MOST: Multi-Formation Soft Masking for Semi-Supervised Medical Image Segmentation

Xinyu Liu[1,*], Zhen Chen[2,*], and Yixuan Yuan[1,✉]

1 Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong SAR
2 Centre for Artificial Intelligence and Robotics (CAIR), Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences, Hong Kong SAR
∗ Equal Contribution
yxyuan@ee.cuhk.edu.hk

Abstract. In semi-supervised medical image segmentation (SSMIS), existing methods typically impose consistency or contrastive regularizations under basic data and network perturbations, and individually segment each voxel/pixel in the image. In fact, a dominating issue in medical scans is the intrinsic ambiguous regions due to unclear boundary and expert variability, whose segmentation requires the information in spatially nearby regions. Thus, these existing works are limited in data variety and tend to overlook the ability of inferring ambiguous regions with contextual information. To this end, we present Multi-Formation Soft Masking (MOST), a simple framework that effectively boosts SSMIS by learning spatial context relations with data regularity conditions. It first applies multi-formation function to enhance the data variety and perturbation space via partitioning and upsampling. Afterwards, each unlabeled data is soft-masked and is constrained to give invariant predictions as the original data. Therefore, the model is encouraged to infer ambiguous regions via varied granularities of contextual information conditions. Despite its simplicity, MOST achieves state-of-the-art performance on four common SSMIS benchmarks. Code and models are released at https://github.com/CUHK-AIM-Group/MOST-SSL4MIS.

Keywords: Multi-Formation Function · Soft Masking · Semi-supervised Learning · Medical Image Segmentation.

1 Introduction

Acquiring a large amount of labeled data is a significant challenge in medical image segmentation due to its high cost and the need for specialized expertise [4, 17, 18]. Hence, numerous studies have been focusing on semi-supervised medical image segmentation (SSMIS) [41, 35, 36, 37, 1, 40, 20], which trains a deep segmentation model by leveraging limited annotated medical images alongside a large number of unlabeled medical images to achieve satisfactory segmentation performance.

The training objectives for existing SSMIS methods can be generally categorized into consistency and contrastive. For consistency-based methods, they enforce the model to produce consistent predictions with geometrical transformed data or different perturbed networks [32, 35, 1, 7, 41, 23]. Nevertheless, the significant disparity between medical scans (*e.g.*, MRI & CT) and natural images hinders the applicability of many transformation and augmentation techniques [8], which leads to a limited data variety and perturbation design space. Therefore, *these methods may be prone to the test data with distribution shift* [1] and suffer from imprecise segmentation. This motivates us to explore alternative techniques that effectively expand the data diversity [10] while satisfying the data regularity conditions [13, 14] in medical imaging. For contrastive-based methods [12, 36, 33], they consider the voxels in medical scans as positive or negative instances, then apply contrastive loss [9, 3] to minimize the embedding distance of positive samples while pushing negative samples apart. These methods cluster each voxel/pixel and assign a class label to each of them. However, it is noted that the intrinsic ambiguity is a dominating issue in medical imaging, due to the unclear boundary and expert variability [31, 34], while overlooking this ambiguity can lead to inaccurate results and unreliable boundaries for lesions or organs [15]. Consequently, the above contrastive learning pattern that only considers voxels individually may *lack of contextual information learning and insufficient in segmenting these ambiguous regions* [28]. Hence, we are committed to boosting the model's capability in inferring these regions via spatial context reasoning [26, 11, 6], which enables the model to leverage contextual information to accurately delineate these regions.

