TriMix: A General Framework for Medical Image Segmentation from Limited Supervision -Appendix-

Zhou Zheng[1,*]   Yuichiro Hayashi[1]   Masahiro Oda[1]
Takayuki Kitasaka[2]   Kensaku Mori[1,3,*]

[1]Nagoya University   [2]Aichi Institute of Technology
[3]National Institute of Informatics
[*]zzheng@mori.m.is.nagoya-u.ac.jp, kensaku@is.nagoya-u.ac.jp

Appendix A: A Brief Summary of Related Weakly-Supervised Segmentation Methods

Table 5. A brief summary of related weakly-supervised segmentation methods. Most existing methods are based on the image-level, box, and scribble annotations and sometimes additionally utilize saliency prior. **Different from existing schemes, our method pioneers exploring a general solution for semi- and scribble-supervised segmentation.**
$\mathcal{I}$: image-level annotation. $\mathcal{P}$: pixel-level annotation. $\mathcal{B}$: box annotation.
$\mathcal{U}$: unlabeled data. $\mathcal{SP}$: saliency prior. $\mathcal{S}$: scribble annotation.

method	annotation	method	annotation
Papandreou *et al.* (ICCV'15) [1]	$\mathcal{I}/\mathcal{I}+\mathcal{P}/\mathcal{B}+\mathcal{P}$	Lee *et al.* (CVPR'21) [2]	$\mathcal{I}/\mathcal{I}+\mathcal{P}$
Kolesnikov *et al.* (ECCV'16) [3]	$\mathcal{I}$	Li *et al.* (CVPR'22) [4]	$\mathcal{I}/\mathcal{I}+\mathcal{SP}$
Khoreva *et al.* (CVPR'17) [5]	$\mathcal{B}$	Pan *et al.* (IJCV'22) [6]	$\mathcal{I}/\mathcal{I}+\mathcal{P}/\mathcal{U}+\mathcal{P}$
Roy *et al.* (CVPR'17) [7]	$\mathcal{I}$	Lin *et al.* (CVPR'16) [8]	$\mathcal{S}$
Wei *et al.* (CVPR'17) [9]	$\mathcal{I}$	Vernaza *et al.* (CVPR'17) [10]	$\mathcal{S}$
Kim *et al.* (ICCV'17) [11]	$\mathcal{I}$	Tang *et al.* (ECCV'18) [12]	$\mathcal{S}$
Chaudhry *et al.* (BMVC'17) [13]	$\mathcal{I}$	Tang *et al.* (CVPR'18) [14]	$\mathcal{S}$
Ge *et al.* (CVPR'18) [15]	$\mathcal{I}$	Wang *et al.* (CVPR'19) [16]	$\mathcal{S}$
Li *et al.* (CVPR'18) [17]	$\mathcal{I}/\mathcal{I}+\mathcal{P}$	Marin *et al.* (CVPR'19) [18]	$\mathcal{S}$
Wang *et al.* (CVPR'18) [19]	$\mathcal{I}$	Ji *et al.* (MICCAI'19) [20]	$\mathcal{S}$
Ahn *et al.* (CVPR'18) [21]	$\mathcal{I}$	Lee *et al.* (MICCAI'20) [22]	$\mathcal{S}$
Huang *et al.* (CVPR'18) [23]	$\mathcal{I}$	Valvano *et al.* (TMI'21) [24]	$\mathcal{S}$
Wei *et al.* (CVPR'18) [25]	$\mathcal{I}/\mathcal{I}+\mathcal{P}$	Zhang *et al.* (TPAMI'21) [26]	$\mathcal{S}$
Lee *et al.* (CVPR'19) [27]	$\mathcal{I}/\mathcal{I}+\mathcal{P}$	Lu *et al.* (PR'21) [28]	$\mathcal{S}$
Song *et al.* (CVPR'19) [29]	$\mathcal{B}$	Chen *et al.* (ICCV'21) [30]	$\mathcal{S}$
Ahn *et al.* (CVPR'19) [31]	$\mathcal{I}$	Xu *et al.* (ICCV'21) [32]	$\mathcal{S}$
Li *et al.* (ICCV'19) [33]	$\mathcal{I}+\mathcal{SP}$	Pan *et al.* (ICCV'21) [34]	$\mathcal{S}$
Shimoda *et al.* (ICCV'19) [35]	$\mathcal{I}$	Liang *et al.* (CVPR'22) [36]	$\mathcal{S}$
Zhang *et al.* (ECCV'20) [37]	$\mathcal{I}$	Zhang *et al.* (CVPR'22) [38]	$\mathcal{S}$
Sun *et al.* (ECCV'20) [39]	$\mathcal{I}/\mathcal{I}+\mathcal{SP}$	Unal *et al.* (CVPR'22) [40]	$\mathcal{S}$
Chen *et al.* (ECCV'20) [41]	$\mathcal{I}$	Luo *et al.* (MICCAI'22) [42]	$\mathcal{S}$
Wang *et al.* (CVPR'20) [43]	$\mathcal{I}$	Zhang *et al.* (MICCAI'22) [44]	$\mathcal{S}$
Chang *et al.* (CVPR'20) [45]	$\mathcal{I}$	Liu *et al.* (PR'22) [46]	$\mathcal{S}$
Zhang *et al.* (NeurIPS'20) [47]	$\mathcal{I}$	Gao *et al.* (MedIA'22) [48]	$\mathcal{S}+\mathcal{U}$
Lee *et al.* (CVPR'21) [49]	$\mathcal{B}$	TriMix (ours)	$\mathcal{S}/\mathcal{U}+\mathcal{P}$

Appendix B: Supplementary of 3D Semi-Supervised Segmentation on ACDC and Hippocampus Datasets

Implementation of 3D semi-supervised methods. When evaluating TriMix on 3D semi-supervised segmentation on the ACDC and hippocampus datasets, we implemented several SSL approaches and compared TriMix to them. We implemented MT[1] [50], UA-MT[2] [51], STS-MT[3] [52], CutMix-Seg[4] [53], and CPS[5] [54] based on their official codebases. UMCT [55] is not open-source, and we reproduced it thanks to its straightforward idea and sufficient technical details in the original paper. There are seven methods for comparison: MT, UA-MT, CutMix-Seg, STS-MT, CPS, UMCT, and TriMix. All methods shared some similar settings: the same backbone V-Net [56], 300 training epochs, an SGD with a weight decay of 0.0001, and a momentum of 0.9. The learning rate was divided by 10 every 100 epochs. Since UA-MT and UMCT require uncertainty estimation, we added dropout layers to V-Net, identical to the work [51]. We extended CutMix [57] to 3D. CutMix-Seg, CPS, and TriMix were trained with the same 3D CutMix augmentation, for which the cropped volume ratio was set to 0.2. UMCT was trained with three different views. We tuned hyperparameters for the methods based on cross-validation.

