Decoupled Training for Semi-supervised Medical Image Segmentation with Worst-Case-Aware Learning

Ankit Das*[1] (✉), Chandan Gautam*[2,3], Hisham Cholakkal[4], Pritee Agrawal, Feng Yang[1], Ramasamy Savitha[2,3], and Yong Liu[1]

[1] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore
{dasak,yangf,liuyong}@ihpc.a-star.edu.sg, priteeagrawal2006@gmail.com
[2] Institute for Infocomm Research (I2R), A*STAR, Singapore
{gautamc,ramasamysa}@i2r.a-star.edu.sg
[3] International Research Laboratory on Artificial Intelligence (IPAL), CNRS@CREATE, Singapore
[4] Mohamed bin Zayed University of Artificial Intelligence (MBZUAI)
hisham.cholakkal@mbzuai.ac.ae

Abstract. While semi-supervised learning (SSL) has demonstrated remarkable success in natural image segmentation, tackling medical image segmentation with limited annotated data remains a highly relevant and challenging research problem. Many existing approaches rely on a shared network for learning from both labeled and unlabeled data, facing difficulties in fully exploiting labeled data due to interference from unreliable pseudo-labels. Additionally, they suffer from degradation in model quality resulting from training with unreliable pseudo-labels. To address these challenges, we propose a novel training strategy that uses two distinct decoders—one for labeled data and another for unlabeled data. This decoupling enhances the model's ability to fully leverage the knowledge embedded within the labeled data. Moreover, we introduce an additional decoder, referred to as the "worst-case-aware decoder," which indirectly assesses potential worst case scenario that might emerge from pseudo-label training. We employ adversarial training of the encoder to learn features aimed at avoiding this worst case scenario. Our experimental results on three medical image segmentation datasets demonstrate that our method shows improvements in range of 5.6% - 28.10% (in terms of dice score) compared to the state-of-the-art techniques. The source code is available at https://github.com/thesupermanreturns/decoupled.

Keywords: SSL · Decoupled Training · Medical Imaging.

1 Introduction

Accurate segmentation of medical images is vital for various clinical applications [18,14,3]. While supervised semantic segmentation methods have demon-

* Joint first author.

strated effectiveness in numerous segmentation tasks [25,28], their reliance on a substantial number of labeled samples poses challenges. Acquiring an adequate amount of labeled data proves to be expensive and arduous, especially within the medical domain. Semi-supervised learning (SSL) [10,13,7] emerges as one of the most practical methods for training models with limited annotated data by providing a mechanism to leverage unlabeled data, consequently diminishing the demand for labeled data. The enhanced performance offered by SSL generally comes at a low cost, given that unlabeled data can be used with minimal human involvement. Consequently, numerous SSL techniques have been developed for the segmentation task [6,9,27] in the last decade. Notably, in recent years, semi-supervised segmentation for medical images (Med-SemSeg) has garnered significant attention from the research community [15,2,24,5], primarily owing to the restricted availability of annotated data in the medical domain.

In the Med-SemSeg literature, the predominant methods often rely on pseudo-labeling and consistency-based regularization (CBR) [11,16,24,23]. These existing methods have demonstrated the ability to harness knowledge from unlabeled data for learning by using a shared network between the labeled and unlabeled data. For example, a CBR-based mean-teacher network: where a network is shared between the teacher and student. Similarly, a typical pseudo-labeled-based approach [19] which also uses a shared network for labeled and unlabeled data. These shared network-based approaches iteratively improve the model's performance by using its predictions on the unlabeled data as pseudo-labels. While these approaches have shown encouraging outcomes, the performance of these shared networks suffers from the inherent unreliability of pseudo-labels, introducing two significant issues:

Prob.1: *Training a shared network with both labeled (true labels) and unlabeled (unreliable pseudo-labels) data interferes with the model's capability to fully exploit the true labels.*

Prob.2: *Training models using unreliable pseudo-labels can degrade model quality, particularly through iterative training with these pseudo-labels.*

The first issue is addressed by introducing a decoupled training approach. It involves the utilization of a common encoder in combination with two specialized decoders for labeled and unlabeled data, ensuring a clear distinction between the tasks of pseudo-label generation and application. This decoupling of decoders facilitates the network in retaining valuable information from labeled data using one decoder, enabling it to generate high-quality pseudo-labels.

