Augmentation Matters: A Simple-yet-Effective Approach to Semi-supervised Semantic Segmentation

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

THE UNIVERSITY OF SYDNEY

Brief Introduction

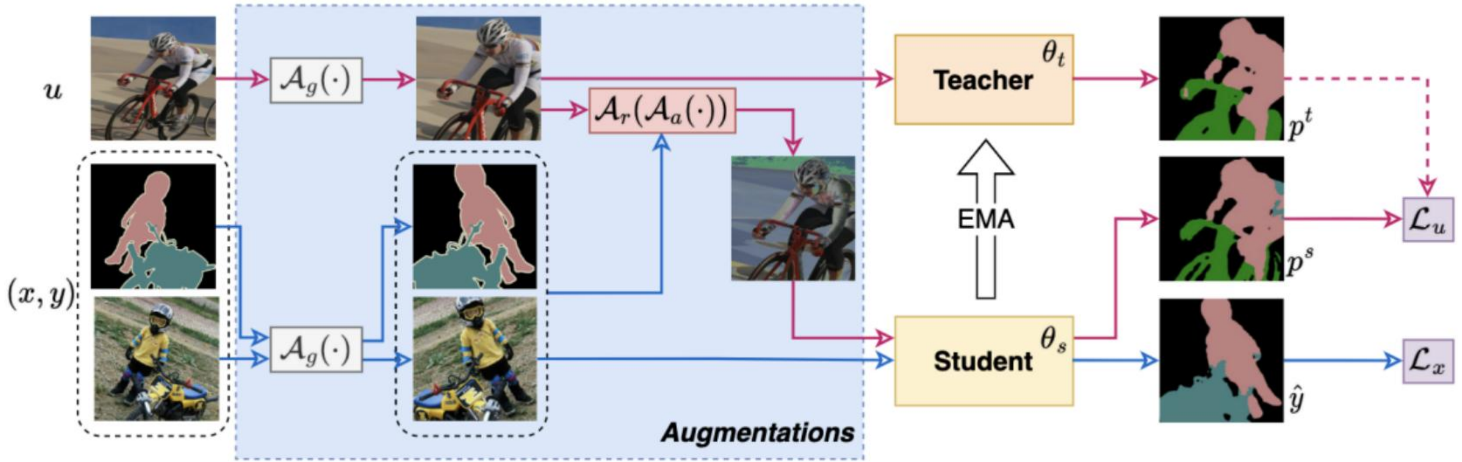


Figure 2. Diagram of AugSeg. In a standard teacher-student framework, AugSeg trains the student model, parameterized by θ_s, on labeled data (x, y) and unlabeled data u simultaneously, via minimizing the corresponding supervised loss $\mathcal{L}_x$ and unsupervised consistency loss $\mathcal{L}_u$, respectively. The teacher model, parameterized by θ_t, is updated by the exponential moving averaging (EMA) of θ_s, and generates the pseudo-label on unlabeled data, p^t. The core of AugSeg is to apply various augmentation techniques on input unlabeled samples, including the weak geometrical augmentation $\mathcal{A}_g$, the random intensity-based augmentation $\mathcal{A}_r$ and the adaptive label-injecting augmentation $\mathcal{A}_a$. The red and blue lines represent the forward path of labeled and unlabeled data, respectively. The dashed line means"stop gradient".

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

➢ Existing studies are Complicated

- ➢ **Augmentations**
  - ➢Image-level
  - ➢Feature-level
- ➢ **Auxiliary tasks**
  - ➢Contrastive loss
  - ➢Multiple branch
  - ➢Co-training
- ➢ **Pseudo-rectifying**
  - ➢Various filtering
  - ➢Correcting networks
  - ➢Prior-based

Method	Augmentations		More Supervision			Pseudo-rectifying		
	SDA	FT	MBSL	CT	UCL	UAFS	ACN	PR
CCT [42]		✓	✓	✓				
ECS [36]			✓				✓	
SSMT [24]	✓		✓			✓		
PseudoSeg [55]	✓					✓		
CAC [29]			✓		✓	✓		
DARS [22]	✓		✓					✓
AEL [23]	✓							✓
PC^{2}Seg [53]	✓		✓		✓	✓		
C3-Semiseg [54]	✓				✓	✓		✓
SimpleBase [48]	✓		✓			✓		
ReCo [32]	✓				✓	✓		
CPS [7]	✓			✓				
ST++ [47]	✓		✓					
ELN [28]	✓		✓				✓	
USRN [19]	✓		✓			✓		✓
PSMT [34]	✓	✓	✓			✓		
U^{2}PL [45]	✓				✓	✓		✓
AugSeg (ours)	✓							

Can we make it simpler?