Comparison of trainable parameters, training and inference time. Table 6 compares our method with other SSL methods regarding trainable parameters, average training time, and average inference time required per fold validation on the ACDC and hippocampus datasets. UMCT and TriMix contain the maximum number of parameters. UA-MT is the most time-consuming for training among the mean-teacher-based methods since it needs several forward passes to model uncertainty. UMCT costs more time for training than CPS and TriMix since it also requires several forward passes to model uncertainty. For the inference time, each network within CPS, UMCT, and TriMix requires a forward pass if predictions from all networks are required.

Table 6. Comparison of number of trainable parameters, training and inference time among semi-supervised methods.

method	backbone	params (M)	hippocampus (1/196), NVIDIA GeForce GTX 1080 Ti		ACDC (16/150), NVIDIA RTX A6000	
			training/fold (s)	inference/fold (s)	training/fold (s)	inference/fold (s)
MT	V-Net	9.44	232.08	2.20	1130.72	45.95
UA-MT	V-Net	9.44	265.21	2.20	1964.63	45.95
CutMix-Seg	V-Net	9.44	239.79	2.20	1229.13	45.95
STS-MT	V-Net	9.44	231.70	2.20	1117.97	45.95
CPS	V-Net	2×9.44	249.24	2×2.20	1917.51	2×45.95
UMCT	V-Net	3×9.44	384.20	3×2.20	5741.31	3×45.95
TriMix	V-Net	3×9.44	270.25	3×2.20	2777.57	3×45.95

---

[1] https://github.com/CuriousAI/mean-teacher

[2] https://github.com/yulequan/UA-MT

[3] https://github.com/tengteng95/Spatial Ensemble

[4] https://github.com/Britefury/cutmix-semisup-seg

[5] https://github.com/charlesCXK/TorchSemiSeg

Table 7. Comparison with semi-supervised state-of-the-arts on LA dataset under 1/10 and 1/5 partition protocols. Results of previous methods are from [58]. Best and second-best results are emphasized in **bold red** and **black**. [†]: method with ensemble strategy. **Our method achieves competitive results with state-of-the-arts.**

method	Dice	Jaccard	95HD	ASD
upper bound	91.14	83.82	5.75	1.52
protocol: 1/10				
baseline	79.99	68.12	21.11	5.48
DAP [59]	81.89	71.23	15.81	3.80
UA-MT [51]	84.25	73.48	13.84	3.36
SASSNet [60]	87.32	77.72	9.62	**2.55**
LG-ER-MT [61]	85.54	75.12	13.29	3.77
DUWM [62]	85.91	75.75	12.67	3.31
DTC [63]	86.57	76.55	14.47	3.74
MC-Net [58]	87.71	78.31	**9.36**	**2.18**
TriMix	**88.46**	**79.59**	9.91	2.94
TriMix[†]	**88.89**	**80.23**	**9.38**	2.69
protocol: 1/5				
baseline	86.03	76.06	14.26	3.51
DAP [59]	87.89	78.72	9.29	2.74
UA-MT [51]	88.88	80.21	7.32	2.26
SASSNet [60]	89.54	81.24	8.24	2.20
LG-ER-MT [61]	89.62	81.31	7.16	2.06
DUWM [62]	89.65	81.35	**7.04**	**2.03**
DTC [63]	89.42	80.98	7.32	2.10
MC-Net [58]	**90.34**	**82.48**	**6.00**	**1.77**
TriMix	89.55	81.22	7.76	2.33
TriMix[†]	**89.90**	**81.78**	7.22	2.07
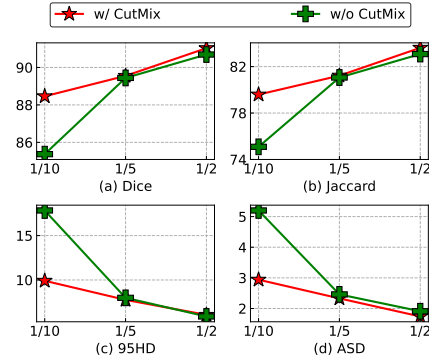
Fig. 5. Ablation study on effectiveness of CutMix augmentation under 1/10, 1/5, and 1/2 partition protocols using LA dataset with four metrics. Without CutMix augmentation, TriMix would degenerate into a tri-training-like scheme, where consistency is only regularized under network diversity. Data augmentation introduced by CutMix blends with network diversity to form a more challenging perturbation. TriMix w/ CutMix achieves better performance than TriMix w/o CutMix across all partition protocols, indicating that **consistency regularization under a stricter perturbation improves model generalization.**

Appendix C: Additional Experiments

C.1: Semi-Supervised 3D Segmentation of Left Atrium

Data and evaluation metrics. We used the **Left atrium (LA) dataset**[6], which contains 100 3D gadolinium-enhanced MRI scans and LA ground truth. All volumes were cropped at the center of the heart region, and intensities were normalized as zero mean and unit variance. We split 100 scans into 80 for training and 20 for validation. We applied Dice, Jaccard, 95% Hausdorff Distance (95HD), and average surface distance (ASD) as evaluation metrics, following previous works [51, 60, 63, 58].

Implementation details. The training setting is identical to previous methods [51, 60, 63, 58]. Concretely, we adopted V-Net [56] as backbone architecture and trained TriMix 6000 iterations using the SGD optimizer with a weight decay

---

[6] http://atriaseg2018.cardiacatlas.org/

of 0.0001 and a momentum of 0.9. We set the initial learning rate as 0.01 and decayed it by a factor of 10 every 2500 iterations. At each training iteration, 2 labeled and 2 unlabeled samples with size $112 \times 112 \times 80$ were fetched. Data augmentation, including flipping and rotation, was applied. The cropped volume ratio of 3D CutMix was empirically set to 0.2.

Experiment results. We compared TriMix under 1/10 and 1/5 partition protocols with seven state-of-the-arts: DAP [59], UA-MT [51], SASSNet [60], LG-ER-MT [61], DUWM [62], DTC [63], and MC-Net [58]. Evaluation results are illustrated in Table 7, where the upper bound indicates the fully-supervised result with all 80 labeled volumes, and the baseline is the result obtained only using the partially labeled volumes. This dataset is a competitive benchmark since the work of [51]. Existing methods successfully improve the baseline by a large margin and reach comparable accuracy to the upper bound result. We note that TriMix achieves competitive results with existing methods. Under the 1/10 partition protocol, TriMix outperforms the second-best method MC-Net [58] by +0.75% in Dice and +1.28% in Jaccard.

Effectiveness of CutMix augmentation. We analyzed the impact of Cut-Mix augmentation within TriMix on the LA dataset under 1/10, 1/5, and 1/2 partition protocols. Ablation results are illustrated in Fig. 5.