To address the second issue of model degradation from training with unreliable pseudo-labels, one way is to compute the error from pseudo-labels and adjust the model to nullify its effect. However, directly measuring this error is not feasible due to the absence of ground truth data for the unlabeled samples. The nature of unreliable pseudo-labels (due to absence of ground truth) may cause the model to effectively learn only from the labeled data while making errors on the unlabeled data. This is the worst case scenario for the model which cannot learn new information from the unlabeled data. As a solution, we introduce a worst-case decoder $\mathcal{D}^w$, that optimizes the encoder to generate features

to avoid this worst case, helping improve the model quality and performance on unlabeled data. Overall, the contributions are summarized as follows:

- We address two challenges in shared training approaches for Med-SemSeg: (i) partial exploitation of labeled samples due to a shared network (addressed by decoupled training), and (ii) Decreased model quality resulting from training with unreliable pseudo-labels (addressed by improved feature generation via introducing a worst-case decoder).
- To validate the efficacy of the proposed method, we performed experiments on three publicly available medical imaging datasets which showcase significant improvements compared to the state-of-the-art (SOTA) methods.

2 Proposed Method

In semi-supervised medical image semantic segmentation, the goal is to achieve generalization from a combined dataset consisting of both pixel-wise labeled images $\mathcal{X}_l = \left\{(x_l^b, y_l^b) : b \in (1, \ldots, B_l)\right\}$ and unlabeled images $\mathcal{X}_u = \left\{x_u^b : b \in (1, \ldots, B_u)\right\}$, where x_l^b denotes training labeled samples and y_l^b represents its corresponding ground truth masks, x_u^b denotes unlabeled samples, B_l and B_u denote the number of batches of labeled and unlabeled data, respectively and $B_l \ll B_u$. The overall objective of semi-supervised medical image semantic segmentation can be expressed as a combination of losses on labeled ($\mathcal{L}^l$) and unlabeled ($\mathcal{L}^u$) data as follows:

$$\mathcal{L} = \mathcal{L}^l + \lambda \mathcal{L}^u, \tag{1}$$

where λ is the trade-off parameter between losses on labeled and unlabeled data. In the realm of semi-supervised methods, $\mathcal{L}^u$ plays a pivotal role in differentiating these methods. Conversely, $\mathcal{L}^l$ typically computes cross-entropy loss between the predicted and ground truth mask.

2.1 Method

The overall architecture of the proposed method is illustrated in Fig. 1. It comprises an encoder and three decoders: (i) the main decoder ($\mathcal{D}^l$), (ii) the strong augmented decoder ($\mathcal{D}^s$), and (iii) the worst-case-aware decoder ($\mathcal{D}^w$). It processes batches containing both labeled and unlabeled data to generate segmentations for medical images. The main decoder ($\mathcal{D}^l$) is trained solely on labeled samples using a supervised loss $\mathcal{L}_{sup}^l$ and is responsible for generating pseudo-labels. The remaining decoders ($\mathcal{D}^s$ and $\mathcal{D}^w$) address challenges related to the partial exploitation of labeled data and the model quality resulting from unreliable pseudo-label training. The strong augmented decoder ($\mathcal{D}^s$) is trained with a weak-to-strong augmentation and consumes the pseudo-labels. Meanwhile, the worst-case-aware decoder ($\mathcal{D}^w$) engages in a min-max game to estimate the worst possible scenario of pseudo-labeling and generates features to avoid this situation. It should be noted that ($\mathcal{D}^l$) is used during inference.