Motivation (Cont.)

- **What is the key in current dominant CR-based approaches?**
  - to produce **prediction disagreements** on unlabeled data,
    - such that unlabeled samples can be leveraged to train models even if their labeled information is un-known.

- **How can we produce such disagreement?**
  - **Data** / model / feature perturbations
    - **Data-level:** auto-augmentations, mix-based augmentations (**our focus**)
    - Model-level: different architectures, dropout, stochastic depth
  - We argue that:
    - *Various data augmentations should be adjusted to better adapt to the semi-supervised scenarios instead of directly applying these techniques from supervised learning.*

Method

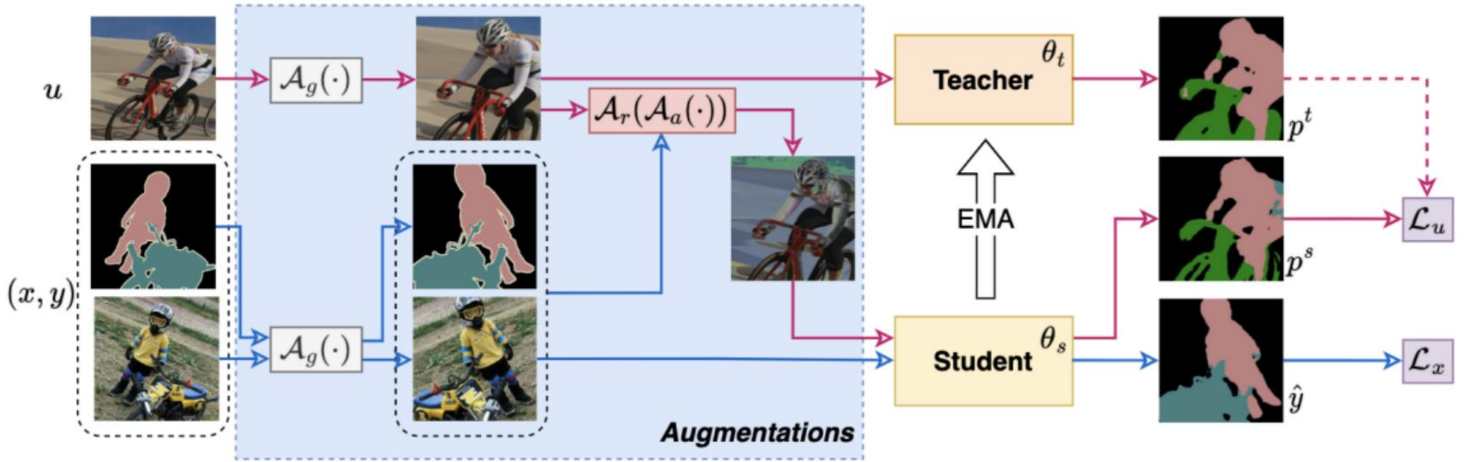


Figure 2. Diagram of AugSeg. In a standard teacher-student framework, AugSeg trains the student model, parameterized by θ_s, on labeled data (x, y) and unlabeled data u simultaneously, via minimizing the corresponding supervised loss $\mathcal{L}_x$ and unsupervised consistency loss $\mathcal{L}_u$, respectively. The teacher model, parameterized by θ_t, is updated by the exponential moving averaging (EMA) of θ_s, and generates the pseudo-label on unlabeled data, p^t. The core of AugSeg is to apply various augmentation techniques on input unlabeled samples, including the weak geometrical augmentation $\mathcal{A}_g$, the random intensity-based augmentation $\mathcal{A}_r$ and the adaptive label-injecting augmentation $\mathcal{A}_a$. The red and blue lines represent the forward path of labeled and unlabeled data, respectively. The dashed line means"stop gradient".

Method (cont.)

