Paper ID #4494 GuidedNet: Semi-Supervised Multi-Organ Segmentation via Labeled Data Guide Unlabeled Data

ACMMultimedia2024
Melbourne, Australia

北京航空航天大學
BEIHANG UNIVERSITY

Haochen Zhao, Hui Meng, Deqian Yang, Xiaozheng Xie, Xiaoze Wu, Qingfeng Li*, Jianwei Niu*

Paper Code WeChat

Abstract 1

Semi-supervised multi-organ medical image segmentation aids physicians in improving disease diagnosis and treatment planning and reduces the time and effort required for organ annotation. Existing state-of-the-art methods train the labeled data with ground truths and train the unlabeled data with pseudo-labels. However, the two training flows are separate, which does not reflect the interrelationship between labeled and unlabeled data. To address this issue, we propose a semi-supervised multi-organ segmentation method called GuidedNet, which leverages the knowledge from labeled data to guide the training of unlabeled data. The primary goals of this study are to improve the quality of pseudo-labels for unlabeled data and to enhance the network's learning capability for both small and complex organs. A key concept is that voxel features from labeled and unlabeled data that are close to each other in the feature space are more likely to belong to the same class. On this basis, a 3D Consistent Gaussian Mixture Model (3D-CGMM) is designed to leverage the feature distributions from labeled data to rectify the generated pseudo-labels. Furthermore, we introduce a Knowledge Transfer Cross Pseudo Supervision (KT-CPS) strategy, which leverages the prior knowledge obtained from the labeled data to guide the training of the unlabeled data, thereby improving the segmentation accuracy for both small and complex organs. Extensive experiments on two public datasets, FLARE22 and AMOS, demonstrated that GuidedNet is capable of achieving state-of-the-art performance.