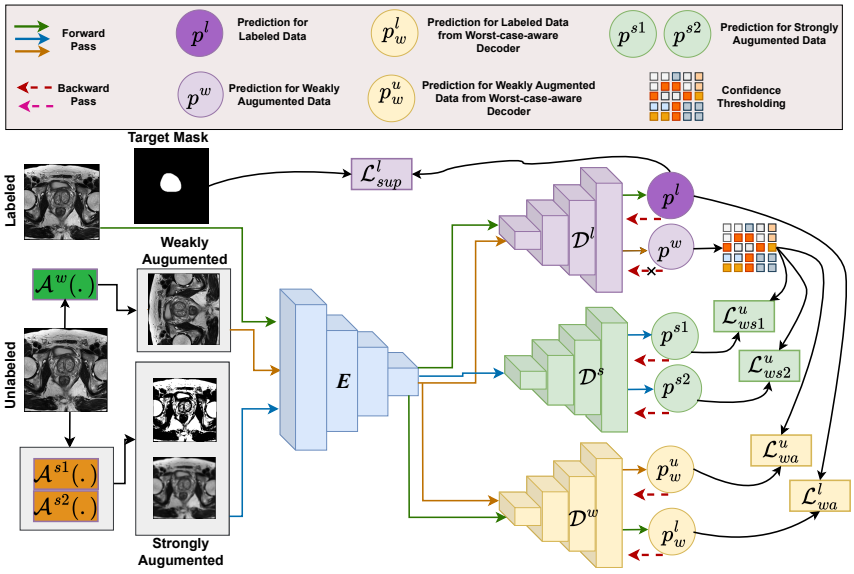

Fig. 1: Illustration of our proposed method.

2.2 Decoupled Training

The effectiveness of shared network-based Med-SemSeg methods is affected by the use of unreliable pseudo-labels, leading to a deterioration in model performance and a further decline in the quality of pseudo-labels. To address this issue, we propose decoupling the generation and application of pseudo-labels through two decoders, $\mathcal{D}^l$ and $\mathcal{D}^s$. The main decoder $\mathcal{D}^l$ (Fig. 1) is exclusively optimized using labeled data, avoiding the influence of unreliable pseudo-labels, while providing predictions (pseudo-labels) for unlabeled data. This prevents the main decoder $\mathcal{D}^l$ from being affected by unreliable pseudo-labels. Additionally, we introduce another decoder, $\mathcal{D}^s$, trained using pseudo-labels obtained from $\mathcal{D}^l$. Pseudo-labels generated by $\mathcal{D}^l$ are employed by $\mathcal{D}^s$ using weak-to-strong consistency regularization for enhanced representational learning. Leveraging two strong views and a weak view of unlabeled data, the model's weak view output serves as pseudo-labels for strongly augmented views. This approach enables the model to generate diverse predictions for the same input, fostering robust feature representation through weak-to-strong consistency. Thus, the proposed method effectively decouples the generation and application of pseudo-labels, minimizing the following loss for unlabeled data in decoupled training:

$$\begin{aligned}\mathcal{L}^u_{ws1}(E,\mathcal{D}^s) = &\frac{1}{B_u}\sum_{i=1}^{B_u}\mathbb{1}(\max(p^w)\geq\eta)\mathbb{H}_c(\mathcal{D}^l(E(\mathcal{A}^w(x^i_u))),\\ &\mathcal{D}^s(E(\mathcal{A}^{s1}(x^i_u)))) + \frac{1}{B_u}\sum_{i=1}^{B_u}\mathbb{1}(\max(p^w)\geq\eta)\mathbb{H}_d(\mathcal{D}^l(E(\mathcal{A}^w(x^i_u))),\mathcal{D}^s(E(\mathcal{A}^{s1}(x^i_u)))),\end{aligned} \tag{2}$$

where $\mathbb{H}_c$ denotes entropy minimization between two probabilistic distributions, $\mathbb{H}_d$ denotes standard dice loss, η represents a threshold, $p_i^w = \mathcal{D}^l(E(\mathcal{A}^w(x_u^i))$ denotes class distribution of pseudo-label, $p_i^{s1} = \mathcal{D}^s(E(\mathcal{A}^{s1}(x_u^i)))$ denotes predicted class distribution of i^{th} unlabeled data for the first strong augmentation $\mathcal{A}^{s1}$. Similar to Equation 2, we compute $\mathcal{L}_{ws2}^u(E, \mathcal{D}^s)$ loss using second strong augmentation $\mathcal{A}^{s2}$ instead of $\mathcal{A}^{s1}$ in Equation 2. The final loss for the decoder $\mathcal{D}^s$ is as follows:

$$\mathcal{L}_{ws}(E, \mathcal{D}^s) = \mathcal{L}_{ws1}^u(E, \mathcal{D}^s) + \mathcal{L}_{ws2}^u(E, \mathcal{D}^s) \tag{3}$$

