Pair Shuffle Consistency for Semi-supervised Medical Image Segmentation

Jianjun He[1], Chenyu Cai[1], Qiong Li[2], and Andy J Ma[1,3,4](✉)

[1] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
{hejj56,caichy8}@mail2.sysu.edu.cn,
majh8@mail.sysu.edu.cn

[2] Department of Radiology, Sun Yat-sen University Cancer Center, Guangzhou, China
liqiong@sysucc.org.cn

[3] Guangdong Province Key Laboratory of Information Security Technology, Guangzhou, China

[4] Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, Guangzhou, China

Abstract. Semi-supervised medical image segmentation is a practical but challenging problem, in which only limited pixel-wise annotations are available for training. While most existing methods train a segmentation model by using the labeled and unlabeled data separately, the learning paradigm solely based on unlabeled data is less reliable due to the possible incorrectness of pseudo labels. In this paper, we propose a novel method namely pair shuffle consistency (PSC) learning for semi-supervised medical image segmentation. The pair shuffle operation splits an image pair into patches, and then randomly shuffle them to obtain mixed images. With the shuffled images for training, local information is better interpreted for pixel-wise predictions. The consistency learning of labeled-unlabeled image pairs becomes more reliable, since predictions of the unlabeled data can be learned from those of the labeled data with ground truth. To enhance the model robustness, the consistency constraint on unlabeled-unlabeled image pairs serves as a regularization term, thereby further improving the segmentation performance. Experiments on three benchmarks demonstrate that our method outperforms the state of the art for semi-supervised medical image segmentation.

Keywords: Semi-supervised learning · Medical image segmentation · Pair shuffle · Consistency learning

1 Introduction

Semi-supervised learning (SSL) has emerged as an important technique in medical image segmentation [10, 12, 14, 20, 22, 23], to reduce the demand of pixel-level manual annotations which are both expensive and time-consuming to obtain. The objective of SSL is to harness the vast amount of unlabeled data, improving the segmentation model trained by only a small amount of labeled data.

In most existing SSL methods developed for medical image segmentation, the labeled and unlabeled data are used separately for training. On the one hand, supervised learning is conducted by the labeled data with pixel-level ground truth. On the other hand, the unlabeled data is utilized under two typical learning paradigms, i.e., self-training [8, 9, 21] and consistency regularization [1, 6, 18, 19, 23]. The former generates pseudo labels for the unlabeled data to train the model, while the latter constrains consistent predictions among different views of an unlabeled input generated by augmentations.

Despite the success, both learning paradigms entirely based on unlabeled data has unavoidable defects due to the lack of ground truth labels. Neither paradigm can guarantee the correctness of the pseudo labels. This dilemma inspires us to consider a new SSL paradigm, which uses both labeled and unlabeled images to synthesize mixed images for guidance in model learning by not only ground truth but also pseudo labels. In this manner, we allow the abundant unlabeled data to benefit from the more precise but limited ground truth labels. Few existing methods have been proposed to address this issue. In [1], BCP extends the copy-paste approach in a bidirectional manner, in which a labeled image is randomly cropped and pasted onto an unlabeled image, and vice versa. MagicNet [4] presents a data augmentation strategy to partition and recover N^3 small cubes cross- and within-labeled and unlabeled images.

For pixel-wise predictions, segmentation models need to pay more attention to local information. In this work, we propose a novel image augmentation operation called pair shuffle, which divides paired images into equal-sized patches, randomly shuffle them, and then combine them into two mixed images. Different from existing methods, pair shuffle modifies the relative positions of patches within an image pair. The global information is deliberately disrupted while the local information is preserved within each patch. By inputting the shuffled images for training, the segmentation model is unable to perceive them as a whole. Instead, local features are extracted from each patch at a finer granularity.