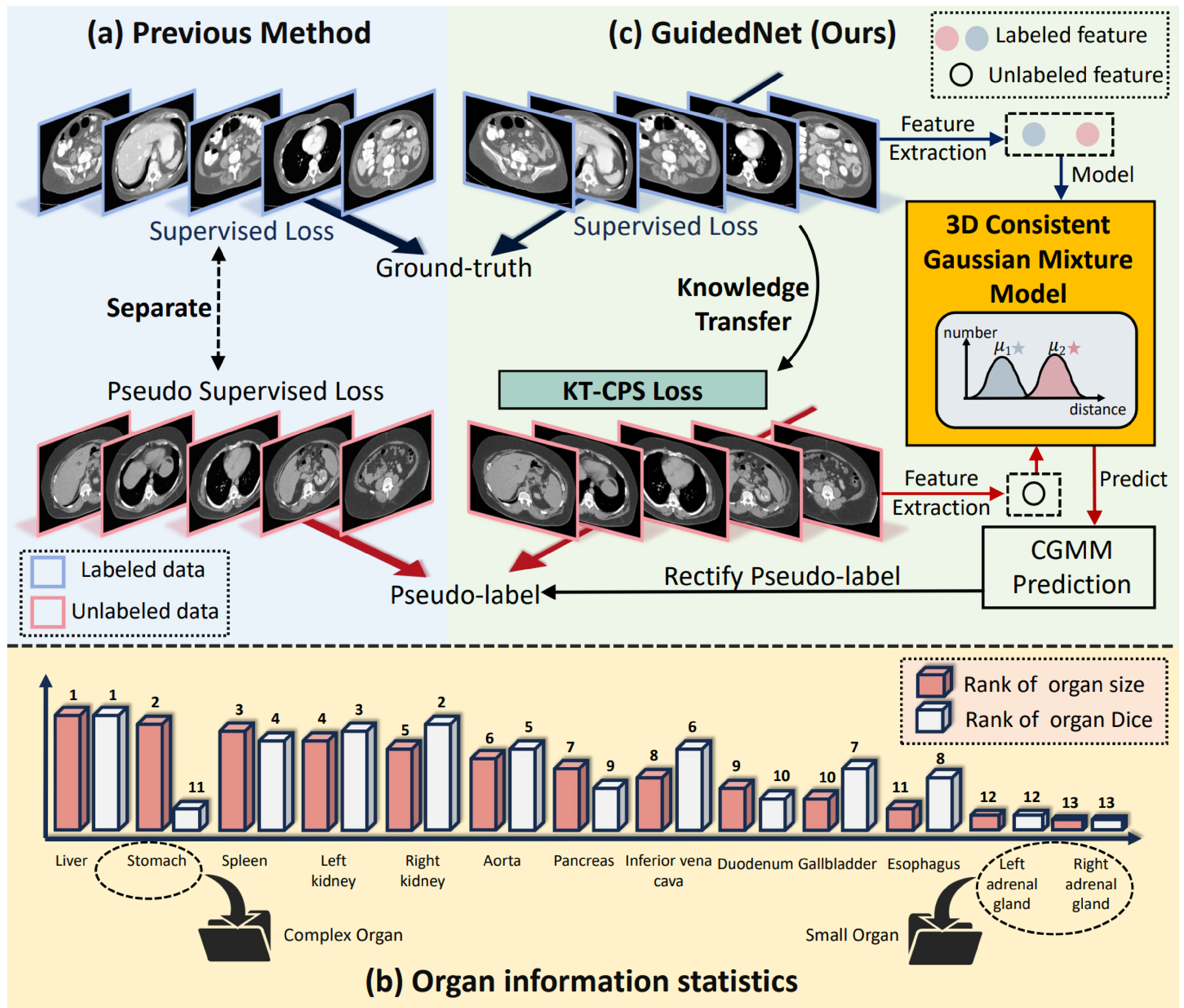


Figure 1. (a) Previously developed pseudo-labeling methods separate the labeled and unlabeled data training flows, which does not reflect the interrelationship between labeled and unlabeled data (*i.e.*, CPS, ARCO, UCMT). (b) In the FLARE22 dataset, rankings are provided for both the sizes of all foreground organs and the Dice. (c) Our GuidedNet, which is comprised of two components: 3D-CGMM and KT-CPS.

Motivation 2

- **Traditional methods employed to train the labeled and unlabeled data in these methods are separated, as depicted in Fig. 1(a).** This separation causes a significant drawback: the quality of the pseudo-labels generated during training depends exclusively on the unlabeled data and the segmentation performance of the network, with no regard for the interrelationship between the labeled and unlabeled data.
- Widely used SSL multi-organ medical image segmentation methods focus on enhancing the learning ability of small organs due to the prevalent issue of class imbalance in medical image datasets. **Specifically, the inherent complexity of organ segmentation must not be overlooked.** For instance, the stomach, which is the second largest organ in abdominal multi-organ dataset, ranks only eleventh in terms of segmentation Dice as shown in Fig.1(b). This discrepancy highlights that the stomach's substantial size does not mitigate the segmentation difficulties posed by its complexity.