2.3 Worst-case-Aware Learning

To address the risk of model degradation stemming from unreliable pseudo-labels, it is essential to estimate the errors and adjust the model accordingly. However, due to the lack of ground truth for unlabeled data, estimating errors becomes infeasible, posing a challenge for the model to adapt and make necessary adjustments. Consequently, the model may only learn to classify labeled data, potentially leading to misclassification of unlabeled data. This represents the worst-case scenario that can arise from unreliable pseudo-label training. To address this challenge in the absence of a direct solution, we employ adversarial learning to avoid this worst case. We introduce a worst-case-aware decoder, $\mathcal{D}^w$, which acts as an adversary who's objective is to correctly classify labeled data while misclassifying unlabeled data. Employing a min-max adversarial setting between the encoder E and this decoder $\mathcal{D}^w$, compels the encoder E to generate features that evade this worst-case scenario. This ensures that the model avoids the aforementioned situation, thereby enhancing overall model quality.

$$\begin{aligned}\mathcal{L}_{wa}(E, \mathcal{D}^w) =& \mathcal{L}_{wa}^u(E, \mathcal{D}^w) + \mathcal{L}_{wa}^l(E, \mathcal{D}^w) \\ =& \min_E \max_{\mathcal{D}^w} \frac{1}{B_u} \sum_{i=1}^{B_u} \mathbb{1}(\max(p_i^w) \geq \eta) \mathbb{H}_c(p^w, \mathcal{D}^w(E(\mathcal{A}^w(x_u^i)))) - \frac{1}{B_l} \sum_{i=1}^{B_l} \mathbb{H}_c(p_i^l, \mathcal{D}^w(E(x_l^i))),\end{aligned} \tag{4}$$

where $p_{adv,i}^u = \mathcal{D}^w(E(\mathcal{A}^w(x_u^i))))$ and $p_{adv,i}^l = \mathcal{D}^w(E(x_l^i)))$ denote predicted class distribution of i^{th} unlabeled and labeled data by $\mathcal{D}^w$, respectively. The predicted class distribution of labeled data by $\mathcal{D}^l$ is denoted by $p_i^l = \mathcal{D}^l(E(x_l^i))$.

Overall loss function: The optimization loss function of the proposed method is formalized as follows:

$$\min_{E, \mathcal{D}^l, \mathcal{D}^S} \max_{\mathcal{D}^w} \lambda_{sup} \mathcal{L}_{sup}^l(E, \mathcal{D}^l) + \lambda_{ws} \mathcal{L}_{ws}(E, \mathcal{D}^s) + \lambda_{wa} \mathcal{L}_{wa}(E, \mathcal{D}^w), \tag{5}$$

where λ_{sup}, λ_{ws}, and λ_{wa} are the coefficient of different losses; $\mathcal{L}_{sup}^l(E, \mathcal{D}^l)$ minimizes sum of cross-entropy loss and dice loss on the labeled data using encoder E and decoder $\mathcal{D}^l$.

3 Experiments and Results

We evaluate our proposed method on three commonly used datasets for medical semantic segmentation, i.e., ProstateX [1], PROMISE12 [12], and CHAOS [8].

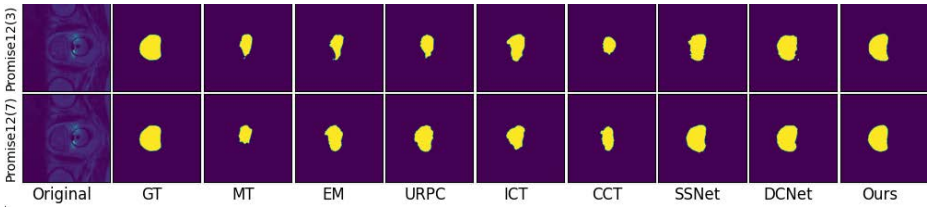

Fig. 2: Visual comparison with different SOTA methods on PROMISE12 dataset for different percentages of labeled data. The visual comparison of ProstateX, and CHAOS datasets are provided in Fig. 1 of supplementary.

3.1 Datasets and Evaluation Metrics

ProstateX. The dataset includes 201 MRI scans with true mask segmentations from the ProstateX Challenge [1]. It is divided into training, validation, and test sets comprising 151, 19, and 31 scans, respectively.

Prostate MR Image Segmentation (PROMISE12). The dataset includes transversal T2-weighted MRI scans of the prostate from 50 patients, with corresponding true mask segmentations [12]. It is split into 35, 5, and 10 cases for training, validation, and testing sets, respectively.