C.2: Semi-Supervised 2D Segmentation of Cardiac Structures

Data and evaluation metrics. We applied TriMix to semi-supervised 2D segmentation of cardiac structures on the ACDC dataset [66]. Identical to the work [42], we performed 5-fold cross-validation and adopted the 1/10 partition protocol. For data prepossessing, all slices were resized to 256×256 pixels, and their intensity was normalized to [0,1]. We will report results in Dice and 95HD.

Implementation details. The 2D U-Net architecture [67] was utilized as the backbone. The cropped area ratio of the CutMix augmentation was set to 0.2. We used SGD with a weight decay of 0.0001 and a momentum of 0.9 to optimize TriMix for a total of 60000 iterations under a poly learning rate with

Table 8. Comparison with other semi-supervised methods on ACDC dataset. Other average (standard deviation) results are from [42]. [†]: method with ensemble strategy.

method	RV		Myo		LV		avg	
	Dice	95HD	Dice	95HD	Dice	95HD	Dice	95HD
upper bound	88.2 (9.5)	6.9 (10.8)	88.3 (4.2)	5.9 (15.2)	93.0 (7.4)	8.1 (20.9)	89.8	7.0
baseline	65.9 (26.1)	26.8 (30.4)	72.4 (17.6)	16.0 (21.6)	79.0 (20.5)	24.5 (30.4)	72.4	22.5
DAN [64]	63.9 (26.0)	20.6 (21.4)	76.4 (14.4)	9.4 (12.4)	82.5 (18.6)	15.9 (20.8)	74.3	15.3
AdvEnt [65]	61.5 (29.6)	20.2 (19.4)	76.0 (15.1)	8.5 (8.3)	84.8 (15.9)	11.7 (18.1)	74.1	13.5
MT [50]	65.3 (27.1)	18.6 (22.0)	78.5 (11.8)	11.4 (17.0)	84.6 (15.3)	19.0 (26.7)	76.1	16.3
UA-MT [51]	66.0 (26.7)	22.3 (22.9)	77.3 (12.9)	10.3 (14.8)	84.7 (15.7)	17.1 (23.9)	76.0	16.6
TriMix	85.6 (4.4)	8.2 (2.4)	86.9 (2.0)	3.8 (1.2)	92.3 (3.6)	5.3 (2.0)	88.2	5.8
TriMix[†]	**86.1 (4.3)**	**7.8 (2.7)**	**87.3 (2.0)**	**3.7 (1.3)**	**92.5 (3.6)**	**5.0 (2.2)**	**88.6**	**5.5**

an initial value of 0.01. The batch size was set to 12 (6 labeled and 6 unlabeled samples were fetched at each iteration.)

Experiment results. The validation results are shown in Table 8, where we compare TriMix with other methods: Deep Adversarial Network (DAN) [64], Adversarial Entropy Minimization (AdvEnt) [65], MT [50], and UA-MT [51]. The backbone U-Net trained with the partitioned labeled data obtained the baseline performance. In addition, we treated the result yielded with the fully labeled data as the upper bound accuracy. We note that TriMix significantly improves the baseline by +15.8% in Dice and -16.7 in 95HD. Besides, TriMix outperforms other existing methods by a large margin, obtaining the closest accuracy to the upper bound performance.
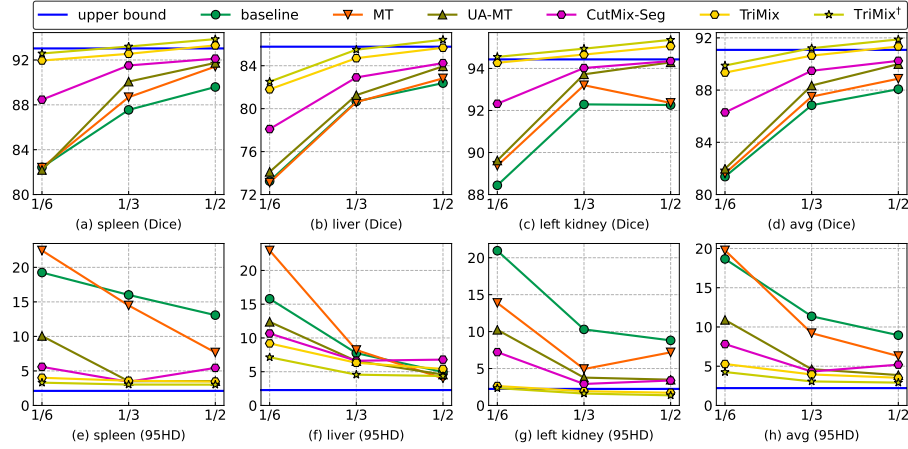
C.3: Semi-Supervised 2D Segmentation of Abdominal Organs

Data and evaluation metrics. We additionally investigated the application of TriMix to 2D segmentation with **BTCV dataset**[7] that provides 50 abdomen CT volumes, 30 of which are training sets containing labels, and the left 20 are testing sets without ground truth, and each volume consists of 85 to 198 slices with the size of 512×512 pixels. The work of [68] has extended the ground truth to 47 volumes. We utilized the 47 volumes for method evaluation, and we focused on three organs (spleen, left kidney, and liver) segmentation in this study. Specifically, 47 volumes were split into 30 for training and 17 for validation. 1/6, 1/3, and 1/2 volumes were further sampled as the labeled data, and the remaining volumes were treated as unlabeled data. All 3D volumes were downsampled by a factor of 0.5 on each axis, and intensities were normalized. Besides, the 3D volumes were cut into 2D axial slices to fulfill the 2D training strategy. The 2D predictions were stacked back to 3D volumes for evaluation in the test phase. We applied Dice and 95HD as metrics.

Implementation details. We applied 2D U-Net [67] as the backbone. During training, an SGD optimizer was adopted to update our model, with a momentum of 0.9 and weight decay of 0.0001. The batch size was set to 8, *i.e.*, eight labeled samples and eight unlabeled samples were fetched at each iteration. Following a poly learning rate policy, the learning rate was set to 0.01 multiplied by $(1 - \frac{k}{K})^{0.9}$ at each iteration, where k is the current iteration, and K is the predefined maximum iteration. TriMix was trained 100 epochs. The cropped area ratio of CutMix was empirically set to 0.2.