Based on the Mean Teacher [18] architecture, we propose the pair shuffle consistency (PSC) learning with both labeled-unlabeled and unlabeled-unlabeled image pairs. The former promotes to learn discriminative features from the labeled to the unlabeled data more reliably, while the latter enhances the model robustness solely based on the large amount of unlabeled data. We feed the shuffled image pair to the student model, facilitating its ability to comprehend local information. Simultaneously, the original images are input to the teacher model for generating pseudo labels by global interpretation. Through the consistency constraint between the teacher and student, both global and local information is effectively integrated, leading to the improved segmentation performance.

In summary, the main contributions of this paper are as follows. 1) We propose a novel image augmentation operation called pair shuffle, which changes the relative positions of patches within image pairs, to train a segmentation modal for better perceiving local information. 2) We propose an innovative method namely pair shuffle consistency (PSC) learning for semi-supervised medical image segmentation, encouraging reliable learning from the labeled to the unlabeled data.

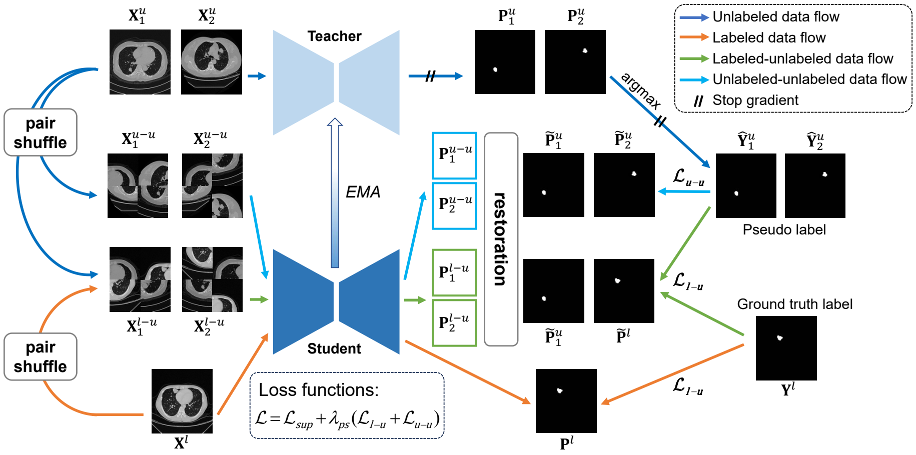

Fig. 1: Overview of our pair shuffle consistency (PSC) learning framework.

3) The proposed method outperforms the state of the art in various tasks such as tumor and organ segmentation, in both binary- and multi-class scenarios.

2 Method

2.1 Overall Framework

Mathematically, we define the 2D slice of a medical image as $\mathbf{X} \in \mathbb{R}^{H\times W}$. The goal of semi-supervised medical image segmentation is to predict the per-pixel label map $\mathbf{Y} \in \{0, 1, \cdots, K-1\}^{H\times W}$, where K is the number of classes and class 0 represents the background. The training set $\mathcal{D}$ consists of N labeled images and M unlabeled images ($N \ll M$), divided into two subsets $\mathcal{D} = \mathcal{D}^l \cup \mathcal{D}^u$, where $\mathcal{D}^l = (\mathbf{X}_i^l, \mathbf{Y}_i^l)_{i=1}^N$ and $\mathcal{D}^u = (\mathbf{X}_i^u)_{i=N+1}^{M+N}$.

The proposed pair shuffle consistency (PSC) learning method follows a teacher-student scheme. The overall framework is depicted in Fig. 1. The teacher and the student models are denoted as $\mathcal{F}_t(\mathbf{X}; \mathbf{\Theta_t})$ and $\mathcal{F}_s(\mathbf{X}; \mathbf{\Theta_s})$ respectively, where $\mathbf{\Theta_t}$ and $\mathbf{\Theta_s}$ are the learnable parameters. The teacher model is updated by using exponential moving average (EMA) based on the student [18], i.e. $\mathbf{\Theta_t} = \omega\mathbf{\Theta_t} + (1-\omega)\mathbf{\Theta_s}, \omega \in (0, 1)$. The student model is optimized through gradient descent of the overall loss function $\mathcal{L}$ given by:

$$\mathcal{L} = \mathcal{L}_{sup} + \lambda_{ps}(\mathcal{L}_{l-u} + \mathcal{L}_{u-u}), \tag{1}$$

where $\mathcal{L}_{sup}$ is the supervised loss, λ_{ps} is a trade-off hyper-parameter, $\mathcal{L}_{l-u}$ and $\mathcal{L}_{u-u}$ denote the loss functions for pair shuffle consistency learning of labeled-unlabeled and unlabeled-unlabeled image pairs, respectively.

For the supervised loss $\mathcal{L}_{sup}$, labeled samples are fed into the student model to obtain prediction outputs $\mathbf{P}^l = \mathcal{F}_s(\mathbf{X}^l; \mathbf{\Theta_s})$. Ground truth labels $\mathbf{Y}^l$ are used

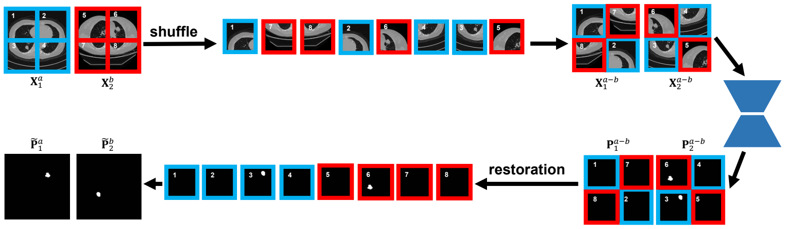

Fig. 2: Illustration of pair shuffle and restoration operation.

to supervise them as shown at the bottom of Fig. 1. This loss function is defined by combining the cross entropy ($\mathcal{L}_{ce}$) and the dice loss [15] ($\mathcal{L}_{dice}$), i.e.:

$$\mathcal{L}_{sup} = \mathcal{L}_{ce}(\mathbf{P}^l; \mathbf{Y}^l) + \mathcal{L}_{dice}(\mathbf{P}^l; \mathbf{Y}^l). \tag{2}$$

Details about the innovative pair shuffle operation and pair shuffle consistency learning approach corresponding to the loss functions $\mathcal{L}_{l-u}$ and $\mathcal{L}_{u-u}$ are provided in the following two subsections.

2.2 Pair Shuffle Operation

The proposed pair shuffle operation is illustrated in Fig. 2. We partition a pair of images, denoted as $\mathbf{X}_1^a$ and $\mathbf{X}_2^b$, into $N \times N$ equal-sized patches. The subscripts a and b assume the values l or u, representing labeled or unlabeled data, respectively. Subsequently, these $2 \times N \times N$ patches are randomly shuffled with a specified permutation order s, resulting in two mixed images, namely $\mathbf{X}_1^{a-b}$ and $\mathbf{X}_2^{a-b}$. The pair shuffle operation can be concisely formulated as follows:

$$\mathbf{X}_1^{a-b}, \mathbf{X}_2^{a-b} = \mathcal{PS}(\mathbf{X}_1^a, \mathbf{X}_2^b, N, s). \tag{3}$$

After shuffling, the mixed images $\mathbf{X}_1^{a-b}$ and $\mathbf{X}_2^{a-b}$ are fed into the student model and obtain their predictions, denoted as $\mathbf{P}_1^{a-b} = \mathcal{F}_s(\mathbf{X}_1^{a-b}; \mathbf{\Theta_s})$ and $\mathbf{P}_2^{a-b} = \mathcal{F}_s(\mathbf{X}_2^{a-b}; \mathbf{\Theta_s})$, respectively. Due to the disorganized positions of the shuffled patches, these predictions cannot be directly used for training. Therefore, we proceed to restore them back to their original orders using the recorded s, resulting in $\widetilde{\mathbf{P}}_1^a$ and $\widetilde{\mathbf{P}}_2^b$. Through our innovative pair shuffle operation, the global information of images is intentionally disturbed. These shuffled images can encourage model learning to exploit local information at a finer granularity from a patch-level perspective, thereby improving the segmentation performance.