To overcome the aforementioned challenges, we propose a simple and effective method coined Multi-Formation Soft Masking (MOST) for SSMIS, a simple consistency framework that uses a single segmentation network. It combines the merit of the aforementioned methods while alleviating their drawbacks, which is achieved by multi-granularity data and contextual information learning. Specifically, (1) we first incorporate the idea of multi-formation function into the framework by partitioning the data and uniformly upsample them to the original input size. It aims to enhance the data diversity, and mitigate the issue of being prone to shifted distributions. As presented in [14], the partitioned data naturally satisfies the data regularity [13], *i.e.*, spatially nearby pixels appear to be similar. Therefore, this operation serves as a structure-preserved data synthetic procedure for medical scans, which effectively improves the data variety, perturbation design space, and model generalization ability. (2) To obtain accurate segmentation results in ambiguous and difficult regions, we present a soft masking strategy that promotes the learning of contextual information within the data. Concretely, our soft masking strategy involves generating a rectangular mask for random 3D patches in the medical scan, and linearly interpolates the voxels around the masked boundaries to form a smooth transition. The model is then forced to produce consistent predictions for both masked and unmasked regions, and has to leverage the spatial context to infer the segmentation for the masked regions. This strategy not only boosts the model to learn spatial con-

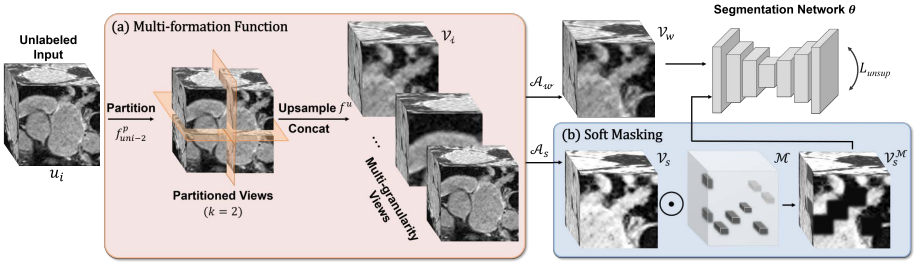

Fig. 1. Illustration of the proposed MOST framework for unlabeled input data u_i. To provide a clear visualization, the figure displays the partition policy with f^p_{uni-2}.

textual clues for inferring ambiguous regions in the image, but also prevents the model from relying on abrupt transition patterns that are essentially artifacts. Extensive experiments are conducted on four datasets, MOST achieves state-of-the-art performance and surpasses existing SSMIS counterparts significantly.

2 Methodology

In the semi-supervised medical image segmentation task, the training set is divided into two subsets: The labeled subset $D_l = \{(x_i, y_i)\}_{i=1}^{N_l}$ and the unlabeled subset $D_u = \{u_i\}_{i=1}^{N_u}$, where $N_u \gg N_l$. The medical images $x_i \in \mathbb{R}^{H\times W\times D}$ in D_l have corresponding ground-truth masks $y_i \in \{0,1\}^{C\times H\times W\times D}$, where C, H, W, D are the number of classes, height, width, and depth respectively. The objective is to train a segmentation model θ with both D_l and D_u and enhance the segmentation performance when deployed to an unseen test set.

The proposed MOST is constructed with a consistency-based framework with strong augmentations. As illustrated in Fig. 1, we adopt *a single segmentation network* (V-Net [25] for 3D volume or U-Net [29] for 2D slice), parameterized by θ. For each unlabeled image u_i, we first apply a multi-formation function f that creates multi-granularity views of the image $\mathcal{V}_i = f(u_i)$ with data regularity, which effectively boosts the variety of the samples. Then, the multi-granularity views $\mathcal{V}_i$ are augmented by weak and strong operators $\mathcal{A}_w$ and $\mathcal{A}_s$ in each iteration simultaneously to obtain the augmented data $\mathcal{V}_w = \mathcal{A}_w(\mathcal{V}_i)$ and $\mathcal{V}_s = \mathcal{A}_s(\mathcal{V}_i)$. In order to train the model to learn contextual information, we propose a soft masking operation that is applied on the multi-granularity strong views of the image $\mathcal{V}_s$ to produce the masked views $\mathcal{V}_s^{\mathcal{M}}$ with randomly masked patches. After feeding the multi-granularity weak views $\mathcal{V}_w$ into the network θ, we obtain the pseudo labels p_w. Finally, a consistency regularization is applied on the segmentation results of the masked strong views $\hat{y}_s$ and the pseudo labels p_w, and the model is trained to infer ambiguous regions in the image with spatial context and achieve better segmentation results.