- Random Intensity-based Augmentations
  - Sample the distorting degree **uniformly** in a **continuous** space instead of a finite discrete space.
  - Sample a **random number** of augmentations, bounded by a maximum value of k, from an augmentation pool instead of using a fixed number.
  - Remove strong transformations e.g. Invert operations

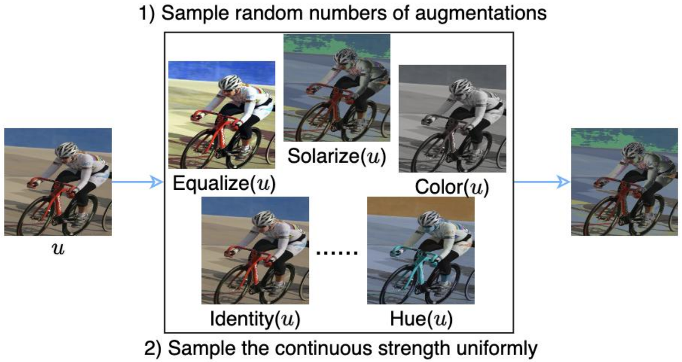


Figure 3. A visualization of random intensity-based augmentation.

Method (cont.)

- Adaptive CutMix-based augmentations

$$\rho_i = \frac{1}{H \times W} \sum_{j=1}^{H \times W} \max(p_i^t(j))\left(1 - \frac{-\sum p_i^t(j) \log p_i^t(j)}{\log N}\right)$$

$$u_n^{'} \leftarrow M_n \odot u_n + (\mathbf{1} - M_n) \odot x_n,$$

$$\mathcal{A}_a(u_m) \leftarrow M_m \odot u_m + (\mathbf{1} - M_m) \odot u_n^{'},$$

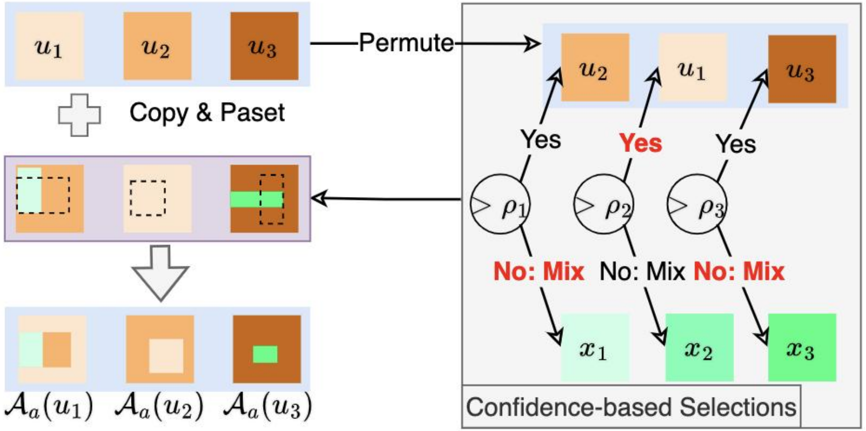


Figure 4. A visualization of adaptive label-injecting CutMix augmentation in a mini-batch. x_i and u_i denote the labeled and unlabeled crops, respectively. ρ_i denote the confidence score for i-th unlabeled sample. The core idea of $\mathcal{A}_a$ is that, these less confident unlabeled samples, with lower values of ρ_i, are more likely to be aided (mixed) by these confident labeled samples.

Experiments: Ablation Study

AugSeg			mIoU	
MT	$\mathcal{A}_r$	$\mathcal{A}_a$	VOC (366)	Citys (744)
			61.65 (supervised)	74.14 (supervised)
✓			69.06 (7.41↑)	75.96 (1.82↑)
✓	✓		72.41 (10.76↑)	77.29 (3.15↑)
✓		✓	74.33 (12.68↑)	77.44 (3.30↑)
✓	✓	✓	**76.17** (14.52↑)	**78.76**(4.62↑)

Table 6. Ablation studies on our AugSeg. "MT" means the standard mean-teacher semi-supervised framework. $\mathcal{A}_r$ and $\mathcal{A}_a$ represent the two main augmentation strategies, the random intensity-based and adaptive label-injecting augmentations, respectively. Improvements over the supervised baseline are highlighted in blue.

