

Meta-hallucinator: Towards few-shot cross-modality cardiac image segmentation

Ziyuan Zhao^{1,2,3}, Fangcheng Zhou^{2,4}, Zeng Zeng^{2,3}, Cuntai Guan¹, and S. Kevin Zhou^{5,6}

¹ Nanyang Technological University, Singapore

² Institute for Infocomm Research (I²R), A*STAR, Singapore

³ Artificial Intelligence, Analytics And Informatics (AI³), A*STAR, Singapore

⁴ National University of Singapore, Singapore

⁵ Center for Medical Imaging, Robotics, Analytic Computing & Learning (MIRACLE), School of Biomedical Engineering & Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, China

⁶ Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, China

Abstract. Domain shift and label scarcity heavily limit deep learning applications to various medical image analysis tasks. Unsupervised domain adaptation (UDA) techniques have recently achieved promising cross-modality medical image segmentation by transferring knowledge from a label-rich source domain to an unlabeled target domain. However, it is also difficult to collect annotations from the source domain in many clinical applications, rendering most prior works suboptimal with the label-scarce source domain, particularly for few-shot scenarios, where only a few source labels are accessible. To achieve efficient few-shot cross-modality segmentation, we propose a novel transformation-consistent meta-hallucination framework, meta-hallucinator, with the goal of learning to diversify data distributions and generate useful examples for enhancing cross-modality performance. In our framework, hallucination and segmentation models are jointly trained with the gradient-based meta-learning strategy to synthesize examples that lead to good segmentation performance on the target domain. To further facilitate data hallucination and cross-domain knowledge transfer, we develop a self-ensembling model with a hallucination-consistent property. Our meta-hallucinator can seamlessly collaborate with the meta-segmenter for learning to hallucinate with mutual benefits from a combined view of meta-learning and self-ensembling learning. Extensive studies on MM-WHS 2017 dataset for cross-modality cardiac segmentation demonstrate that our method performs favorably against various approaches by a lot in the few-shot UDA scenario.

Keywords: Domain adaptation · Meta-learning · Semi-supervised learning · Segmentation

1 Introduction

Deep learning has made tremendous advancements in recent years, achieving promising performance in a wide range of medical imaging applications, such as segmentation. [15,19,31]. However, the clinical deployment of well-trained models to unseen domains remains a severe problem due to the distribution shifts across different imaging protocols, patient populations, and even modalities. While it is a simple but effective approach to fine-tune models with additional target labels for domain adaptation, this would inevitably increase annotation time and cost. In medical image segmentation, it is known that expert-level pixel-wise annotations are usually difficult to acquire and even infeasible for some applications. In this regard, considerable efforts have been devoted in unsupervised domain adaptation (UDA), including feature/pixel-level adversarial learning [32,23,6,4], self-training [34,14], and disentangled representation learning [24,20,16]. Current UDA methods mainly focus on leveraging source labeled and target unlabeled data for domain alignment. Source annotations, however, are also not so easy to access due to expert requirements and privacy problems. Therefore, it is essential to develop a UDA model against the low source pixel-annotation regime. For label-efficient UDA, Zhao *et al.* [29] proposed an MT-UDA framework, advancing self-ensembling learning in a dual-teacher manner for enforcing dual-domain consistency. In MT-UDA, rich synthetic data was generated to diversify the training distributions for cross-modality medical image segmentation, thereby requiring an extra domain generation step in advance. In addition, images generated from independent networks have a limited potential to capture complex structural variations across domains.

On the other hand, many not-so-supervised methods, including self-supervised learning [25,7], semi-supervised learning [1,30,12,11], and few-shot learning [17,21] have been developed to reduce the dependence on large-scale labeled datasets for label-efficient medical image segmentation. However, these methods have not been extensively investigated for either extremely low labeled data regime, *e.g.*, one-shot scenarios or the severe domain shift phenomena, *e.g.*, cross-modality scenarios. Recent works suggest that atlas-based registration and augmentation techniques advance the development of few-shot segmentation [3,27] and pixel-level domain adaptation [10,18]. By approximating styles/deformations between different images, these methods can generate the augmented images with plausible distributions to increase the training data and improve the model generalizability. However, image registration typically increases computational complexity, while inaccurate registrations across modalities can negatively impact follow-up segmentation performance, especially with limited annotations. In this regard, we pose a natural question: *How can we generate useful samples to quickly and reliably train a good cross-modality segmentation model with only a few source labels?* Recently, model-agnostic meta-learning (“learning to learn”) [5] with the goal of improving the learning model itself via the gradient descent process is flexible and independent of any model, leading to broad applications in few-shot learning [26,9] and domain generalization [2,13,8]. Motivated by these observations, we argue that meta-learning can also enable the genera-