Experiment results. We re-implemented several consistency-based methods and compared them with ours. These methods were Mean Teacher (MT) [50], uncertainty-aware Mean Teacher (UA-MT) [51], and CutMix-seg [53]. Quantitative results of the BTCV dataset are shown in Fig. 6. We can note that TriMix substantially outperforms the baseline with gains of 7.9%, 3.8%, and 3.2% in Dice, and reductions of 13.4, 7.5, and 5.4 in 95HD, under the 1/6, 1/3, and 1/2 partition protocols, respectively. Besides, under 1/3 and 1/2 partition protocols,

---

[7] https://www.synapse.org/#!Synapse:syn3193805/wiki/89480

Fig. 6. Application of TriMix to semi-supervised 2D segmentation on BTCV dataset under different partition protocols: 1/6, 1/3 and 1/2. Resutlts are in Dice and 95HD. [†]: method with ensemble strategy. **Overall, TriMix consistently improves baseline and outperforms other existing methods across all partition protocols. TriMix obtains closer results to upper bound and sometimes apparently surpasses it.**

TriMix obtains closer results to the upper bound accuracy and sometimes apparently surpasses it. TriMix outperforms other methods by a large margin, especially with the fewest labeled samples (1/6 partition protocol). For instance, our method outperforms the second-best CutMix-Seg by about 3.0% in Dice under this setting. This validation shows that TriMix can be applied to 2D segmentation and achieves competitive results.

Appendix D: Visualization Results

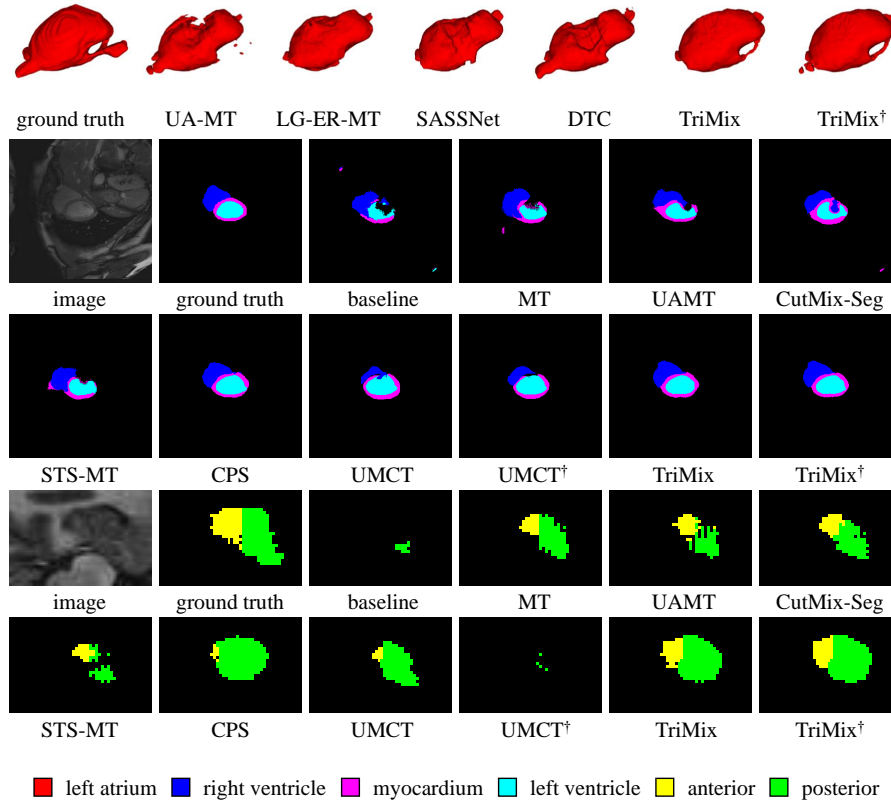### D.1: Visualization Results of Semi-Supervised Segmentation

The visualization comparison of semi-supervised 3D and 2D segmentation is shown in Fig. 7 and Fig. 8. For the hippocampus case, the performance gap among triple networks in UMCT is large, and thus the ensemble strategy worsens the result. We can note that TriMix obtains closer results to the ground truth than previous methods on all datasets.

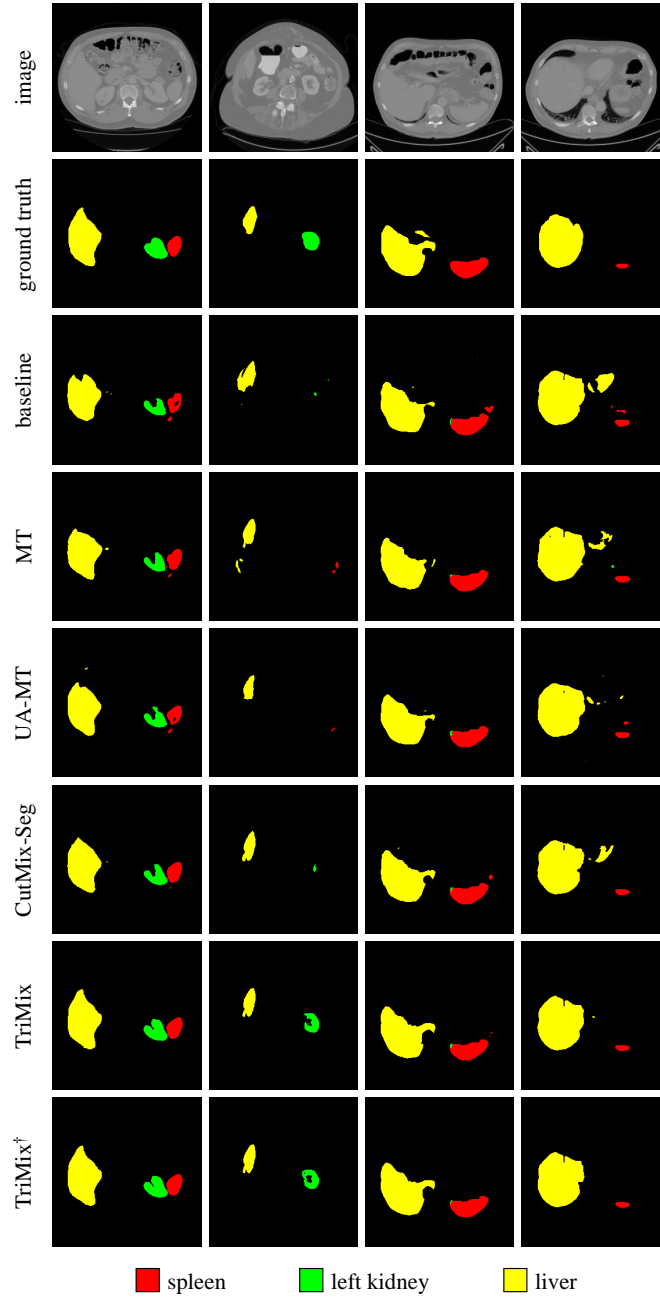### D.2: Visualization Results of Scribble-Supervised Segmentation

We show some visualization examples of the ablation study on different loss combinations in Fig. 9, as supplementary material for Fig. 4. We can note that

$L_{pce}^{unmix} + L_{ce}^{mix}$ achieves better segmentation than L_{pce}^{unmix} and $L_{pce}^{unmix} + L_{pce}^{mix}$, and $L_{pce}^{unmix} + L_{pce}^{mix} + L_{ce}^{mix}$ realizes the best performance.