Key Concept

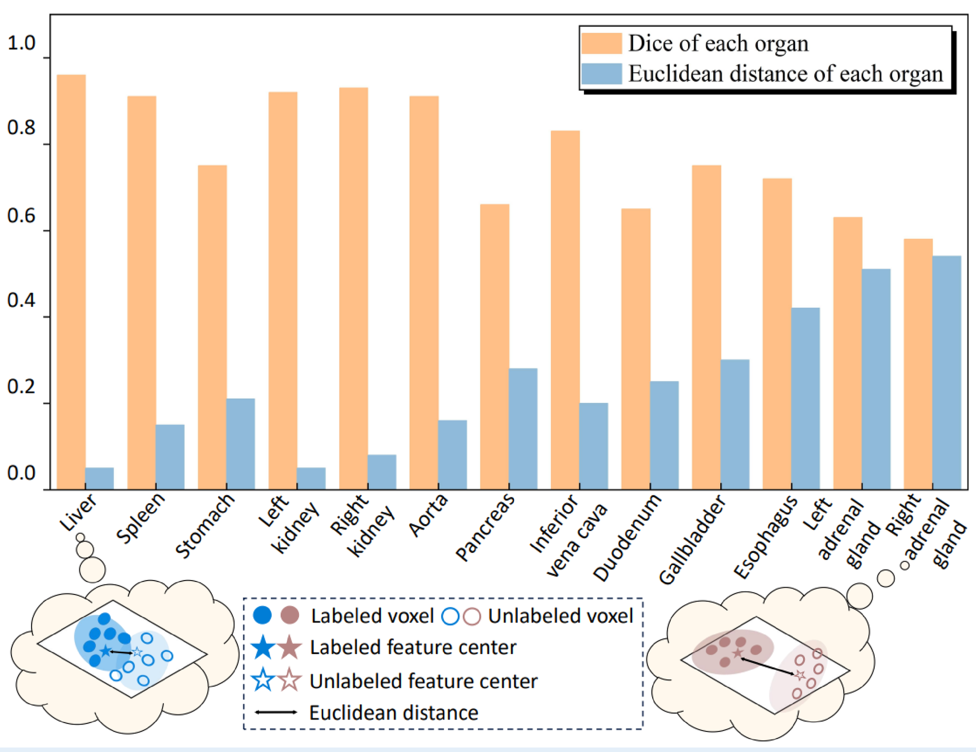


Figure 2. Category-wise performance for the FLARE22 dataset. The orange bars represent the Dice for each organ, and the blue bars denote the Euclidean distances between the feature centers of the labeled and unlabeled data for each organ.

Conclusions 3

- We propose to leverage the knowledge from labeled data to rectify the generated pseudo-labels by using a **3D-CGMM** that utilizes the feature distribution of the labeled data to generate CGMM predictions to guide the learning of unlabeled data.
- We design a **KT-CPS strategy**, which guides the training of unlabeled data to learn prior knowledge from the labeled data and enhances the learning ability of small and complex organs.
- Extensive experiments have been conducted to validate the effectiveness of our proposed GuidedNet. The results of these experiments have demonstrated solid performance gains across two public datasets.

Formulation 4

- **For the k_{th} Gaussian mixture:**

$$\mu_k = \frac{1}{|O_k|}\sum_{\forall x \in O_k} f(x) \qquad \sigma_k = \sqrt{\frac{1}{|P_k|}\sum_{\forall x \in O_k} P_k\left(f(x)-\mu_k\right)^2}$$

- **Each voxel x_{ij} follows the probability density of the k_{th} Gaussian mixture, which is calculated using the Gaussian probability density function:**

$$\mathcal{N}(x_{ij} \mid \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_k}\exp\left(-\frac{1}{2}\left(x_{ij}-\mu_k\right)^2\sigma_k^{-2}\right)$$

- **Then, following Bayes' rule, the posterior is derived as**

$$P(x_{ij} \mid k) = \frac{p^{tra}(k)\cdot\mathcal{N}(x_{ij}|\mu_k\Sigma_{k'})}{\sum_{k'=1}^{K} p^{tra}(k')\cdot\mathcal{N}(x_{ij}|\mu_{k'}\Sigma_{k'})}$$

- **Each voxel G_{ij} in the CGMM prediction is then produced using:**

$$G_{ij} = \mathrm{argmax}_k P(x_{ij} \mid k)$$

Training Loss 5

- **Training Loss:**

$$\mathcal{L}_{gt} = -\frac{1}{|\mathcal{B}_l|}\sum_{i=1}^{\mathcal{B}_l} y_i \log(G_i) \qquad \mathcal{L}_{self} = -\frac{1}{|\mathcal{B}_l|}\sum_{i=1}^{\mathcal{B}_l}\left[G_i * \log(P_i) + (1-G_i) * \log(1-P_i)\right]$$