2.1 Multi-formation Function

In SSMIS, applying consistency regularization for unlabeled data under different perturbations serves as an effective method for learning invariant and robust representations [32, 35, 1, 19]. However, as 3D medical images and natural images have significant disparity, only limited data transformation techniques are directly applicable [8]. Furthermore, medical images demonstrate significant variability in terms of anatomical structures and scanning protocols. Consequently, a substantial distribution gap often exists between the training and testing data. Hence, inspired by [14], we develop a multi-formation function that effectively enhances the diversity while satisfying the data regularity.

As displayed in Fig. 1 (a), given batch-wise unlabeled images $\{u_i\}_{i=1}^{\mathcal{B}_u}$, we partition each image u_i via a certain policy f^p, which can be a uniform partition f^p_{uni-k} with a specific factor k ($k > 1, k \in N^+$) or a multi-scale partition f^p_{ms-k} (Refer to Sec. 3.3 for the policy choice). Without loss of generality, we use f^p_{uni-k} as a representation. Therefore, the image is partitioned with the mapping function $f^p_{uni-k} : \mathbb{R}^{1\times H\times W\times D} \rightarrow \mathbb{R}^{k^2\times \frac{H}{k}\times \frac{W}{k}\times D}$, and is augmented to k^2 partitioned views[1]. Afterwards, an upsample operation f^u is applied on each partition to recover the original size: $f^u : \mathbb{R}^{k^2\times \frac{H}{k}\times \frac{W}{k}\times D} \rightarrow \mathbb{R}^{k^2\times H\times W\times D}$. After concatenating the partitioned views and the original view, we can obtain the **multi-granularity views** of the image,

$$\mathcal{V}_i = \mathrm{Cat}[u_i, f^u(f^p_{uni-k}(u_i))], \tag{1}$$

where Cat refers to the concatenation operation, and $\mathcal{V}_i \in \mathbb{R}^{(k^2+1)\times H\times W\times D}$. Different from other data or network perturbation methods [35, 22, 41] which exhibit limited perturbation variability, our method provides a wider range of data diversity, i.e., the cardinality of the transformed set of data $\{\mathcal{V}_i\}_{i=1}^{\mathcal{B}_u}$ displays a quadratic growth when compared to the input set of data $\{u_i\}_{i=1}^{\mathcal{B}_u}$, and the augmented views inherently satisfy the data regularity conditions in medical imaging [13]. Therefore, models trained using our approach are expected to possess enhanced robustness and generalisability.

2.2 Soft Masking

The contextual information is crucial for inferring ambiguous regions during segmenting medical images, as the human body tends to display a typical structure with prior anatomy information [27]. For example, locations of certain organs or surrounding tissues can help determine their boundaries even when they are not clearly defined. However, how to enhance the model's capability in inferring ambiguous regions via spatial context reasoning remains unexplored in SSMIS.

To this end, a soft-masking strategy is proposed to address the aforementioned deficiency. With the given multi-granularity views $\mathcal{V}_i$, they are first augmented via weak and strong operators $\mathcal{A}_w$ and $\mathcal{A}_s$ to produce $\mathcal{V}_w = \mathcal{A}_w(\mathcal{V}_i)$ and $\mathcal{V}_s = \mathcal{A}_s(\mathcal{V}_i)$. Then, we define the parameters for the soft masking template $\mathcal{M}$,

[1] Practically, a threshold η_{max} is set to replace k^2 to prevent excessive augmentation when k is set to a very large value: $\eta_{max} = \min(k^2, \eta_{max})$.

with 3D mask patch size $s \times s \times s$ and mask ratio r. To apply the soft mask, we begin by resizing an all-ones volume with the same size as input to $\mathcal{M}'$ with the dimension of $H/s \times W/s \times D/s$. A proportion of r voxels are randomly masked out, by setting their corresponding values to 0. Next, we are able to obtain the soft mask $\mathcal{M} \in \mathbb{R}^{H \times W \times D}$ via a trilinear upsampling $\mathcal{M} = \text{Trilinear}(\mathcal{M}')$, and we perform a Hadamard product between $\mathcal{M}$ and $\mathcal{V}_s$ for the soft masking,

$$\mathcal{V}_s^{\mathcal{M}} = \mathcal{M} \bigodot \mathcal{V}_s, \tag{2}$$

where $\mathcal{V}_s^{\mathcal{M}}$ refers to the strong views after soft masking. Compared to the hard masking methods [11, 6, 38, 5], soft masking prohibits the model from learning the abrupt transition characteristics within the image, meanwhile improves the model's capacity in learning spatial contextual information. Moreover, naive masked image modeling pretraining [38] is not sufficient to capture the complex context dependencies [11], thus the proposed soft masking is effective and practical for SSMIS.