We also show some visualization examples of good, medium, and worse segmentation cases obtained by our method on the ACDC and MSCMRseg datasets in Fig. 10.


Fig. 7. Visualization results of semi-supervised 3D segmentation, from top to bottom: left atrium segmentation, cardiac multi-structure segmentation, and hippocampus segmentation. [†] indicates method with ensemble strategy. **TriMix obtains closer results to ground truth than other methods.**
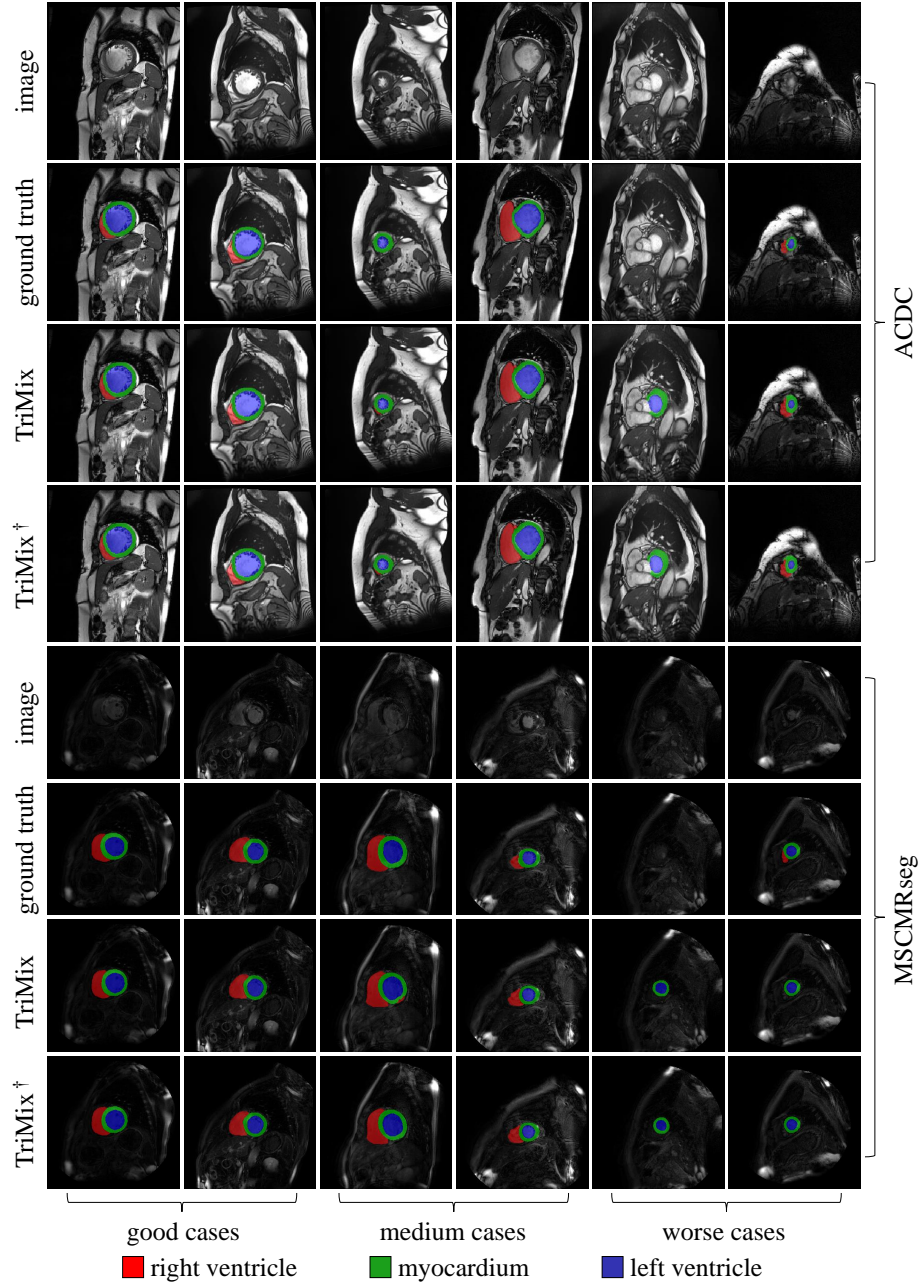
Fig. 8. Visualization results of semi-supervised 2D segmentation on BTCV dataset. † indicates method with ensemble strategy. **TriMix obtains closer results to ground truth than other methods.**
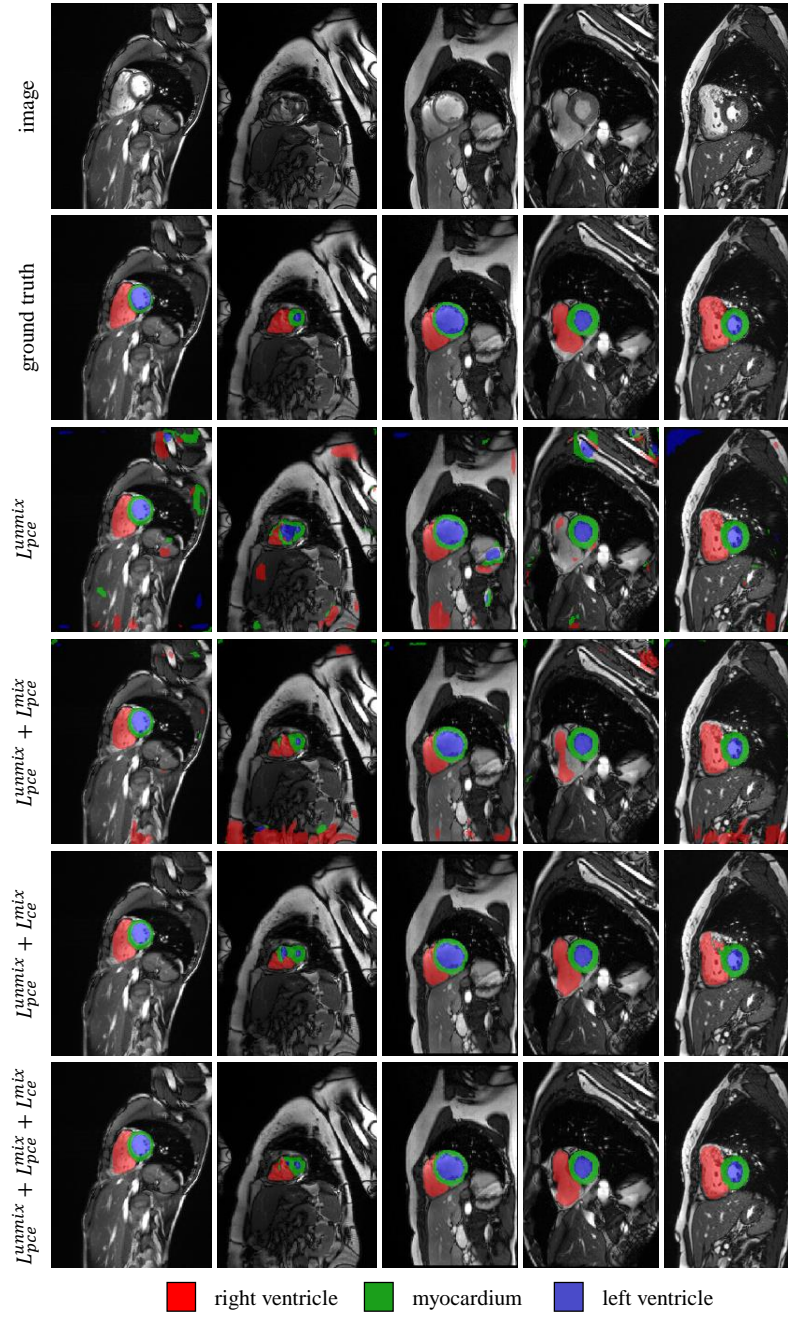
Fig. 9. Visualization results of ablation study on different loss combinations using ACDC dataset. $L_{pce}^{unmix} + L_{ce}^{mix}$ achieves better segmentation than L_{pce}^{unmix} and $L_{pce}^{unmix} + L_{pce}^{mix}$. $L_{pce}^{unmix} + L_{pce}^{mix} + L_{ce}^{mix}$ realizes the best performance.