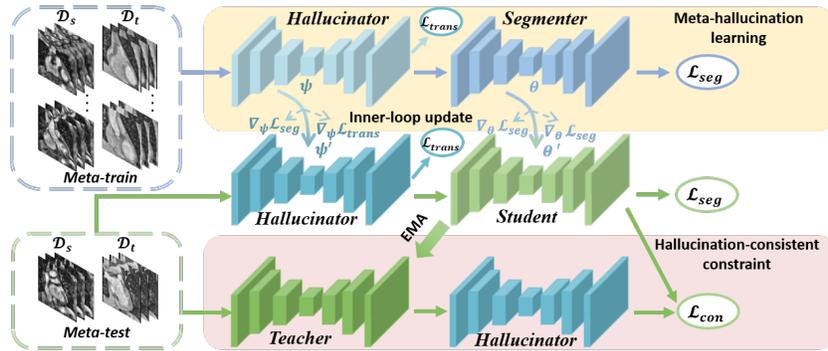


Fig. 1. Overview of our transformation-consistent meta-hallucination framework. In meta-training, hallucinator \mathcal{G} and segmenter \mathcal{F} are optimized together with collaborative objectives. In meta-testing, the transformations generated by \mathcal{G} are used for hallucination-consistent self-ensembling learning to boost cross-modality performance.

tor/hallucinator to “learn to hallucinate” meaningful images and obtain better segmentation models under few-shot UDA settings. Therefore, we aim to build a meta-hallucinator for useful sample generation to advance model generalizability on the target domain using limited source annotations.

In this work, we propose a novel transformation-consistent meta-hallucination scheme for unsupervised domain adaptation under source label scarcity. More specifically, we introduce a meta-learning episodic training strategy to optimize both the hallucination and segmentation models by explicitly simulating structural variances and domain shifts in the training process. Both the hallucination and segmentation models are trained concurrently in a collaborative manner to consistently improve few-shot cross-modality segmentation performance. The hallucination model generates helpful samples for segmentation, whereas the segmentation model leverages transformation-consistent constraints and segmentation objectives to facilitate the hallucination process. We extensively investigate the proposed method with the application of cross-modality cardiac substructure segmentation using the public MM-WHS 2017 dataset. Experimental results and analysis have demonstrated the effectiveness of meta-hallucinator against domain shift and label scarcity in the few-shot UDA scenario.

2 Method

Let there be two domains: source \mathcal{D}_s and target \mathcal{D}_t , sharing the joint input and label space $\mathcal{X} \times \mathcal{Y}$. Source domain contains N labeled samples $\{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^N$ and M unlabeled samples $\{(\mathbf{x}_i^s)\}_{i=1}^M$, where N is much less than M , while target domain includes P unlabeled samples $\{(\mathbf{x}_i^t)\}_{i=1}^P$. We aim to develop a segmentation model (segmenter) $\mathcal{F}_\theta : \mathcal{X} \rightarrow \mathcal{Y}$ by leveraging available data and labels so that it can adapt well to the target domain. The overview of the proposed

transformation-consistent meta-hallucination framework is presented in Fig. 1, which we will discuss in detail in this section.

2.1 Gradient-based meta-hallucination learning

In each iteration of gradient-based meta-learning [5], the training data is randomly split into two subsets, *i.e.*, meta-train set \mathcal{D}_{tr} and meta-test set \mathcal{D}_{te} to simulate various tasks, *e.g.*, domain shift or few-shot scenarios, for episodic training to promote robust optimization. Specifically, each episode includes a meta-train step and a meta-test step. In meta-training, the gradient of a meta-train loss $\mathcal{L}_{meta-train}$ on \mathcal{D}_{tr} is first back-propagated to update the model parameters $\theta \rightarrow \theta'$. During the meta-test stage, the resulting model $\mathcal{F}_{\theta'}$ is further used to explore \mathcal{D}_{te} via a meta-test loss $\mathcal{L}_{meta-test}$ for fast optimization towards the original parameters θ . Intuitively, such meta-learning schemes not only learn the task on \mathcal{D}_{tr} , but also learn how to generalize on \mathcal{D}_{te} for fast adaptation.