Combined Healthy Abdominal Organ Segmentation (CHAOS). The dataset consists of 20 MRI volumes from the 2019 CHAOS Challenge [8], divided into subsets of 10, 5, and 5 cases for training, validation, and testing, respectively.

Evaluation Metrics. During the inference stage, predictions are produced slice by slice and then stacked into a 3D volume. Our results are presented based on three commonly used evaluation metrics: the Dice Similarity Score (DSC), Hausdorff Distance 95 (HD95), and Average Symmetric Distance (ASD) [4].

3.2 Implementation Details

We implemented the proposed method using PyTorch and conducted experiments on an NVIDIA A40 GPU with 128 GB RAM. The encoder-decoder architecture utilized a standard U-Net [18] backbone, trained from scratch across all datasets. Model convergence was achieved using an ADAM optimizer, running for 600 epochs with a batch size of 12 and a learning rate of 0.01 for all experiments. Consistent weights were maintained for different losses throughout the experiments: $\lambda_{sup} = 1$, $\lambda_{ws} = 0.5$, $\lambda_{wa} = 2$. The ablation of λ_{ws} and λ_{wa} is provided in Table 1 and Table 2 of supplementary, respectively.

Data Augmentation. In our implementation, we utilize five types of data augmentation techniques, namely (1) random rotation, (2) random flipping, (3) color jitter, (4) random Gaussian blur, and (5) CutMix [26]. Random rotation and flipping are applied as weak augmentation methods, while color jitter, random Gaussian blur, and CutMix are utilized as strong augmentation techniques.

3.3 Comparison with State-of-the-Art Methods

In our experiments, we vary the labeled data percentages and compare the performance with state-of-the-art (SOTA) methods, including MT [20], EM [22], URPC [16], ICT [21], CCT [17], SSNet [24] and DCnet [5]. This study enables us to evaluate the impact of varying amounts of labeled data on the model's performance. We present the results of all three datasets with different percentages in Table 1. In this table, the proposed method surpasses all SOTA methods for all three datasets by at least 5.6%, 16.6%, and 8.1% margin in the case of 13%, 30%, and 20% labeled data for ProstateX, CHAOS, and PROMISE12 datasets, respectively. Further, when we experiment with a lesser percentage of the labeled data, then the proposed method outperformed the SOTA methods with even greater extent compared to earlier, i.e., outperformed by at least 11.9%, 28.10%, and 24.3% margin in the case of 3%, 11%, and 8% labeled data for ProstateX, CHAOS, and PROMISE12 datasets, respectively. In the scenario of a smaller number of labeled data, the proposed method performs well, however, most of the SOTA methods that utilize a shared network become even more corrupted by unreliable pseudo-labels. Further, the qualitative assessment of the outcomes of the proposed and (SOTA) methods across various label percentages is visually presented in Fig. 2. Here, it can be observed that our method obtains better predictions on all datasets against the SOTA methods. In both instances of the PROMISE12, specifically with the three cases, it is evident that the other competing methods struggle to perform accurate segmentation. Conversely, our method demonstrates proficiency in successfully segmenting the data.

3.4 Ablation Studies

Impact of different losses. The proposed method integrates five distinct losses: supervised loss ($\mathcal{L}sup$), worst-case-aware losses ($\mathcal{L}wa^l$ and $\mathcal{L}wa^u$), and