Fig. 10. Visualization results of good, medium, and worse segmentation cases on ACDC and MSCMRseg datasets. [†] indicates method with ensemble strategy.

Algorithm 1 Training scheme of TriMix for semi-supervised segmentation.

Input: networks f_1, f_2, and f_3 with different weights $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$, labeled dataset D_l, unlabeled dataset D_u, and total training epoch T

Output: networks f_1, f_2, and f_3 with the updated weights $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$

1: **while** T is not reached **do**
2:     Fetch mini-batch $\{\mathbf{x}_l, \mathbf{y}_l\} \in D_l$ and mini-batch $\{\mathbf{x}_u\} \in D_u$
        # Step 1: first forward pass.
3:     **for** $i = 1, 2, 3$ **do**
4:         $\mathbf{p}_{l_i} = f_i(\mathbf{x}_l, \mathbf{w}_i)$
5:         $\mathbf{p}_{u_i} = f_i(\mathbf{x}_u, \mathbf{w}_i)$
6:     **end for**
        # Step 2: mix augmentation.
7:     **for** $i = 1, 2, 3$ **do**
8:         $\{\tilde{\mathbf{x}}_{u_i}, \tilde{\mathbf{p}}_{u_i}\} = $ random-shuffle-mini-batch $\{\mathbf{x}_u, \mathbf{p}_{u_i}\}$
9:     **end for**
10:     $\bar{\mathbf{x}}_{u_1} = \mathrm{Mix}(\tilde{\mathbf{x}}_{u_2}, \tilde{\mathbf{x}}_{u_3})$, $\hat{\mathbf{y}}_{u_1} = \mathrm{Mix}(\tilde{\mathbf{p}}_{u_2}, \tilde{\mathbf{p}}_{u_3})$
11:     $\bar{\mathbf{x}}_{u_2} = \mathrm{Mix}(\bar{\mathbf{x}}_{u_1}, \bar{\mathbf{x}}_{u_3})$, $\hat{\mathbf{y}}_{u_2} = \mathrm{Mix}(\tilde{\mathbf{p}}_{u_1}, \tilde{\mathbf{p}}_{u_3})$
12:     $\bar{\mathbf{x}}_{u_3} = \mathrm{Mix}(\tilde{\mathbf{x}}_{u_1}, \tilde{\mathbf{x}}_{u_2})$, $\hat{\mathbf{y}}_{u_3} = \mathrm{Mix}(\tilde{\mathbf{p}}_{u_1}, \tilde{\mathbf{p}}_{u_2})$
13:     **for** $i = 1, 2, 3$ **do**
14:         $\hat{\mathbf{y}}_{u_i} = \mathrm{argmax}\,(\hat{\mathbf{y}}_{u_i})$ # hard labels for pseudo supervision consistency.
15:     **end for**
        # Step 3: second forward pass.
16:     **for** $i = 1, 2, 3$ **do**
17:         $\bar{\mathbf{p}}_{u_i} = f_i(\bar{\mathbf{x}}_{u_i}, \mathbf{w}_i)$
18:     **end for**
        # loss calculation and update network parameters.
19:     **for** $i = 1, 2, 3$ **do**
20:         $L_i = L_{dice}(\mathbf{p}_{l_i}, \mathbf{y}_l) + \lambda L_{dice}(\bar{\mathbf{p}}_{u_i}, \hat{\mathbf{y}}_{u_i})$
21:         $\mathbf{w}_i = \mathrm{SGD}(L_i, \mathbf{w}_i)$
22:     **end for**
23: **end while**

Algorithm 2 Training scheme of TriMix for scribble-supervised segmentation.

Input: networks f_1, f_2, and f_3 with different weights $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$, dataset D_s
    with scribble annotations, and total training epoch T

Output: networks f_1, f_2, and f_3 with the updated weights $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$

1: **while** T is not reached **do**

2:    Fetch mini-batch $\{\mathbf{x}_s, \mathbf{y}_s\} \in D_s$

       # Step 1: first forward pass.

3:    **for** $i = 1, 2, 3$ **do**

4:        $\mathbf{p}_{s_i} = f_i(\mathbf{x}_s, \mathbf{w}_i)$

5:    **end for**

       # Step 2: mix augmentation.

6:    **for** $i = 1, 2, 3$ **do**

7:        $(\tilde{\mathbf{x}}_{s_i}, \tilde{\mathbf{y}}_{s_i}, \tilde{\mathbf{p}}_{s_i}) = $ random-shuffle-mini-batch $(\mathbf{x}_s, \mathbf{y}_s, \mathbf{p}_{s_i})$

8:    **end for**

9:    $\bar{\mathbf{x}}_{s_1} = \mathrm{Mix}(\tilde{\mathbf{x}}_{s_2}, \tilde{\mathbf{x}}_{s_3})$, $\bar{\mathbf{y}}_{s_1} = \mathrm{Mix}(\tilde{\mathbf{y}}_{s_2}, \tilde{\mathbf{y}}_{s_3})$, $\hat{\mathbf{y}}_{s_1} = \mathrm{Mix}(\tilde{\mathbf{p}}_{s_2}, \tilde{\mathbf{p}}_{s_3})$

10:    $\bar{\mathbf{x}}_{s_2} = \mathrm{Mix}(\tilde{\mathbf{x}}_{s_1}, \tilde{\mathbf{x}}_{s_3})$, $\bar{\mathbf{y}}_{s_2} = \mathrm{Mix}(\tilde{\mathbf{y}}_{s_1}, \tilde{\mathbf{y}}_{s_3})$, $\hat{\mathbf{y}}_{s_2} = \mathrm{Mix}(\tilde{\mathbf{p}}_{s_1}, \tilde{\mathbf{p}}_{s_3})$

