* training set (augmented by SBD dataset), including 10, 583 samples in total. ‡ means the results are obtained by setting the output_stride as 8 in DeepLabV3+ (16 for others). * denotes our reproduced results. Best results are highlighted in **bold**.

Method	1/16 (92)	1/8 (183)	1/4 (366)	1/2 (732)	Full (1464)
Supervised *	45.5	57.5	66.6	70.4	72.9
CutMix-Seg [12]	52.2	63.5	69.5	73.7	76.5
PseudoSeg [55]	57.6	65.5	69.1	72.4	73.2
PC^2Seg [52]	57.0	66.3	69.8	73.1	74.2
CPS [7]	64.1	67.4	71.7	75.9	-
PSMT [26]	65.8	69.6	76.6	78.4	80.0
ST++ [45]	65.2	71.0	74.6	77.3	79.1
iMAS (ours)	68.8	74.4	78.5	79.5	81.2
U^2PL‡ [41]	68.0	69.2	73.7	76.2	79.5
iMAS‡(ours)	**70.0**	**75.3**	**79.1**	**80.2**	**82.0**

Table 3. Comparison with SOTA methods on *classic* **PASCAL VOC 2012** `val` set under different partition protocols. Labeled images are sampled from the official VOC `train` set, including 1, 464 samples in total. Results are reported using Resnet-101. All notations are the same as in Table 2.



rbg gt sup ours

Experiments (Cont.)

Method	1/16 (186)	1/8 (372)	1/4 (744)	1/2 (1488)
Supervised *	64.0	69.2	73.0	76.4
MT [39]	66.1	72.0	74.5	77.4
CCT [33]	66.4	72.5	75.7	76.8
GCT [22]	65.8	71.3	75.3	77.1
CPS [7]	74.4	76.6	77.8	78.8
CPS† [41]	69.8	74.3	74.6	76.8
PSMT [26]	-	75.8	76.9	77.6
ELN [23]	-	70.3	73.5	75.3
ST++ [45]	-	72.7	73.8	-
U^{2}PL * [41]	67.8	72.5	74.8	77.1
iMAS (ours)	74.3	77.4	78.1	79.3
U^{2}PL‡* [41]	69.0	73.0	76.3	78.6
iMAS (ours)‡	**75.2**	**78.0**	**78.2**	**80.2**

Table 4. Comparison with SOTA methods on **Cityscapes** `val` set under different partition protocols. Labeled images are sampled from the Cityscapes `train` set, including 2,975 samples in total. Results are reported using Resnet-50. * and † represent reproduced results in iMAS and U^{2}PL, respectively. Results with ‡ are obtained by setting the output_stride as 8 in DeepLabV3+.

iMAS on: Loss $\mathcal{L}_u$	Augs of $\mathcal{A}_s^I$	Augs of $\mathcal{A}_s^C$	mIOU (%)
			72.1 (supervised)
✓			75.5 (3.4↑)
	✓		76.5 (4.4↑)
		✓	76.9 (4.8↑)
✓	✓	✓	**77.9** (**5.8↑**)

Table 5. Ablation studies on the effectiveness of the instance-specific model-adaptive supervision on the unsupervised loss, intensity-based and CutMix augmentations, respectively. Results are reported on **PASCAL VOC 2012** under the 1/8 (1323) partition using Resnet-101 as the backbone. Improvements over the supervised baseline are marked in blue.

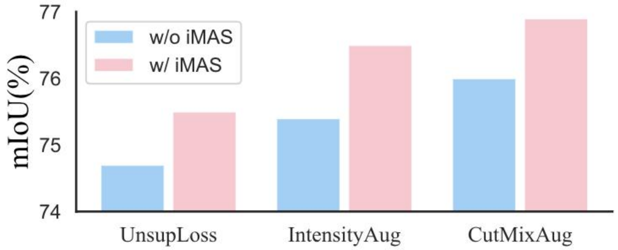


Figure 2. Effectiveness of iMAS on the unsupervised loss, intensity-based and CutMix augmentations, respectively.

Conclusion

- In this paper, we highlight the instance uniqueness and propose iMAS, an instance-specific and model-adaptive supervision for semi-supervised semantic segmentation.

- Relying on our class-weighted symmetric hardness-evaluating strategies, iMAS treats each unlabeled instance discriminatively and employ model-adaptive augmentation and loss weighting strategies on each instance.

- Without introducing additional networks or losses, iMAS can remarkably improve the SSS performance.

Thank You

THE UNIVERSITY OF SYDNEY

Experiments (Cont.)