2.3 Pair Shuffle Consistency Learning

The pair shuffle operation is performed on both labeled-unlabeled and unlabeled-unlabeled image pairs for consistency learning. To mitigate the error of pseudo label estimation, a labeled image $\mathbf{X}^l$ and an unlabeled image $\mathbf{X}_1^u$ are randomly paired to generate shuffled images $\mathbf{X}_1^{l-u}$ and $\mathbf{X}_2^{l-u}$. After that, we feed the shuffled images into the student model and obtain the shuffled predictions $\mathbf{P}_1^{l-u}$ and $\mathbf{P}_2^{l-u}$. By restoring to the original order, they are converted to $\widetilde{\mathbf{P}}^l$ and $\widetilde{\mathbf{P}}_1^u$ as the predictions of the labeled and unlabeled images, respectively. Simultaneously, $\mathbf{X}_1^u$ is input to the teacher model which outputs predictions $\mathbf{P}_1^u$. By applying the *argmax* function to $\mathbf{P}_1^u$, pseudo labels for the unlabeled data are generated, denoted as $\hat{\mathbf{Y}}_1^u$. They are then combined with the ground truth label $\mathbf{Y}^l$ to supervise $\widetilde{\mathbf{P}}_1^u$ and $\widetilde{\mathbf{P}}^l$ in the loss function $\mathcal{L}_{l-u}$. Similar to $\mathcal{L}_{sup}$, $\mathcal{L}_{l-u}$ is a combination of the cross entropy and the dice loss, given by:

$$\mathcal{L}_{l-u} = \mathcal{L}_{ce}(\widetilde{\mathbf{P}}^l; \mathbf{Y}^l) + \mathcal{L}_{ce}(\widetilde{\mathbf{P}}_1^u; \hat{\mathbf{Y}}_1^u) + \mathcal{L}_{dice}(\widetilde{\mathbf{P}}^l; \mathbf{Y}^l) + \mathcal{L}_{dice}(\widetilde{\mathbf{P}}_1^u; \hat{\mathbf{Y}}_1^u). \tag{4}$$

It yields great benefit for semi-supervised medical image segmentation by utilizing pair shuffle consistency learning with labeled-unlabeled data pairs. With ground truth labels, available only for a limited amount of labeled data, the unlabeled data can be guided by not only the pseudo labels from the teacher model but also the accurate annotations. As a result, the training process with both the labeled and unlabeled data becomes more reliable.

In addition, the proposed pair shuffle consistency learning is extended to exclusively using unlabeled data, preventing from over-fitting to the labeled data. Specifically, we apply the pair shuffle operation to unlabeled-unlabeled image pairs $\mathbf{X}_1^u$ and $\mathbf{X}_2^u$, resulting in mixed images denoted as $\mathbf{X}_1^{u-u}$ and $\mathbf{X}_2^{u-u}$. Similar to labeled-unlabeled pairs, these mixed images are fed into the student model to generate their corresponding restored predictions, denoted as $\widetilde{\mathbf{P}}_1^u$ and $\widetilde{\mathbf{P}}_2^u$. These predictions are then supervised by the pseudo labels $\hat{\mathbf{Y}}_1^u$ and $\hat{\mathbf{Y}}_2^u$ generated by the teacher model for optimization. Analogous to $\mathcal{L}_{l-u}$, we design the loss function $\mathcal{L}_{u-u}$ to facilitate the pair shuffle consistency learning of unlabeled-unlabeled image pairs, which can be formulated as follows:

$$\mathcal{L}_{u-u} = \mathcal{L}_{ce}(\widetilde{\mathbf{P}}_1^u; \hat{\mathbf{Y}}_1^u) + \mathcal{L}_{ce}(\widetilde{\mathbf{P}}_2^u; \hat{\mathbf{Y}}_2^u) + \mathcal{L}_{dice}(\widetilde{\mathbf{P}}_1^u; \hat{\mathbf{Y}}_1^u) + \mathcal{L}_{dice}(\widetilde{\mathbf{P}}_2^u; \hat{\mathbf{Y}}_2^u). \tag{5}$$