11:    $\bar{\mathbf{x}}_{s_3} = \mathrm{Mix}(\tilde{\mathbf{x}}_{s_1}, \tilde{\mathbf{x}}_{s_2})$, $\bar{\mathbf{y}}_{s_3} = \mathrm{Mix}(\tilde{\mathbf{y}}_{s_1}, \tilde{\mathbf{y}}_{s_2})$, $\hat{\mathbf{y}}_{s_3} = \mathrm{Mix}(\tilde{\mathbf{p}}_{s_1}, \tilde{\mathbf{p}}_{s_2})$

12:    **for** $i = 1, 2, 3$ **do**

13:        $\hat{\mathbf{y}}_{s_i} = \mathrm{argmax}\,(\hat{\mathbf{y}}_{s_i})$ # hard labels for pseudo supervision consistency.

14:    **end for**

       # Step 3: second forward pass.

15:    **for** $i = 1, 2, 3$ **do**

16:        $\bar{\mathbf{p}}_{s_i} = f_i(\bar{\mathbf{x}}_{s_i}, \mathbf{w}_i)$

17:    **end for**

       # loss calculation and update network parameters.

18:    **for** $i = 1, 2, 3$ **do**

19:        $L_i = L_{pce}^{unmix}(\mathbf{p}_{s_i}, \mathbf{y}_s) + \lambda_1 L_{pce}^{mix}(\bar{\mathbf{p}}_{s_i}, \bar{\mathbf{y}}_{s_i}) + \lambda_2 L_{ce}^{mix}(\bar{\mathbf{p}}_{s_i} . \hat{\mathbf{y}}_{s_i})$

20:        $\mathbf{w}_i = \mathrm{SGD}(L_i, \mathbf{w}_i)$

21:    **end for**

22: **end while**

References

1. Papandreou, G., Chen, L.C., Murphy, K.P., Yuille, A.L.: Weakly- and semi-supervised learning of a deep convolutional network for semantic image segmentation. In: *ICCV.* (2015)
2. Lee, J., Kim, E., Yoon, S.: Anti-adversarially manipulated attributions for weakly and semi-supervised semantic segmentation. In: *CVPR.* (2021)
3. Kolesnikov, A., Lampert, C.H.: Seed, expand and constrain: Three principles for weakly-supervised image segmentation. In Leibe, B., Matas, J., Sebe, N., Welling, M., eds.: *ECCV.* (2016)
4. Li, J., Fan, J., Zhang, Z.: Towards noiseless object contours for weakly supervised semantic segmentation. In: *CVPR.* (2022)
5. Khoreva, A., Benenson, R., Hosang, J., Hein, M., Schiele, B.: Simple does it: Weakly supervised instance and semantic segmentation. In: *CVPR.* (2017)
6. Pan, J., Zhu, P., Zhang, K., Cao, B., Wang, Y., Zhang, D., Han, J., Hu, Q.: Learning self-supervised low-rank network for single-stage weakly and semi-supervised semantic segmentation. *IJCV* **130** (2022)
7. Roy, A., Todorovic, S.: Combining bottom-up, top-down, and smoothness cues for weakly supervised image segmentation. In: *CVPR.* (2017)
8. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In: *CVPR.* (2016)
9. Wei, Y., Feng, J., Liang, X., Cheng, M.M., Zhao, Y., Yan, S.: Object region mining with adversarial erasing: A simple classification to semantic segmentation approach. In: *CVPR.* (2017)
10. Vernaza, P., Chandraker, M.: Learning random-walk label propagation for weakly-supervised semantic segmentation. In: *CVPR.* (2017)
11. Kim, D., Cho, D., Yoo, D., So Kweon, I.: Two-phase learning for weakly supervised object localization. In: *ICCV.* (2017)
12. Tang, M., Perazzi, F., Djelouah, A., Ben Ayed, I., Schroers, C., Boykov, Y.: On regularized losses for weakly-supervised cnn segmentation. In: *ECCV.* (2018)
13. Chaudhry, A., Dokania, P., Torr, P.: Discovering class-specific pixels for weakly-supervised semantic segmentation. In: *BMVC.* (2017)
14. Tang, M., Djelouah, A., Perazzi, F., Boykov, Y., Schroers, C.: Normalized cut loss for weakly-supervised cnn segmentation. In: *CVPR.* (2018)
15. Ge, W., Yang, S., Yu, Y.: Multi-evidence filtering and fusion for multi-label classification, object detection and semantic segmentation based on weakly supervised learning. In: *CVPR.* (2018)
16. Wang, B., Qi, G., Tang, S., Zhang, T., Wei, Y., Li, L., Zhang, Y.: Boundary perception guidance: A scribble-supervised semantic segmentation approach. In: *IJCAI.* (2019)
17. Li, K., Wu, Z., Peng, K.C., Ernst, J., Fu, Y.: Tell me where to look: Guided attention inference network. In: *CVPR.* (2018)
18. Marin, D., Tang, M., Ayed, I.B., Boykov, Y.: Beyond gradient descent for regularized segmentation losses. In: *CVPR.* (2019)
19. Wang, X., You, S., Li, X., Ma, H.: Weakly-supervised semantic segmentation by iteratively mining common object features. In: *CVPR.* (2018)
20. Ji, Z., Shen, Y., Ma, C., Gao, M.: Scribble-based hierarchical weakly supervised learning for brain tumor segmentation. In: *MICCAI.* (2019)
21. Ahn, J., Kwak, S.: Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In: *CVPR.* (2018)