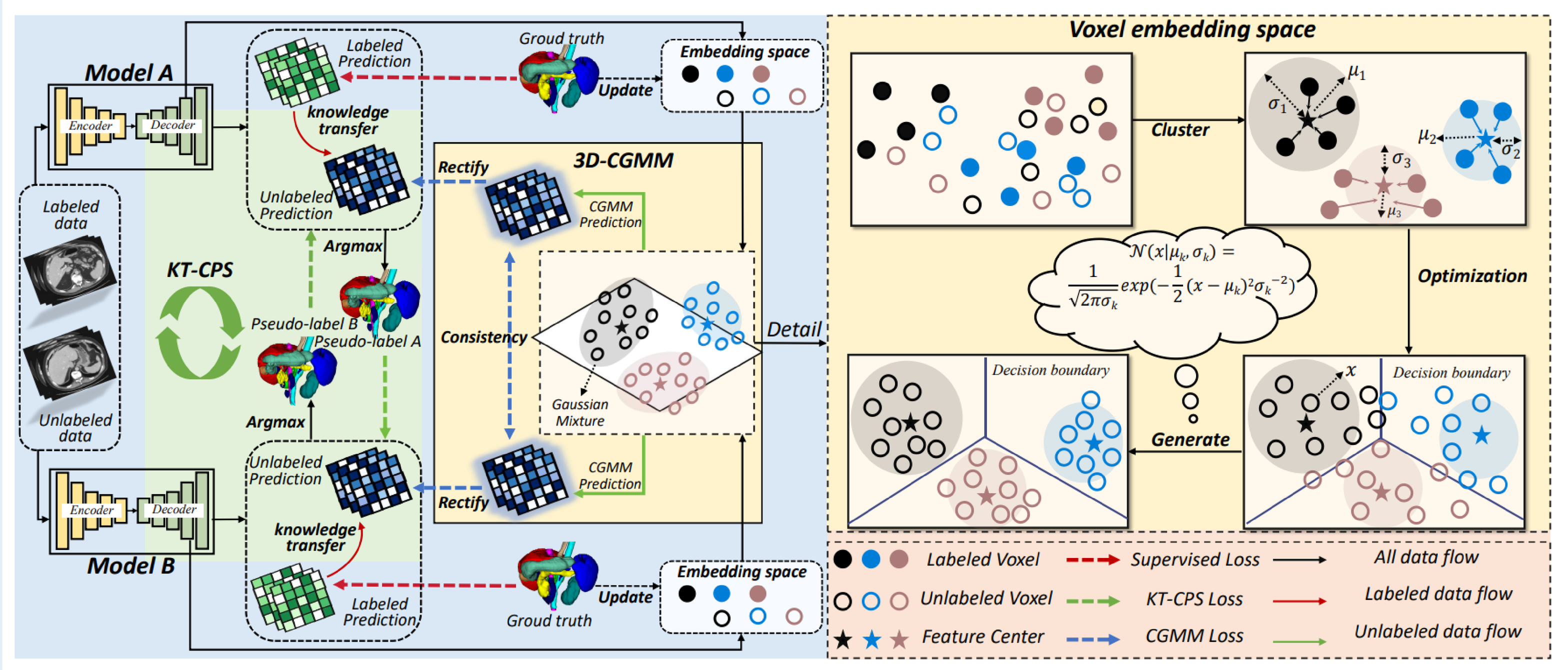
$$\mathcal{L}_{max} = \frac{2}{K(K-1)}\sum_{\forall k,v \in K, k \neq v} e^{-(\mu_k-\mu_v)^2} \qquad \mathcal{L}_{cons} = \frac{1}{|\mathcal{B}_u|}\sum_{i=1}^{\mathcal{B}_u}\left(\hat{G}_i^A - \hat{G}_i^B\right)^2$$

- **Rectify Pseudo-labels:**

$$\mathcal{L}_{rectify} = \frac{1}{|\mathcal{B}_u|}\sum_{i=1}^{\mathcal{B}_u}\left[\mathcal{L}_{ce}(p_i^A, \hat{G}_i^A) + \mathcal{L}_{ce}(p_i^B, \hat{G}_i^B)\right]$$

Overall Architecture 6

Figure 3. The workflow of GuidedNet involves processing input data from model A and model B to yield predictions. The feature distributions of the labeled predictions are utilized to train the 3D-CGMM, and the generated CGMM predictions are used to rectify the initial pseudo-labels. The prior knowledge obtained from the labeled predictions are transferred to the unlabeled predictions using the KT-CPS strategy for cross pseudo supervised training.