By exclusively relying on unlabeled data, the consistency learning on unlabeled-unlabeled image pairs serves as a complementary term to the former part $\mathcal{L}_{l-u}$. The loss function $\mathcal{L}_{u-u}$ fully utilizes the massive unlabeled data for model regularization to enhance the overall robustness.

3 Experiments

3.1 Dataset

We evaluate our method on three datasets, i.e., MLT, ACDC, and Promise12. The division is conducted in units of 3D volumes, and we convert them into 2D slices for experiments. We collect a large-scale dataset for malignant lung tumor (MLT) segmentation. It consists of 960 CT volumes from 949 patients, with a total of 2,760 2D-slices. For training, validation, and testing, we used 2,400, 160, and 200 slices, respectively. The dataset was partitioned according to

Table 1: Comparison with state-of-the-art methods on MLT dataset with Dice (%) and HD95. The best and the second-best results are marked in bold and underlined, respectively.

Setting		5%		10%	
Method	Publication	Dice↑	HD95↓	Dice↑	HD95↓
MT [18]	NIPS 2017	67.30	17.93	71.89	14.75
UAMT [22]	MICCAI 2019	67.85	16.21	71.62	14.92
CCT [16]	CVPR 2020	65.80	15.13	72.55	15.94
CPS [7]	CVPR 2021	68.68	13.92	73.88	15.45
URPC [14]	MICCAI 2021	69.18	14.13	74.15	15.75
MC-Net+ [20]	MIA 2022	69.41	14.42	73.04	14.62
SLC-Net [12]	MICCAI 2022	67.85	17.26	71.78	14.92
ICT [19]	NN 2022	69.23	14.98	74.06	14.68
BCP [1]	CVPR 2023	70.41	$\underline{13.86}$	$\underline{75.62}$	$\underline{12.31}$
DCNet [5]	MICCAI 2023	67.57	14.48	73.83	15.74
ICL [24]	MIDL 2023	$\underline{70.91}$	15.08	74.13	16.51
PSC (**ours**)	This paper	**71.89**	**13.19**	**77.44**	**11.10**

the protocol where 5% and 10% of the images were annotated, resulting in 120 and 240 annotated images. The Automated cardiac diagnosis challenge dataset (ACDC) [3] is a multi-class segmentation dataset including the right ventricle, the left ventricle cavities, and the myocardium (epicardial contour more specifically). It contains multi-slice 2D cine cardiac MR imaging samples from 100 patients, and is split into 70/10/20 for train/val/test. We set the amount of labeled patients to 7 (10%) following [13]. The Prostate MR Image Segmentation 2012 (Promise12) [11] is the dataset of the prostate segmentation challenge in MICCAI 2012. We divide 50 T2-weighted MRI volumes into 35/5/10 for train/val/test. We experiment with 7 annotated training volumes (20%) following [12].

3.2 Evaluation Metrics and Implementation Details

We use several well-known metrics in medical image segmentation, including Dice similarity coefficient (Dice), which is a set similarity measurement function. Additionally, we use the 95% Hausdorff Distance (HD95) and Average Surface Distance (ASD) which can reflect the segmentation accuracy of the boundary.