In label-scarce domain shift scenarios, we are encouraged to hallucinate useful samples for diversifying training distributions to deal with label scarcity and domain shift. To this end, we introduce a ‘‘hallucinator’’ module \mathcal{G}_{Ψ} to augment the training set. The objective of the hallucinator is to narrow the domain gap at the image level and generate useful samples for boosting the segmentation performance. We advance the hallucinator into the meta-learning process and promote it to learn how to hallucinate useful samples for the following segmentation model. Specially, in a meta-train step, the parameters Ψ and θ of the hallucinator \mathcal{G}_{Ψ} and the segmenter \mathcal{F}_{θ} , respectively, are updated with the meta-train set \mathcal{D}_{tr} via an inner-loop update, defined as:

$$\begin{aligned}\psi' &\leftarrow \psi - \alpha \nabla_{\psi} \mathcal{L}_{meta-train}(\psi, \theta); \\ \theta' &\leftarrow \theta - \alpha \nabla_{\theta} \mathcal{L}_{meta-train}(\psi, \theta),\end{aligned}\tag{1}$$

where α denotes the learning rate of the hyperparameters. For the meta-train loss, the segmenter is optimized using the segmentation loss \mathcal{L}_{seg} on the enlarged dataset, whereas the hallucinator objective is to minimize the transformation loss \mathcal{L}_{trans} between source and target images. It is noted that the gradient of the segmentation loss is back-propagated to both hallucinator parameters Ψ and segmenter parameters θ . Therefore, the total meta-train objective is defined as:

$$\mathcal{L}_{meta-train} = \mathcal{L}_{seg} + \lambda_{trans} \mathcal{L}_{trans},\tag{2}$$

where λ_{trans} is the weighting trade-off parameter. For an input pair consisting of a moving source image and a fixed target image $\{x_i^s, x_i^t\}$, the hallucinator aims to generate a moved target-like image $x_i^{s \rightarrow t}$. We promote fast and robust optimization of both hallucinator and segmenter by sampling tasks of different input pairs for meta-training and meta-testing to simulate structural variances and distribution shifts across domains.

2.2 Hallucination-consistent self-ensembling learning

To effectively leverage the rich knowledge hidden in the unlabeled data, we take advantage of the mean-teacher model based on self-ensembling [22]. Specially, we construct a teacher \mathcal{F}^{tea} with the same architecture as the segmenter and update it with an exponential moving average (EMA) of the segmenter parameters θ at different training steps, *i.e.*, $\theta_t^{tea} = \beta\theta_{t-1}^{tea} + (1 - \beta)\theta_t$, where t and β represent the current step and the EMA smoothing rate, respectively. With a larger β , the teacher model is less reliant on the student model parameters. In general self-ensembling learning, the predictions of the student and teacher models with inputs under different perturbations, such as noises are encouraged to be consistent for model regularization, *i.e.*, $\mathcal{F}^{tea}(x_i; \theta_t^{tea}, \xi') = \mathcal{F}(x_i; \theta_t, \xi)$, where ξ' and ξ represent different perturbations. In contrast to the geometric transformation-invariant property in the context of classification tasks, segmentation is desired to be transformation equivariant at the spatial level. In other words, if the input is transformed with a function f , the output should be transformed in the same manner. Several previous studies [12,28] have demonstrated that the transformation consistency is beneficial for enhancing the regularization of self-ensembling models via various transformation operations, such as rotation. In light of these, we introduce a hallucination-consistent self-ensembling scheme to further promote unsupervised regularization. We apply the same spatial transformations produced by the hallucinator to the student inputs and the teacher outputs, and enable the alignment between their final outputs, *i.e.*, $\mathcal{G}_\Psi(\mathcal{F}^{tea}(x_i; \theta_t^{tea})) = \mathcal{F}(\mathcal{G}_\Psi(x_i); \theta_t)$. The student model is regularized by minimizing the difference between the outputs of the student and teacher models with a mean square error (MSE) loss. Then, the hallucination-consistent loss is defined as:

$$\mathcal{L}_{con} = \frac{1}{N} \sum_{i=1}^N \|\mathcal{G}_\Psi(\mathcal{F}^{tea}(x_i; \theta_t^{tea}, \xi')) - \mathcal{F}(\mathcal{G}_\Psi(x_i); \theta_t, \xi)\|^2, \quad (3)$$

where N denotes the number of samples. Different from stochastic transformations, such as random rotation, our hallucination process is learned via meta-learning, producing more meaningful target-like samples in spatial and appearance for domain adaptation. In addition, the hallucination consistency can be used to regularize the meta-optimization of the hallucinator. Note that we only impose the hallucination-consistent loss in the meta-test step since we expect such regularization on unseen data for robust adaptation, thereby improving the network generalization capacity. Then, the meta-test loss is defined as:

$$\mathcal{L}_{meta-test} = \mathcal{L}_{seg} + \lambda_{con}\mathcal{L}_{con} + \lambda_{trans}\mathcal{L}_{trans}, \quad (4)$$

where λ_{con} is to control the strength of the unsupervised consistency loss. Finally, the total objective of meta-learning is defined as:

$$\operatorname{argmin}_{\psi, \theta} \mathcal{L}_{meta-train}(\mathcal{D}_{tr}; \psi, \theta) + \mathcal{L}_{meta-test}(\mathcal{D}_{te}; \psi', \theta'). \quad (5)$$