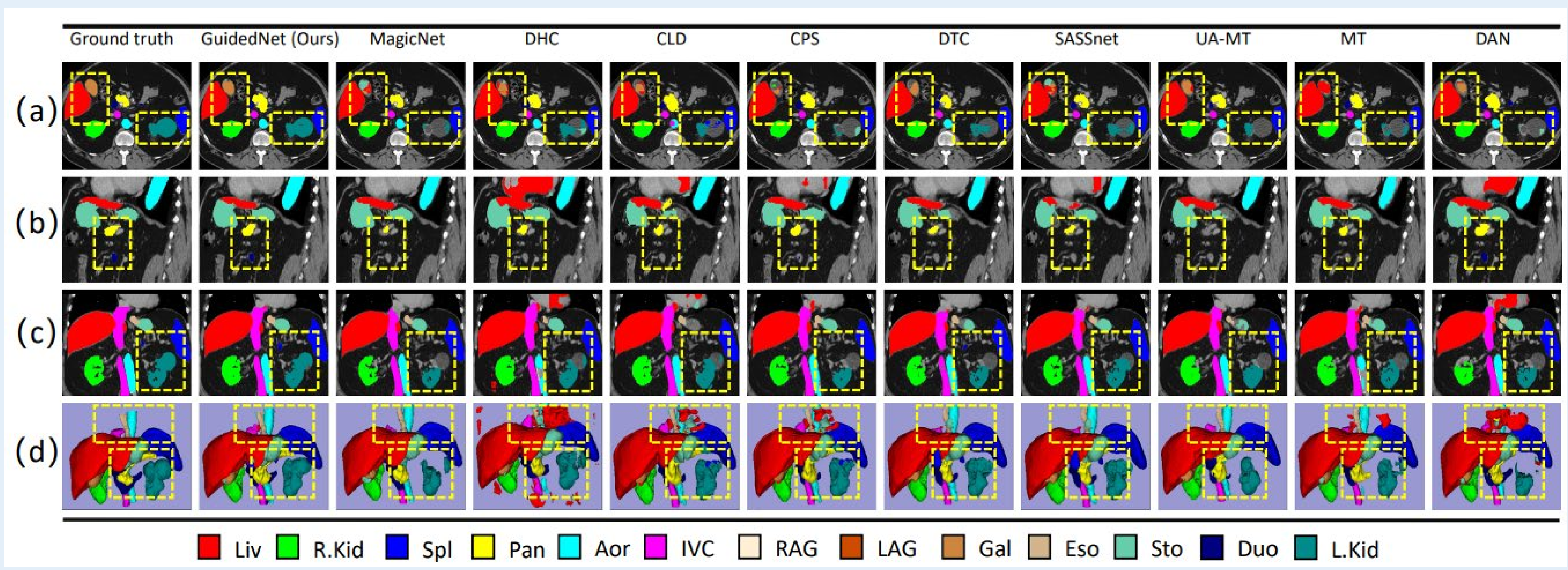


Experiments 7

- **Comparison to SOTA Methods:**

Methods	Labeled/Unlabeled	Liv	Sto	Spl	L.kid	R.kid	Aor	Bla	IVC	Pan	Duo	P/U	Gal	Eso	RAG	LAG	Mean Dice	Mean Jaccard
		Mean Dice for each organ																
Sup only	18/0(100%)	85.99	40.08	82.19	79.57	80.24	75.51	11.63	59.86	26.80	20.22	22.40	15.06	37.05	32.46	9.68	45.24 ± 0.65	39.06 ± 0.42
DAN [48] [MICCAI'17]	18/162(10%)	86.54	50.00	83.98	86.97	85.76	85.46	54.81	67.26	48.97	43.84	52.23	33.08	40.28	28.00	18.09	57.80 ± 0.88	47.38 ± 0.89
MT [35] [Neurips'17]	18/162(10%)	89.26	56.87	84.27	84.42	85.98	85.57	51.22	70.13	48.91	48.04	39.46	42.17	50.71	43.92	30.46	61.44 ± 1.28	51.00 ± 1.10
UA-MT [43] [MICCAI'19]	18/162(10%)	88.25	52.49	86.39	86.34	87.73	86.14	68.22	70.76	47.19	42.79	49.54	32.49	52.87	43.98	37.76	61.73 ± 1.12	50.97 ± 0.90
SASSnet [18] [MICCAI'20]	18/162(10%)	**90.33**	48.61	86.87	<u>87.66</u>	**88.17**	<u>87.09</u>	43.55	73.85	50.29	48.56	8.38	36.15	43.18	41.36	28.85	58.35 ± 1.42	51.15 ± 0.83
DTC [23] [AAAI'21]	18/162(10%)	<u>89.81</u>	50.49	<u>87.48</u>	85.20	85.84	85.83	64.49	72.72	43.44	47.36	39.19	38.62	50.56	42.53	37.02	60.81 ± 1.27	50.84 ± 1.24
CPS [3] [CVPR'21]	18/162(10%)	88.52	55.52	83.25	86.30	<u>87.97</u>	85.36	60.53	71.71	50.11	46.05	60.33	37.95	52.37	46.33	37.48	63.52 ± 0.36	51.82 ± 0.49
CLD [20] [MICCAI'22]	18/162(10%)	88.43	<u>63.71</u>	84.90	85.85	86.07	85.16	64.15	<u>75.56</u>	55.21	49.67	60.62	39.47	<u>56.71</u>	<u>50.91</u>	40.56	<u>65.81 ± 1.24</u>	54.00 ± 1.69
DHC [37] [MICCAI'23]	18/162(10%)	83.27	63.39	83.60	84.11	85.66	84.40	**74.52**	74.88	<u>56.02</u>	<u>51.89</u>	<u>65.47</u>	47.53	43.21	48.28	42.59	65.17 ± 1.47	52.46 ± 1.30
MagicNet [2] [CVPR'23]	18/162(10%)	88.99	61.20	83.52	**88.39**	87.24	83.69	62.47	74.83	54.11	51.18	54.62	**56.69**	55.68	46.87	<u>43.16</u>	65.31 ± 1.31	<u>54.89 ± 0.78</u>