UNet [17] is employed as the backbone network $\mathcal{F}$. For all the experiments, we train for 300 epochs using the SGD optimizer with a momentum of 0.9 and weight decay of 1×10^{-4}. The learning rate η_0 of $\mathcal{F}_s(X; \theta_s)$ is set to 1×10^{-2}, with a poly scheduler $\eta = \eta_0 \times (1 - iter/max\_iter)^{0.9}$ as in [19]. We set $\lambda_{ps} = 1.0$ in Eq. 1 for all the datasets and experiment settings. The N in Eq. 3 is determined empirically as $N = 4$ for ACDC and $N = 8$ for MLT and Promise12. For preprocessing, we randomly apply rotation and flipping to all images, and then scale them to the same size of 256×256.

3.3 Comparison with State-of-the-Art Methods

Results on MLT. We compare the proposed PSC on the MLT dataset with the existing competitors. Results of different methods are reproduced in the same

Table 2: Comparison with state-of-the-art methods on ACDC dataset with Dice (%), HD95 and ASD.

Method	Publication	Dice↑	HD95↓	ASD ↓
MT [18]	NIPS 2017	87.53	4.58	1.64
UA-MT [22]	MICCAI 2019	86.34	5.32	1.55
CCT [16]	CVPR 2020	87.74	4.65	1.41
CPS [7]	CVPR 2021	85.03	9.06	2.84
URPC [14]	MICCAI 2021	86.93	4.56	1.36
MC-Net+ [20]	MEDIA 2022	86.98	5.36	1.51
ICL [24]	MIDL 2023	88.18	2.46	0.69
ICT [19]	NN 2022	87.15	6.03	1.78
BCP [1]	CVPR 2023	88.84	3.98	1.17
CL [2]	CVPR 2023	89.10	4.98	1.80
DCNet [5]	MICCAI 2023	89.55	4.69	1.58
PSC (**ours**)	This paper	**90.29**	**1.75**	**0.61**

Table 3: Comparison with state-of-the-art methods on Promise12 dataset with Dice (%), HD95 and ASD.

Method	Publication	Dice↑	HD95↓	ASD ↓
MT [18]	NIPS 2017	75.30	13.87	3.95
UA-MT [22]	MICCAI 2019	73.16	14.31	4.15
CCT [16]	CVPR 2020	74.88	9.12	4.09
CPS [7]	CVPR 2021	73.84	11.92	3.55
URPC [14]	MICCAI 2021	75.83	9.83	2.72
MC-Net+ [20]	MEDIA 2022	75.64	8.58	3.25
SLC-Net [12]	MICCAI 2022	75.91	10.23	3.13
ICT [19]	NN 2022	76.74	10.49	3.26
BCP [1]	CVPR 2023	81.09	7.37	3.49
DCNet [5]	MICCAI 2023	78.89	7.41	3.45
ICL [24]	MIDL 2023	78.86	7.15	2.51
PSC (**ours**)	This paper	**83.64**	**4.58**	**2.04**

experimental setting for fair comparisons. As shown in Table 1, our method achieves the best performance on both evaluation metrics, outperforming other competitors on Dice and HD95 (i.e., surpassing the second best by 0.98 percentage point (pp) and 0.67 on 5% setting, 1.82pp and 1.21 on 10% setting, respectively). By employing our PSC, the local information within patches is effectively integrated with the global information of images through consistency learning, thereby improving the performance of tumor segmentation.

Results on ACDC. Table 2 shows the averaged performance of the four-class segmentation results on the ACDC dataset with 10% labeled ratio. Our PSC surpasses the state of the arts. This suggests that during the training phase, blending the information from labeled and unlabeled data proves more beneficial in harnessing the limited labeled ground truth, compared to the paradigm of training them isolatedly. Specifically, the proposed PSC improves the segmentation performance over BCP, proving that altering the relative positions to better exploit local information at a finer granularity is beneficial for segmentation.