2.3 Overall Objective

With the weak views $\mathcal{V}_w$ and masked strong views $\mathcal{V}_s^{\mathcal{M}}$, we first obtain the pseudo labels p_w and segmentation predictions $\hat{y}_s$ for the unlabeled data:

$$\begin{aligned} &\hat{y}_w = f(\mathcal{V}_w; \theta), \hat{y}_s = f(\mathcal{V}_s^{\mathcal{M}}; \theta), \\ &p_w = \mathbb{1}\left(\max\left(\hat{y}_w\right) \geq \tau\right) \text{argmax}(\hat{y}_w), \end{aligned} \tag{3}$$

where τ is a hyperparameter that denotes the threshold above which we retain a pseudo-label. Therefore, the unsupervised loss is formulated as,

$$\mathcal{L}_{unsup} = \frac{1}{\mathcal{B}_u} \sum_{i=1}^{\mathcal{B}_u} \frac{1}{HWD} \sum_{j=1}^{HWD} \text{CE}\left(\hat{y}_{s,i}(j), p_{w,i}(j)\right), \tag{4}$$

where CE denotes the cross-entropy. The training stage of MOST framework is conducted end-to-end, with the following overall optimization objective,

$$\mathcal{L} = \mathcal{L}_{sup} + \mathcal{L}_{unsup} + \mathcal{L}_{cp}, \tag{5}$$

where $\mathcal{L}_{sup}$ is the loss for supervised data, which is a combination of cross-entropy and dice loss between the segmentation prediction of the labeled data and the ground-truth [41, 21]. $\mathcal{L}_{unsup}$ is the loss for unsupervised data. Furthermore, we apply copy-paste as one augmentation operation in our framework, which is jointly trained with the labeled and unlabeled data via Dice loss $\mathcal{L}_{cp}$ following [1, 42]. During inference, the sample is directly fed into the trained model to produce the segmentation results. Therefore, no extra inference cost is required.

3 Experiments

3.1 Datasets and Implementation Details

To evaluate the effectiveness of the proposed MOST, we conduct experiments on four widely used SSMIS datasets: LA [39], Pancreas-CT [30], ACDC [2], and BraTS 2019 [24]. The data splits strictly follow common practice [41, 1, 37, 40].

Table 1. Results comparisons on LA dataset with 5% and 10% data labeled. ↑ denotes the higher the better and ↓ denotes the lower the better.

Method	Metrics (5% labeled)				Metrics (10% labeled)			
	Dice↑	Jac↑	95HD↓	ASD↓	Dice↑	Jac↑	95HD↓	ASD↓
Supervised	91.47	84.36	5.48	1.51	91.47	84.36	5.48	1.51
UA-MT [41]	82.26	70.98	13.71	3.82	87.79	78.39	8.68	2.12
SASSNet [16]	81.60	69.63	16.16	3.58	87.54	78.05	9.84	2.59
DTC [22]	81.25	69.33	14.90	3.99	87.51	78.17	8.23	2.36
URPC [23]	82.48	71.35	14.65	3.65	86.92	77.03	11.13	2.28
MC-Net [35]	83.59	72.36	14.07	2.70	87.62	78.25	10.03	1.82
SS-Net [36]	86.33	76.15	9.97	2.31	88.55	79.62	7.49	1.90
BCP [1]	88.02	78.72	7.90	2.15	89.62	81.31	6.81	1.76
CAML [7]	87.34	77.65	9.76	2.49	89.62	81.28	8.76	2.02
MOST	**89.51**	**81.10**	**5.92**	**2.02**	**91.17**	**83.85**	**5.63**	**1.76**

Table 2. Results comparison on Pancreas-CT with 20% data labeled.