GuidedNet (Ours)	18/162(10%)	89.08	**66.44**	**87.50**	85.86	87.25	**87.93**	<u>70.65</u>	**76.32**	**58.38**	**55.55**	**67.68**	<u>48.95</u>	**59.87**	**54.11**	**43.40**	**69.19 ± 0.17**	**56.97 ± 0.15**
Sup only	90/0(100%)	89.25	55.60	84.23	87.40	88.58	87.32	53.49	73.71	48.56	48.21	52.68	38.43	50.27	38.48	30.30	61.29 ± 1.74	51.62 ± 1.35
DAN [48] [MICCAI'17]	90/90(50%)	90.49	55.91	89.63	90.08	88.74	86.71	47.44	72.09	54.98	50.33	53.04	39.13	58.34	29.57	6.49	61.39 ± 1.16	52.06 ± 1.45
MT [35] [Neurips'17]	90/90(50%)	<u>92.08</u>	62.02	89.83	90.23	89.24	89.12	63.05	<u>78.11</u>	53.46	52.85	40.93	51.63	59.64	45.41	37.35	66.17 ± 0.75	57.06 ± 1.00
UA-MT [43] [MICCAI'19]	90/90(50%)	90.86	58.55	88.92	88.93	88.83	88.49	54.86	74.28	51.88	54.54	44.73	40.99	58.58	51.31	41.78	65.48 ± 0.80	55.62 ± 1.10
SASSnet [18] [MICCAI'20]	90/90(50%)	91.65	53.00	**91.54**	89.61	89.72	88.50	50.43	74.87	46.34	52.48	55.92	37.93	60.57	45.62	39.17	63.77 ± 1.13	54.68 ± 0.55
DTC [23] [AAAI'21]	90/90(50%)	91.25	56.49	90.68	88.88	89.30	<u>89.16</u>	<u>67.37</u>	76.50	48.13	54.67	54.23	41.88	62.49	47.67	42.91	66.93 ± 1.78	55.92 ± 1.78
CPS [3] [CVPR'21]	90/90(50%)	90.94	61.90	89.97	90.25	89.67	88.77	65.03	75.27	52.34	45.15	54.76	42.87	62.44	49.96	47.74	66.65 ± 1.24	56.56 ± 0.54
CLD [20] [MICCAI'22]	90/90(50%)	91.23	66.18	89.34	89.50	<u>89.86</u>	88.85	66.40	76.97	55.63	53.35	58.82	45.78	<u>62.93</u>	<u>54.24</u>	43.79	<u>69.09 ± 1.14</u>	57.99 ± 1.14
DHC [37] [MICCAI'23]	90/90(50%)	86.68	58.39	86.62	85.57	87.48	87.28	67.04	74.38	<u>60.88</u>	<u>56.91</u>	<u>58.87</u>	53.75	54.14	51.59	<u>51.03</u>	68.60 ± 0.56	56.05 ± 0.51
MagicNet [2] [CVPR'23]	90/90(50%)	91.69	<u>66.33</u>	88.59	<u>90.28</u>	89.64	86.80	61.80	74.39	59.94	52.88	57.28	**58.83**	59.53	52.74	42.35	68.94 ± 0.56	<u>58.33 ± 0.52</u>
GuidedNet (Ours)	90/90(50%)	**92.32**	**72.99**	<u>91.21</u>	**90.82**	**89.87**	**89.31**	**74.00**	**78.41**	**61.75**	**58.26**	**65.23**	<u>56.72</u>	**66.62**	**54.71**	**51.10**	**72.94 ± 0.09**	**61.53 ± 0.12**