22. Lee, H., Jeong, W.K.: Scribble2label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: *MICCAI.* (2020)
23. Huang, Z., Wang, X., Wang, J., Liu, W., Wang, J.: Weakly-supervised semantic segmentation network with deep seeded region growing. In: *CVPR.* (2018)
24. Valvano, G., Leo, A., Tsaftaris, S.A.: Learning to segment from scribbles using multi-scale adversarial attention gates. *TMI* **40** (2021)
25. Wei, Y., Xiao, H., Shi, H., Jie, Z., Feng, J., Huang, T.S.: Revisiting dilated convolution: A simple approach for weakly- and semi-supervised semantic segmentation. In: *CVPR.* (2018)
26. Zhang, B., Xiao, J., Jiao, J., Wei, Y., Zhao, Y.: Affinity attention graph neural network for weakly supervised semantic segmentation. *TPAMI* (2021)
27. Lee, J., Kim, E., Lee, S., Lee, J., Yoon, S.: Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference. In: *CVPR.* (2019)
28. Lu, W., Gong, D., Fu, K., Sun, X., Diao, W., Liu, L.: Boundarymix: Generating pseudo-training images for improving segmentation with scribble annotations. *PR* **117** (2021)
29. Song, C., Huang, Y., Ouyang, W., Wang, L.: Box-driven class-wise region masking and filling rate guided loss for weakly supervised semantic segmentation. In: *CVPR.* (2019)
30. Chen, H., Wang, J., Chen, H.C., Zhen, X., Zheng, F., Ji, R., Shao, L.: Seminar learning for click-level weakly supervised semantic segmentation. In: *ICCV.* (2021)
31. Ahn, J., Cho, S., Kwak, S.: Weakly supervised learning of instance segmentation with inter-pixel relations. In: *CVPR.* (2019)
32. Xu, J., Zhou, C., Cui, Z., Xu, C., Huang, Y., Shen, P., Li, S., Yang, J.: Scribble-supervised semantic segmentation inference. In: *ICCV.* (2021)
33. Li, K., Zhang, Y., Li, K., Li, Y., Fu, Y.: Attention bridging network for knowledge transfer. In: *ICCV.* (2019)
34. Pan, Z., Jiang, P., Wang, Y., Tu, C., Cohn, A.G.: Scribble-supervised semantic segmentation by uncertainty reduction on neural representation and self-supervision on neural eigenspace. In: *ICCV.* (2021)
35. Shimoda, W., Yanai, K.: Self-supervised difference detection for weakly-supervised semantic segmentation. In: *ICCV.* (2019)
36. Liang, Z., Wang, T., Zhang, X., Sun, J., Shen, J.: Tree energy loss: Towards sparsely annotated semantic segmentation. In: *CVPR.* (2022)
37. Zhang, T., Lin, G., Liu, W., Cai, J., Kot, A.: Splitting vs. merging: Mining object regions with discrepancy and intersection loss for weakly supervised semantic segmentation. In Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., eds.: *ECCV.* (2020)
38. Zhang, K., Zhuang, X.: Cyclemix: A holistic strategy for medical image segmentation from scribble supervision. In: *CVPR.* (2022)
39. Sun, G., Wang, W., Dai, J., Van Gool, L.: Mining cross-image semantics for weakly supervised semantic segmentation. In: *ECCV.* (2020)
40. Unal, O., Dai, D., Van Gool, L.: Scribble-supervised lidar semantic segmentation. In: *CVPR.* (2022)
41. Chen, L., Wu, W., Fu, C., Han, X., Zhang, Y.: Weakly supervised semantic segmentation with boundary exploration. In: *ECCV.* (2020)
42. Luo, X., Hu, M., Liao, W., Zhai, S., Song, T., Wang, G., Zhang, S.: Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In: *MICCAI.* (2022)

43. Wang, Y., Zhang, J., Kan, M., Shan, S., Chen, X.: Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In: *CVPR*. (2020)
44. Zhang, K., Zhuang, X.: Shapepu: A new pu learning framework regularized by global consistency for scribble supervised cardiac segmentation. In: *MICCAI*. (2022)
45. Chang, Y.T., Wang, Q., Hung, W.C., Piramuthu, R., Tsai, Y.H., Yang, M.H.: Weakly-supervised semantic segmentation via sub-category exploration. In: *CVPR*. (2020)
46. Liu, X., Yuan, Q., Gao, Y., He, K., Wang, S., Tang, X., Tang, J., Shen, D.: Weakly supervised segmentation of covid19 infection with scribble annotation on ct images. *PR* **122** (2022)
47. Zhang, D., Zhang, H., Tang, J., Hua, X.S., Sun, Q.: Causal intervention for weakly-supervised semantic segmentation. In: *NeurIPS*. (2020)
48. Gao, F., Hu, M., Zhong, M.E., Feng, S., Tian, X., Meng, X., yi-di-li Ni-jia ti, M., Huang, Z., Lv, M., Song, T., Zhang, X., Zou, X., Wu, X.: Segmentation only uses sparse annotations: Unified weakly and semi-supervised learning in medical images. *MedIA* **80** (2022)
49. Lee, J., Yi, J., Shin, C., Yoon, S.: Bbam: Bounding box attribution map for weakly supervised semantic and instance segmentation. In: *CVPR*. (2021)
50. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: *NeurIPS*. (2017)
51. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: *MICCAI*. (2019)
52. Huang, T., Sun, Y., Wang, X., Yao, H., Zhang, C.: Spatial ensemble: a novel model smoothing mechanism for student-teacher framework. In: *NeurIPS*. (2021)
53. French, G., Laine, S., Aila, T., Mackiewicz, M., Finlayson, G.: Semi-supervised semantic segmentation needs strong, varied perturbations. In: *BMVC*. (2020)
54. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: *CVPR*. (2021)
55. Xia, Y., Yang, D., Yu, Z., Liu, F., Cai, J., Yu, L., Zhu, Z., Xu, D., Yuille, A., Roth, H.: Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. *MedIA* **65** (2020)
56. Milletari, F., Navab, N., Ahmadi, S.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: *3DV*. (2016)
57. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: cutmix: Regularization strategy to train strong classifiers with localizable features. In: *ICCV*. (2019)
58. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: *MICCAI*. (2021)
59. Zheng, H., Lin, L., Hu, H., Zhang, Q., Chen, Q., Iwamoto, Y., Han, X., Chen, Y.W., Tong, R., Wu, J.: Semi-supervised segmentation of liver using adversarial learning with deep atlas prior. In: *MICCAI*. (2019)
60. Li, S., Zhang, C., He, X.: Shape-aware semi-supervised 3d semantic segmentation for medical images. In: *MICCAI*. (2020)
61. Hang, W., Feng, W., Liang, S., Yu, L., Wang, Q., Choi, K.S., Qin, J.: Local and global structure-aware entropy regularized mean teacher model for 3d left atrium segmentation. In: *MICCAI*. (2020)
62. Wang, Y., Zhang, Y., Tian, J., Zhong, C., Shi, Z., Zhang, Y., He, Z.: Double-uncertainty weighted method for semi-supervised learning. In: *MICCAI*. (2020)

63. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: *AAAI*. (2021)
64. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: *MICCAI*. (2017)
65. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: *CVPR*. (2019)
66. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M.A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V.A., Krishnamurthi, G., Rohé, M.M., Pennec, X., Sermesant, M., Isensee, F., Jäger, P., Maier-Hein, K.H., Full, P.M., Wolf, I., Engelhardt, S., Baumgartner, C.F., Koch, L.M., Wolterink, J.M., Išgum, I., Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., Jodoin, P.M.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *TMI* **37** (2018)
67. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: *MICCAI*. (2015)
68. Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., Davidson, B., Pereira, S.P., Clarkson, M.J., Barratt, D.C.: Automatic multi-organ segmentation on abdominal CT with dense V-Networks. *TMI* **37** (2018)