Method	Metrics			
	Dice↑	Jac↑	95HD↓	ASD↓
Supervised	82.60	70.81	5.61	1.33
UA-MT [41]	77.26	63.82	11.90	3.06
SASSNet [16]	77.66	64.08	10.93	3.05
DTC [22]	78.27	64.75	8.36	2.25
URPC [23]	80.02	67.30	8.51	1.98
FUSSNet [37]	81.82	69.76	5.42	1.51
BCP [1]	82.91	70.97	6.43	2.25
MOST	**83.84**	**72.40**	**4.42**	**1.10**

Table 3. Results comparison on ACDC with 10% data labeled.

Method	Metrics			
	Dice↑	Jac↑	95HD↓	ASD↓
Supervised	91.44	84.59	4.30	0.99
UA-MT [41]	81.65	70.64	6.88	2.02
SASSNet [16]	84.50	74.34	5.42	1.86
DTC [22]	84.29	73.92	12.81	4.01
URPC [23]	83.10	72.41	4.84	1.53
SS-Net [36]	86.78	77.67	6.07	1.40
BCP [1]	88.84	80.62	3.98	1.17
MOST	**89.29**	**81.23**	**3.28**	**0.98**

We use V-Net [25] for 3D datasets (LA, Pancreas-CT, BraTS 2019) and U-Net [29] for 2D dataset (ACDC). We train 15k iterations with batch size 4 on 3D datasets and 30k iterations with batch size 24 on 2D datasets, using the SGD optimizer with initial learning rate 0.01 and a cosine scheduler. Random crop and copy-paste [42, 1] are used as the basic weak augmentations $\mathcal{A}_w$, and we apply random gamma adjustment [3] as the specific strong augmentation in $\mathcal{A}_s$. During training, the scans are cropped to $112 \times 112 \times 80$ in LA, $96 \times 96 \times 96$ in Pancreas-CT and BraTS 2019, and 256×256 in ACDC, respectively, which strictly follow [1, 37, 40]. Threshold τ for pseudo-labeling is fixed to 0.75. During inference, a sliding window is used to obtain segmentation results for 3D datasets, with a stride of $18 \times 18 \times 4$ on LA, $16 \times 16 \times 4$ on Pancreas-CT, and $64 \times 64 \times 64$ on BraTS 2019. The experiments of MOST are repeated with three different random seeds [7], and the mean performances are reported. For evaluation, we use the four metrics including Dice Coefficient (*Dice*), Jaccard Score (*Jac*), 95% Hausdorff Distance (*95HD*), and Average Surface Distance (*ASD*).

Table 4. Ablation study on the proposed components in our MOST.

Module			Metrics			
SA	MF	SM	Dice↑	Jac↑	95HD↓	ASD↓
✓			88.08	79.48	6.68	1.98
✓	✓		89.57	81.30	7.54	1.77
	✓	✓	90.53	82.78	5.70	1.80
✓		✓	90.11	82.18	6.65	1.97
✓	✓	✓	**91.17**	**83.85**	**5.63**	**1.76**

Table 5. Ablation study on the multi-formation function policy.

Function	Dice↑	Jac↑	95HD↓	ASD↓
None	90.11	82.18	6.65	1.97
Uniform-2	91.17	83.85	5.63	1.76
Uniform-3	**91.22**	**83.93**	**5.01**	**1.54**
Uniform-4	90.92	83.48	5.59	1.71
Multi-scale-4	90.90	83.40	5.22	1.56