- **Qualitative Results:**



- **Ablation Studies:**

Baseline	KT-CPS	3D-CGMM	Liv	Spl	Sto	L.kid	R.kid	Aor	Pan	IVC	Duo	Gal	Eso	RAG	LAG	Mean Dice	Mean Jaccard
			Mean Dice for each organ														
✔			<u>96.92</u>	91.86	77.02	92.70	92.71	92.25	69.39	81.91	65.94	75.12	72.78	63.56	58.96	79.32 ± 0.46	68.14 ± 0.61
✔	✔		96.71	<u>93.12</u>	<u>81.53</u>	<u>94.47</u>	92.96	<u>92.62</u>	73.53	83.98	<u>70.43</u>	75.61	72.61	66.50	<u>62.53</u>	81.28 ± 0.24	70.45 ± 0.60
✔		✔	**97.15**	91.12	77.26	94.33	**93.87**	92.15	<u>74.48</u>	**84.81**	69.79	<u>84.60</u>	<u>75.23</u>	<u>69.24</u>	61.89	<u>82.00 ± 0.16</u>	<u>71.31 ± 0.16</u>
✔	✔	✔	96.77	**93.48**	**83.19**	**94.51**	<u>93.48</u>	**92.95**	**75.97**	<u>84.63</u>	**71.92**	**85.87**	**75.74**	**71.10**	**67.77**	**83.64 ± 0.42**	**73.08 ± 0.38**

$\mathcal{L}_{gt}$	$\mathcal{L}_{self}$	$\mathcal{L}_{max}$	$\mathcal{L}_{cons}$	Mean Dice	Mean Jaccard
✔				81.29 ± 0.95	70.76 ± 1.11
✔	✔			83.08 ± 0.17	72.48 ± 0.23
✔		✔		83.02 ± 0.21	72.29 ± 0.19
✔	✔	✔		<u>83.28 ± 0.26</u>	<u>72.58 ± 0.46</u>
✔	✔	✔	✔	**83.64 ± 0.42**	**73.08 ± 0.38**