Method	Labeled%	ProstateX			Labeled%	Chaos			Labeled%	Promise12		
		DSC	Hd95	ASSD		DSC	Hd95	ASSD		DSC	Hd95	ASSD
FS (UNet)	20(13%)	0.696	12.11	3.77	3(30%)	0.561	51.83	17.37	7(20%)	0.539	26.25	5.57
MT		0.717	11.47	3.71		0.582	44.29	17.58		0.610	36.44	6.27
EM		0.739	7.57	2.68		0.574	49.38	17.98		0.646	14.38	3.93
URPC		0.726	7.32	2.29		0.601	32.81	12.84		0.677	30.32	7.03
ICT		0.755	7.51	2.72		0.634	42.63	13.52		0.684	16.74	4.15
CCT		0.729	6.89	1.99		0.609	43.17	18.62		0.624	16.65	2.06
SSnet		0.760	6.16	1.44		0.666	38.52	12.02		0.730	34.24	10.98
DCNet		0.749	7.98	1.84		0.673	38.23	14.40		0.760	6.48	2.10
Ours		**0.805**	**2.83**	**0.53**		**0.807**	**13.73**	**3.71**		**0.827**	**2.51**	**0.33**
FS (UNet)	5(3%)	0.604	25.4	7.03	1(11%)	0.381	60.27	23.70	3(8%)	0.319	45.23	19.39
MT		0.633	25.8	6.23		0.357	59.63	23.66		0.313	40.86	15.28
EM		0.630	33.7	11.89		0.365	62.38	25.61		0.353	49.28	21.47
URPC		0.614	17.64	3.87		0.385	52.65	17.97		0.377	31.27	9.64
ICT		0.636	33.13	11.63		0.429	65.61	26.72		0.474	37.53	5.86
CCT		0.595	36.26	9.92		0.445	50.92	21.30		0.282	21.59	10.10
SSnet		0.676	15.99	4.22		0.494	42.98	18.18		0.499	25.46	11.87
DCNet		0.678	12.31	3.1		0.477	67.30	25.42		0.599	16.03	6.86
Ours		**0.77**	**3.38**	**0.55**		**0.687**	**19.72**	**7.38**		**0.791**	**3.37**	**0.93**

Table 1: Comparison of the proposed method with SOTA Med-SemSeg methods.

strong-augmentation losses ($\mathcal{L}ws^1$ and $\mathcal{L}ws^2$). In our experimentation, we analyze the impact of including or excluding each loss while keeping $\mathcal{L}sup$ fixed as the base loss. Results in Table 2 show that removing either of the strong augmentations leads to a similar decline in model performance, as expected due to their similar functions. Excluding worst-case-aware losses based on labeled ($\mathcal{L}wa^l$) and unlabeled ($\mathcal{L}wa^u$) data is explored in the third and fourth rows. The absence of $\mathcal{L}^u_{wa}$ has a more significant impact, primarily due to its role in fine-tuning the encoder with unlabeled data to avoid the worst case scenario and generate higher-quality features.

Methods	**3 (8%)**		**7 (20%)**	
	DSC↑	**HD95↓**	**DSC↑**	**HD95↓**
w/o $\mathcal{L}^2_{ws}$	0.776	3.95	0.801	3.10
w/o $\mathcal{L}^1_{ws}$	0.787	3.88	0.803	2.72
w/o $\mathcal{L}^l_{wa}$	0.748	3.71	0.799	3.26
w/o $\mathcal{L}^u_{wa}$	0.756	5.03	0.781	3.58
Ours	**0.791**	**3.37**	**0.827**	**2.51**

Table 2: Ablation on different losses on PROMISE12.

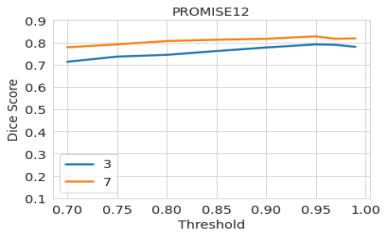

Fig. 3: Ablation on confidence threshold values for PROMISE12.

	3 (8%)		**7 (20%)**	
	DSC↑	**HD95↓**	**DSC↑**	**HD95↓**
w/o $\mathcal{D}^w$	0.748	8.25	0.778	4.13
with $\mathcal{D}^w$	0.791	3.37	0.827	2.51

Table 3: Ablation on PROMISE12 with and without $\mathcal{D}^w$.

# Aug	**3 (8%)**		**7 (20%)**	
	DSC↑	**HD95↓**	**DSC↑**	**HD95↓**
2	0.791	3.37	0.83	2.51
3	0.803	2.95	0.814	2.74
4	0.792	3.16	0.823	2.50

Table 4: Ablation with different augmentations on PROMISE12.

Impact of the worst-case-aware decoder. Table 3 highlights the significant impact of excluding a worst-case-aware decoder on the proposed method's performance. Without this decoder, the encoder-generated features may result in pseudo-labels that deviate from their true labels due to sub-optimal hyperplane.

Impact of threshold. By varying the threshold from 0.7 to 0.99 in Fig. 3, it can be observed that the model consistently achieved the best performance at 0.95 for both the cases of 3 and 7 on the PROMISE12 dataset.