3.2 Comparison with State-of-the-art Methods

We compare the proposed MOST with the state-of-the-art SSMIS methods [41, 16, 22, 23, 35, 36, 37, 1, 7]. The results on LA with with 5% and 10% data labeled are shown in Table 1. MOST gives 91.17% Dice and 5.63 95HD with 10% labels, surpassing CAML [7] by 1.55% and 3.13. Compared to BCP [1], MOST comprehensively surpasses it with 1.55% Dice, 2.24% Jaccard, and 1.18 95HD, respectively. Our method also surpasses previous methods with only 5% labels available and achieves a remarkable Dice of 89.51%, verifying its effectiveness in different labeling scenarios.Results on Pancreas-CT is presented in Table 2. Specifically, MOST achieves the best results in terms of all metrics, with 83.84% Dice, 72.40% Jaccard, 4.42 95HD, and 1.10 ASD, respectively, demonstrating the effectiveness in segmenting ambiguous boundaries of the proposed design. Moreover, MOST can also be extended on 2D datasets by setting scan depth $D = 1$. As presented in Table 3 on ACDC, MOST outperforms previous state-of-the-art BCP [1] on all metrics, with 89.29% Dice, 81.23% Jaccard, 3.28 95HD, and 0.98 ASD, demonstrating its capacity and generalization ability in different SSMIS tasks. Lastly, on BraTS 2019 with 10% data labeled, MOST still achieves 84.17% Dice, surpassing the supervised performance with 83.84% Dice. More detailed results are given in the supplementary.

3.3 Further Analysis

Ablation study on the components in MOST. We first investigate the effectiveness of each component in MOST in Table 4. “SA” means the strong augmentation framework, "MF" represents the multi-formation function, and "SM" means the soft masking strategy, respectively. With SA, the model shows 88.08% Dice on LA. When combining SA with MF, the Dice improves by 1.49%, and SM further improves the performance. The best overall result is obtained when combining the three proposed components, which shows their complementary benefits, and verifies the superiority of the proposed framework.

Impact of the multi-formation function policy. As the multi-formation function augments the data into partitioned views, we ablate 5 choices of the policy: uniform partition with factor 2, 3, 4; multi-scale partition with maximum factor 4; no multi-formation. Results are provided in Table 5. Compared to no multi-formation, all other methods surpass it remarkably, highlighting the im-

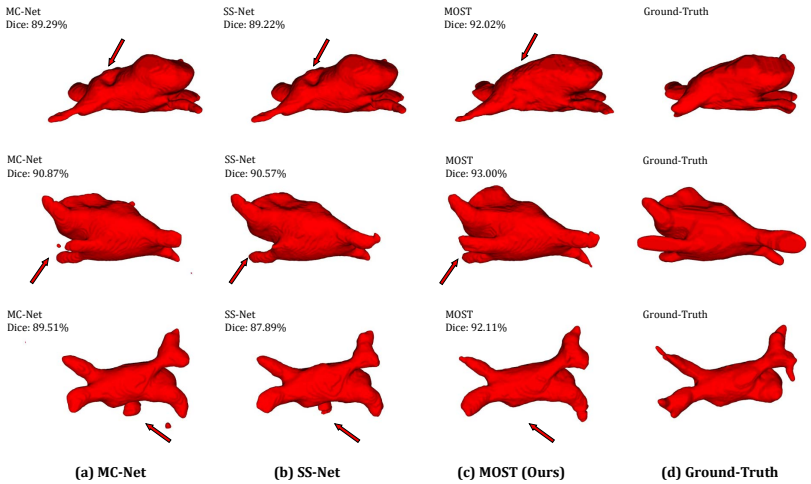

Fig. 2. Segmentation performance comparison on 3D LA dataset among (a) MC-Net [35], (b) SS-Net [36], (c) the proposed MOST, (d) ground-truth.

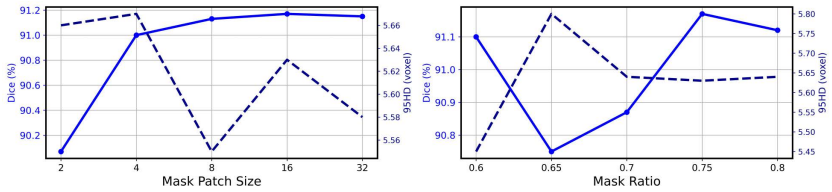

Fig. 3. Ablation study on the hyperparameters in soft masking. Left: different mask patch sizes. Right: different mask ratios.

portance of the proposed transformation. The function designs exhibit similar performance, with Uniform-3 achieving slightly superior results, which indicates the robustness of the proposed method. Considering the simplicity and consistency, Uniform-2 f^{p}_{uni-2} policy is used in all experiments.