3 Experiments and Results

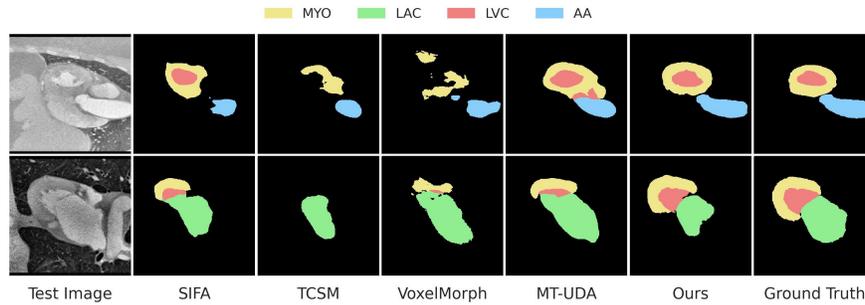
Dataset and evaluation metrics. In light of our emphasis on cross-modality segmentation with distinct distribution shifts, we employ the public available Multi-Modality Whole Heart Segmentation (MM-WHS) 2017 dataset [33] to evaluate our meta-hallucinator framework. The dataset contains unpaired 20 MR and 20 CT scans with segmentation maps corresponding to different cardiac substructures. For unsupervised domain adaptation, MR and CT are employed as source \mathcal{D}_s and target \mathcal{D}_t , respectively. Following [4], the volumes in each domain are randomly divided into a training set (16 scans) and a testing set (4 scans). For the study on label-scarce scenarios, experiments are conducted with 1-shot and 4-shots source labels. We repeat 4 times with different samples for one-shot scenarios to avoid randomness. For pre-processing, each volume is resampled with unit spacing, and the slices centered on the heart region in the coronal view are cropped to 256×256 and then normalized with z -score into zero mean and unit standard deviation. Four substructures of interest are used for evaluation, *i.e.*, ascending aorta (AA), left atrium blood cavity (LAC), left ventricle blood cavity (LVC), and myocardium of the left ventricle (MYO). Two commonly used metrics for segmentation, *i.e.*, Dice score (Dice) and Average Surface Distance (ASD) [29] are employed to evaluate different methods. Both metrics are reported with the mean performance and the cross-subject variations.

Implementation details. We employ the 2D U-Net [19] as the segmentation model due to the large variation on slice thickness cross domains. For the hallucinator, we consider image-and-spatial transformer networks (ISTNs) [18], including a CycleGAN-like model [32] for style translation and a spatial transformer network for image registration. Considering memory limitations, we only involve the spatial transformer network in the meta-learning process. More Specifically, we first follow CycleGAN [32] to achieve unpaired image translation for image adaptation. Since limited labels are provided in the source domain, we transform target images to source-like images for training and testing. Then, the pairs of source images and source-like images are fed into our scheme for augmentation and segmentation. For segmentation loss, we use the combination of Dice loss and Cross-entropy loss [29], while the transformation loss involved in meta-learning is based on MSE loss between source images and source-like images [18]. We train the whole framework for 150 epochs using Adam optimizer. The batch size is set as 32, including 8 labeled data, 8 augmented data, and 16 unlabeled data. The number of pairs for meta-train and meta-test are set as 16 and 8, respectively. The learning rate for inner-loop update is set as 0.001. The learning rate for meta-optimization is linearly warmed up from 0 to 0.005 over the first 30 epochs to avoid volatility in early iterations. The EMA decay rate β is set as 0.99, and hyperparameters λ_{con} and λ_{trans} are ramped up individually with the function $\lambda(t) = 10 \times \exp\{-5(1 - t/150)^2\}$ (t denotes the current iteration). We apply data augmentations, like random rotation, and extract the largest connected component for each substructure in the 3D mesh for post-processing in all experiments.

Table 1. Segmentation performance of different approaches.