Effectiveness of multiple augmentations. We analyzed our method by increasing the number of strong augmentations on decoder $\mathcal{D}^s$ and presented results in Table 4. Additional augmentations do not consistently improve performance, reaching saturation where they no longer provide new information.

4 Conclusion

In this paper, we recognized the inherent challenges associated with shared network-based approaches and introduced a novel semi-supervised medical image segmentation method to overcome these issues. The proposed method is developed based on decoupled training combined with worst-case-aware learning. Experimental results on three medical imaging datasets demonstrated that our approach attains state-of-the-art performance, highlighting its effectiveness compared to existing methods.

Acknowledgments. This work was supported by the Agency for Science, Technology and Research (A*STAR) through its AME Programmatic Funding Scheme Under Project A20H4b0141. This research is also part of the programme DesCartes and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

Disclosure of Interests. Author Yong Liu has received research grants from Company Agency for Science, Technology and Research (A*STAR), and Author Ramasamy Savitha has received research grants from Company National Research Foundation.

References

1. Armato III, S.G., Huisman, H., Drukker, K., Hadjiiski, L., Kirby, J.S., Petrick, N., Redmond, G., Giger, M.L., Cha, K., Mamonov, A., et al.: Prostatex challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. Journal of Medical Imaging **5**(4), 044501–044501 (2018)
2. Bortsova, G., Dubost, F., Hogeweg, L., Katramados, I., De Bruijne, M.: Semi-supervised medical image segmentation via learning consistency under transformations. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. pp. 810–818. Springer (2019)
3. Cai, H., Li, S., Qi, L., Yu, Q., Shi, Y., Gao, Y.: Orthogonal annotation benefits barely-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3302–3311 (2023)
4. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. Advances in neural information processing systems **33**, 12546–12558 (2020)
5. Chen, F., Fei, J., Chen, Y., Huang, C.: Decoupled consistency for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 551–561. Springer (2023)
6. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2613–2622 (2021)
7. Chen, Y., Tan, X., Zhao, B., Chen, Z., Song, R., Liang, J., Lu, X.: Boosting semi-supervised learning by exploiting all unlabeled data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7548–7557 (2023)

8. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al.: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021)
9. Lai, X., Tian, Z., Jiang, L., Liu, S., Zhao, H., Wang, L., Jia, J.: Semi-supervised semantic segmentation with directional context-aware consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1205–1214 (2021)
10. Li, J., Xiong, C., Hoi, S.C.: Comatch: Semi-supervised learning with contrastive graph regularization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9475–9484 (2021)
11. Li, X., Yu, L., Chen, H., Fu, C.W., Xing, L., Heng, P.A.: Transformation-consistent self-ensembling model for semisupervised medical image segmentation. IEEE Transactions on Neural Networks and Learning Systems **32**(2), 523–534 (2020)
12. Litjens, G., Toth, R., Van De Ven, W., Hoeks, C., Kerkstra, S., Van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al.: Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. Medical image analysis **18**(2), 359–373 (2014)
13. Lucas, T., Weinzaepfel, P., Rogez, G.: Barely-supervised learning: Semi-supervised learning with very few labeled images. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 1881–1889 (2022)
14. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 8801–8809 (2021)
15. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 8801–8809 (2021)
16. Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, Shichuan, D.N.M., Zhang, S.: Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. Medical Image Analysis **80**, 102517 (2022)
17. Ouali, Y., Hudelot, C., Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12674–12684 (2020)
18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
19. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
20. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
21. Verma, V., Lamb, A., Kannala, J., Bengio, Y., Lopez-Paz, D.: Interpolation consistency training for semi-supervised learning. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19. pp. 3635–3641 (2019)
22. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2517–2526 (2019)

23. Wu, H., Wang, Z., Song, Y., Yang, L., Qin, J.: Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11666–11675 (2022)
24. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 34–43. Springer (2022)
25. Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N.: Bisenet: Bilateral segmentation network for real-time semantic segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 325–341 (2018)
26. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6023–6032 (2019)
27. Zhong, Y., Yuan, B., Wu, H., Yuan, Z., Peng, J., Wang, Y.X.: Pixel contrastive-consistent semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7273–7282 (2021)
28. Zhou, T., Wang, W., Konukoglu, E., Van Gool, L.: Rethinking semantic segmentation: A prototype view. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2582–2593 (2022)