Qualitative segmentation results. We depict the qualitative segmentation results on LA dataset of (a) MC-Net [35], (b) SS-Net [36], (c) MOST (Ours), and (d) ground-truth in Fig. 2. Compared with MC-Net [35], the proposed method recognizes the anatomy structure precisely and alleviates incomplete segmentation regions (2^{nd} and 3^{rd} rows). Moreover, compared with SS-Net [36], MOST not only produces high-quality segmentation results on the ambiguous boundaries (1^{st} row), but also accurately captures the fine details (2^{nd} row).

Impact of different hyperparameters. We evaluate the effect of varied mask patch sizes and ratios in Fig. 3. MOST is relatively stable for hyperparameters overall. For the patch size, only a small masked size ($\leq$2) will cause a significant Dice decrease. This may due to the large area of available context simplifies the learning of contextual information. We fix the masked patch size

to 16 on all datasets. For the ratio of masked region, the model achieves the highest Dice at 0.75, and we use this value in all our experiments.

4 Conclusion

In this work, we aim to address the limited data variety and intrinsic ambiguity issues in SSMIS. To this end, we propose MOST, a simple and effective method that jointly enhances the data variety and learns the contextual information to infer the ambiguous regions. Built upon a framework with strong augmentation, it first adopts a multi-formation function via partitioning and upsampling. Then, a soft masking is applied on the unlabeled images, and the model is constrained to provide consistent predictions as the original input. Extensive experiments on four benchmark datasets are conducted, and the proposed method shows significant performance superiority over existing approaches.

Acknowledgments. This work was supported by Hong Kong Research Grants Council (RGC) General Research Fund 14204321.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y.: Bidirectional copy-paste for semi-supervised medical image segmentation. In: CVPR. pp. 11514–11524 (2023)
2. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging **37**(11), 2514–2525 (2018)
3. Chen, T., Kornblith, S., Swersky, K., Norouzi, M., Hinton, G.E.: Big self-supervised models are strong semi-supervised learners. NeurIPS **33**, 22243–22255 (2020)
4. Cheplygina, V., de Bruijne, M., Pluim, J.P.: Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in med. image anal. Med. Image Anal. **54**, 280–296 (2019)
5. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
6. Gao, S., Zhou, P., Cheng, M.M., Yan, S.: Masked diffusion transformer is a strong image synthesizer. arXiv preprint arXiv:2303.14389 (2023)
7. Gao, S., Zhang, Z., Ma, J., Li, Z., Zhang, S.: Correlation-aware mutual learning for semi-supervised medical image segmentation. In: MICCAI. pp. 98–108. Springer (2023)
8. Garcea, F., Serra, A., Lamberti, F., Morra, L.: Data augmentation for medical imaging: A systematic literature review. Comput. Biol. Med. p. 106391 (2022)
9. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: CVPR. pp. 9729–9738 (2020)
10. Ho, D., Liang, E., Chen, X., Stoica, I., Abbeel, P.: Population based augmentation: Efficient learning of augmentation policy schedules. In: ICML. pp. 2731–2741. PMLR (2019)
11. Hoyer, L., Dai, D., Wang, H., Van Gool, L.: Mic: Masked image consistency for context-enhanced domain adaptation. In: CVPR. pp. 11721–11732 (2023)