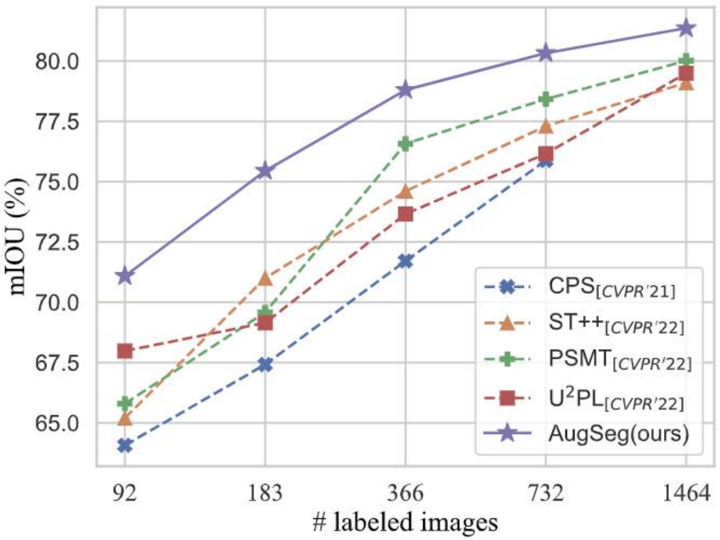
λ_u	0.0	0.5	1.0	1.5	2.0
VOC (366)	61.65	75.21	76.17	75.95	**77.05**
Citys (744)	74.14	77.02	78.76	**78.99**	78.68

Table 7. Ablations on the loss weight λ_u, set as 1.0 by default.

k	0	1	2	3	4
VOC (366)	74.38	75.50	76.10	76.17	**76.32**
Citys (186)	71.26	72.10	73.42	**73.73**	73.03
Citys (744)	77.44	78.34	78.11	**78.76**	78.48

Table 8. Ablations on the maximum number of selected intensity-based augmentations, using R50 as the encoder. $k = 3$ by default.

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	63.72	68.49	72.46	67.76	72.13	75.04
MT [44]	66.77	70.78	73.22	70.59	73.20	76.62
CCT [42]	65.22	70.87	73.43	67.94	73.00	76.17
GCT [27]	64.05	70.47	73.45	69.77	73.30	75.25
CPS [7]	68.21	73.20	74.24	72.18	75.83	77.55
CPS w/ CutMix [7]	71.98	73.67	74.90	74.48	76.44	77.68
ST++ [47]	72.60	74.40	75.40	74.50	76.30	76.60
PS-MT [34]	72.83	75.70	76.43	75.50	78.20	78.72
AugSeg	74.66	75.99	77.16	77.01	77.31	78.82
Supervised‡	67.66	71.91	74.53	70.63	75.02	76.47
U^2PL‡∗ [45]	74.74	77.44	77.51	77.21	79.01	79.30
AugSeg‡	77.28	78.27	78.24	79.29	81.46	80.50

Method	ResNet-50				ResNet-101			
	1/16(186)	1/8(372)	1/4(744)	1/2(1488)	1/16(186)	1/8(372)	1/4(744)	1/2(1488)
Supervised	63.34	68.73	74.14	76.62	64.77	71.64	75.24	78.03
MT [44]	66.14	72.03	74.47	77.43	68.08	73.71	76.53	78.59
CCT [42]	66.35	72.46	75.68	76.78	69.64	74.48	76.35	78.29
GCT [27]	65.81	71.33	75.30	77.09	66.90	72.96	76.45	78.58
CPS [7]	69.79	74.39	76.85	78.64	70.50	75.71	77.41	80.08
CPS * [7]	-	-	-	-	69.78	74.31	74.58	76.81
PS-MT† [34]	-	75.76	76.92	77.64	-	76.89	77.60	79.09
U^2PL [45]	69.03	73.02	76.31	78.64	70.30	74.37	76.47	79.05
AugSeg	73.73	76.49	78.76	79.33	75.22	77.82	79.56	80.43

Conclusion

- We propose a simple-yet-effective SSS framework, **AugSeg**, which follows a standard two-branch teacher-student framework to train models on labeled and unlabeled data jointly.

- We break the trend of SSS studies that integrate increasingly complex designs, and revise the widely-adopted data augmentations to better **adapt to SSS** tasks by injecting labeled information adaptively and simplifying the standard RandomAug with a highly random design.

- Without any additional complicated designs, AugSeg readily obtains **new SOTA** performance on popular SSS benchmarks under different partition protocols.

Thank You

THE UNIVERSITY OF SYDNEY