Results on Promise12. In Table 3, we summarize the results on the Promise12 dataset. Noticeable improvements compared with second-best method can be seen for Dice(↑), HD95(↓) and ASD(↓) with 2.55pp, 2.57 and 0.47 respectively. It demonstrates that PSC successfully transfers ground truth information to unlabeled data and concentrating on the local information within the patches improves the segmentation performance.

3.4 Ablation Studies

We conduct ablation studies to show the impact of each component in PSC on the ACDC dataset with 10% labeled ratio.

Different Choices of Shuffle Strategies. As shown in Table 4, we investigate the effectiveness of the loss functions in Eq. 1. By using $\mathcal{L}_{l-u}$ and $\mathcal{L}_{u-u}$ independently, the performance improves over the supervised-loss-only baseline by 11.69pp and 11.15pp respectively. When using both of them, PSC significantly outperforms the baseline with only $\mathcal{L}_{Sup}$ by a larger margin of 12.82pp. We

Table 4: Results with different shuffle strategies on ACDC.

$\mathcal{L}_{sup}$	$\mathcal{L}_{l-u}$	$\mathcal{L}_{u-u}$	$\mathcal{L}_{u}$	Dice ↑	HD95 ↓	ASD ↓
✓				77.47	10.44	2.82
✓			✓	88.56	3.74	1.31
✓	✓			89.16	3.42	1.24
✓		✓		88.62	4.67	1.25
✓	✓	✓		**90.29**	**1.75**	**0.61**

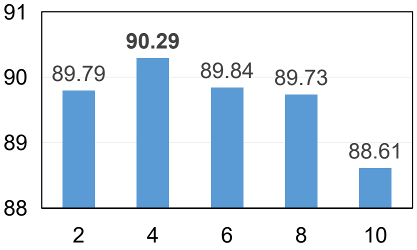

Fig. 3: Effect of different splitting numbers N.

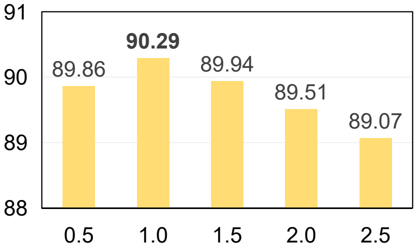

Fig. 4: Effect of different values for the factor λ_{ps}.

additionally evaluate a loss function $\mathcal{L}_u$, which entails performing the random shuffle operation and consistency learning to individual unlabeled images instead of image pairs. Using $\mathcal{L}_u$ performs poorly comparing with the results of using either $\mathcal{L}_{l-u}$ or $\mathcal{L}_{u-u}$. All comparisons demonstrate the validity and rationality of our proposed pair shuffle operation and the pair shuffle consistency learning that jointly utilizes labeled-unlabeled and unlabeled-unlabeled image pairs to effectively leverage labeled and unlabeled data within the same training process.

Different Number of Patches. As depicted in Fig. 3, the impact of the splitting number (N in the Eq. 3) in the pair shuffle operation is studied. The best performance is achieved by taking $N = 4$. On one hand, when N is small, the images cannot be sufficiently disrupted, causing the model to still focus on global information rather than the local details. On the other hand, a large N runs the risk of inadvertently dividing objects into different patches, resulting in the loss of necessary holistic information associated with those objects.

Different Weight in Loss Function. We further evaluate the sensitivity of the hyper-parameter λ_{ps} in Eq. 1. As illustrated in Fig. 4, it is observed that λ_{ps} should neither be excessively small nor overly large, with the optimal performance achieved when $\lambda_{ps} = 1.0$. From a theoretical perspective, as our method integrates labeled and unlabeled data within the same training paradigm and the specific forms of all three loss functions ($\mathcal{L}_{sup}$, $\mathcal{L}_{l-u}$, and $\mathcal{L}_{u-u}$) are a combination of cross entropy and dice loss, these loss functions hold equal importance. Therefore, by setting $\lambda_{ps} = 1.0$ without specifying different values, PSC achieves the best performance.