12. Hu, X., Zeng, D., Xu, X., Shi, Y.: Semi-supervised contrastive learning for label-efficient medical image segmentation. In: MICCAI. pp. 481–490. Springer (2021)
13. Huang, J., Mumford, D.: Statistics of natural images and models. In: CVPR. vol. 1, pp. 541–547. IEEE (1999)
14. Kim, J.H., Kim, J., Oh, S.J., Yun, S., Song, H., Jeong, J., Ha, J.W., Song, H.O.: Dataset condensation via efficient synthetic-data parameterization. In: ICML. pp. 11102–11118. PMLR (2022)
15. Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K., Eslami, S., Jimenez Rezende, D., Ronneberger, O.: A probabilistic u-net for segmentation of ambiguous images. NeurIPS **31** (2018)
16. Li, S., Zhang, C., He, X.: Shape-aware semi-supervised 3d semantic segmentation for medical images. In: MICCAI. pp. 552–561. Springer (2020)
17. Liu, X., Guo, X., Liu, Y., Yuan, Y.: Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images. Medical image analysis **71**, 102052 (2021)
18. Liu, X., Li, W., Yuan, Y.: Decoupled unbiased teacher for source-free domain adaptive medical object detection. IEEE Trans. Neural Netw. Learn. Syst. (2023)
19. Liu, X., Yuan, Y.: A source-free domain adaptive polyp detection framework with style diversification flow. IEEE Trans. Med. Imagi **41**(7), 1897–1908 (2022)
20. Lu, W., Lei, J., Qiu, P., Sheng, R., Zhou, J., Lu, X., Yang, Y.: Upcol: Uncertainty-informed prototype consistency learning for semi-supervised medical image segmentation. In: MICCAI. pp. 662–672. Springer (2023)
21. Luo, X.: SSL4MIS. `https://github.com/HiLab-git/SSL4MIS` (2020)
22. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: AAAI. vol. 35, pp. 8801–8809 (2021)
23. Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Metaxas, D.N., Zhang, S.: Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. Med. Image Anal. **80**, 102517 (2022)
24. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2014)
25. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3DV. pp. 565–571. Ieee (2016)
26. Mottaghi, R., Chen, X., Liu, X., Cho, N.G., Lee, S.W., Fidler, S., Urtasun, R., Yuille, A.: The role of context for object detection and semantic segmentation in the wild. In: CVPR. pp. 891–898 (2014)
27. Paragios, N.: A level set approach for shape-driven segmentation and tracking of the left ventricle. IEEE Trans. Med. Imaging **22**(6), 773–776 (2003)
28. Rai, S.N., Cermelli, F., Fontanel, D., Masone, C., Caputo, B.: Unmasking anomalies in road-scene segmentation. arXiv preprint arXiv:2307.13316 (2023)
29. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
30. Roth, H.R., Lu, L., Farag, A., Shin, H.C., Liu, J., Turkbey, E.B., Summers, R.M.: Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In: MICCAI. pp. 556–564. Springer (2015)
31. Singh, S.P., Wang, L., Gupta, S., Goli, H., Padmanabhan, P., Gulyás, B.: 3d deep learning on medical images: a review. Sensors **20**(18), 5097 (2020)
32. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. NeurIPS **33** (2020)

33. Wang, K., Zhan, B., Zu, C., Wu, X., Zhou, J., Zhou, L., Wang, Y.: Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning. Med. Image Anal. **79**, 102447 (2022)
34. Wang, R., Chen, S., Ji, C., Fan, J., Li, Y.: Boundary-aware context neural network for medical image segmentation. Med. Image Anal. **78**, 102395 (2022)
35. Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J.: Mutual consistency learning for semi-supervised medical image segmentation. Med. Image Anal. **81**, 102530 (2022)
36. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: MICCAI. pp. 34–43. Springer (2022)
37. Xiang, J., Qiu, P., Yang, Y.: Fussnet: Fusing two sources of uncertainty for semi-supervised medical image segmentation. In: MICCAI. pp. 481–491. Springer (2022)
38. Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., Dai, Q., Hu, H.: Simmim: A simple framework for masked image modeling. In: CVPR. pp. 9653–9663 (2022)
39. Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., et al.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Med. Image Anal. **67**, 101832 (2021)
40. Xu, Z., Wang, Y., Lu, D., Luo, X., Yan, J., Zheng, Y., Tong, R.K.y.: Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation. Med. Image Anal. **88**, 102880 (2023)
41. 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. pp. 605–613. Springer (2019)
42. Zhao, Z., Yang, L., Long, S., Pi, J., Zhou, L., Wang, J.: Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In: CVPR. pp. 11350–11359 (2023)