4 Conclusion

We present a novel augmentation method called pair shuffle and propose the pair shuffle consistency (PSC) learning for semi-supervised medical image segmentation. Our method leverages both the labeled and unlabeled data concurrently, to improve the learning reliability and model robustness. With the pair shuffle operation, local information within patches is effectively utilized by training with the shuffled images. Through consistency constraints between the shuffled and non-shuffled images, local information is effectively integrated with the global information. Extensive experiments demonstrate that the proposed method outperforms the state of the arts for semi-supervised medical image segmentation. The shuffling operation could split the objects and introduce artefacts at the boundaries of patches in some scenarios. We will explore this potential limitation in our future work.

Acknowledgments. This work was supported partially by the National Natural Science Foundation of China (62276281) and Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515011882).

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: IEEE Conference on Computer Vision and Pattern Recognition. pp. 11514–11524 (2023)
2. Basak, H., Yin, Z.: Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 19786–19797 (2023)
3. Bernard, O., *et al.*: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Transactions on Medical Imaging **37**(11), 2514–2525 (2018)
4. Chen, D., Bai, Y., Shen, W., Li, Q., Yu, L., Wang, Y.: Magicnet: Semi-supervised multi-organ segmentation via magic-cube partition and recovery. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 23869–23878 (2023)
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 (2023)
6. Chen, J., Zhang, H., Mohiaddin, R., Wong, T., Firmin, D., Keegan, J., Yang, G.: Adaptive hierarchical dual consistency for semi-supervised left atrium segmentation on cross-domain data. IEEE Transactions on Medical Imaging **41**(2), 420–433 (2022)
7. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2613–2622 (2021)
8. Fan, D.P., Zhou, T., Ji, G.P., Zhou, Y., Chen, G., Fu, H., *et al.*: Inf-net: Automatic covid-19 lung infection segmentation from ct images. IEEE Transactions on Medical Imaging **39**(8), 2626–2637 (2020)

9. Lee, D.H.: Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. In: International Conference on Machine Learning. pp. 896–901 (2013)
10. Lei, T., Zhang, D., Du, X., Wang, X., Wan, Y., Nandi, A.K.: Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network. IEEE Transactions on Medical Imaging **42**(5), 1265–1277 (2023)
11. Litjens, G.J.S., *et al*: Evaluation of prostate segmentation algorithms for mri: The promise12 challenge. Medical Image Analysis **18**(2), 359–373 (2014)
12. Liu, J., Desrosiers, C., Zhou, Y.: Semi-supervised medical image segmentation using cross-model pseudo-supervisionwith shape awareness and local context constraints. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. vol. 13438, pp. 140–150 (2022)
13. Luo, X.: SSL4MIS. https://github.com/HiLab-git/SSL4MIS (2020)
14. Luo, X., *et al.*: Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2021)
15. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: International Conference on 3D Vision. pp. 565–571 (2016)
16. Ouali, Y., Hudelot, C., Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 12671–12681 (2020)
17. 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 (2015)
18. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in Neural Information Processing Systems. pp. 1195–1204 (2017)
19. Verma, V., *et al.*: Interpolation consistency training for semi-supervised learning. Neural Networks **145**, 90–106 (2022)
20. Wu, Y., *et al.*: Mutual consistency learning for semi-supervised medical image segmentation. Medical Image Analysis **81**, 102530 (2022)
21. Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y.: St++: Make self-trainingwork better for semi-supervised semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 4258–4267 (2022)
22. Yu, L., Wang, S., Li, X., Fu, C., Heng, P.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 605–613 (2019)
23. Zhang, Z., *et al.*: Self-aware and cross-sample prototypical learning for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. vol. 14221, pp. 192–201 (2023)
24. Zhu, Y., Yang, J., Liu, S., Zhang, R.: Inherent consistent learning for accurate semi-supervised medical image segmentation. In: International Conference Medical Imaging with Deep Learning. pp. 1581–1601 (2024)