Method	Dice (%) \uparrow					ASD (voxel) \downarrow				
	AA	LAC	LVC	MYO	Average	AA	LAC	LVC	MYO	Average
4-shots										
Supervised-only	85.0 _{9.2}	87.1 _{3.7}	75.2 _{11.6}	63.0 _{20.2}	77.6 _{11.2}	2.2 _{0.7}	3.3 _{1.3}	5.1 _{3.0}	5.1 _{3.8}	3.9 _{2.2}
W/o adaptation	18.9 _{19.7}	4.6 _{4.6}	20.7 _{17.0}	11.6 _{12.5}	14.0 _{13.5}	48.2 _{34.0}	30.8 _{15.0}	35.2 _{12.7}	40.6 _{22.6}	38.7 _{21.1}
ADDA [23]	35.5 _{24.3}	4.2 _{4.2}	2.1 _{3.6}	36.9 _{3.9}	19.7 _{9.0}	30.9 _{38.0}	47.9 _{31.2}	57.4 _{21.7}	11.8 _{3.5}	37 _{23.6}
CycleGAN [32]	43.7 _{27.7}	49.8 _{12.2}	43.2 _{26.9}	23.1 _{24.7}	40.0 _{22.9}	25.4 _{18.3}	12.9 _{4.7}	17.9 _{19.0}	36.4 _{31.5}	23.1 _{18.4}
SIFA [4]	42.3 _{17.4}	61.0 _{6.6}	46.4 _{21.1}	42.0 _{20.2}	47.9 _{16.3}	10.2 _{3.4}	8.0 _{2.8}	10.5 _{5.6}	8.2 _{4.3}	9.2 _{4.0}
MT [22]	59.0 _{1.8}	59.3 _{25.8}	45.3 _{19.2}	35.9 _{19.2}	49.9 _{16.5}	6.8 _{1.0}	6.6 _{3.6}	10.3 _{6.8}	9.9 _{6.2}	8.4 _{4.4}
TCSM [12]	65.3 _{3.1}	62.7 _{16.9}	50.9 _{13.0}	38.3 _{7.8}	54.3 _{10.2}	5.6 _{0.6}	6.2 _{2.3}	9.6 _{6.6}	8.2 _{2.8}	7.4 _{3.1}
ISTN [18]	34.0 _{12.0}	61.0 _{14.6}	47.1 _{17.3}	32.9 _{13.6}	43.8 _{14.4}	10.0 _{1.6}	5.4 _{1.3}	9.7 _{4.2}	10.7 _{4.2}	9.0 _{2.8}
VoxelMorph [3]	57.6 _{7.2}	67.2 _{12.9}	41.1 _{21.0}	35.7 _{9.3}	50.4 _{12.6}	6.6 _{1.1}	7.2 _{2.7}	9.9 _{6.3}	8.2 _{3.8}	8.4 _{3.2}
MT-UDA [29]	67.2 _{6.6}	80.0 _{4.1}	72.1 _{8.4}	56.2 _{11.8}	68.9 _{7.8}	6.3 _{2.5}	4.1 _{1.0}	5.7 _{2.6}	6.8 _{2.5}	5.7 _{2.2}
Ours	75.6 _{8.3}	75.1 _{11.6}	82.3 _{4.6}	69.6 _{6.8}	75.6 _{11.3}	4.8 _{2.9}	5.1 _{2.5}	4.3 _{1.7}	4.9 _{0.9}	4.8 _{2.0}
1-shot										
ADDA [23]	17.3 _{12.4}	12.7 _{7.1}	15.7 _{12.1}	15.2 _{11.7}	15.2 _{10.8}	47.1 _{12.8}	34.5 _{4.7}	40.5 _{12.0}	37.3 _{10.4}	39.9 _{10.0}
CycleGAN [32]	8.9 _{6.3}	10.0 _{11.7}	14.2 _{13.2}	7.1 _{6.9}	10.1 _{9.5}	28.5 _{2.8}	31.7 _{7.1}	22.0 _{6.7}	21.7 _{8.7}	26.0 _{6.3}
SIFA [4]	15.3 _{12.0}	26.3 _{21.5}	16.8 _{12.5}	13.0 _{10.3}	17.9 _{14.0}	37.6 _{15.4}	25.3 _{21.4}	21.7 _{12.7}	18.5 _{9.3}	25.8 _{14.7}
MT [22]	20.1 _{16.2}	18.2 _{9.4}	24.1 _{13.9}	21.1 _{5.5}	21.0 _{11.3}	41.8 _{22.0}	25.7 _{6.0}	24.5 _{9.5}	26.8 _{8.8}	29.5 _{11.1}
TCSM [12]	32.7 _{15.8}	30.8 _{9.3}	37.7 _{13.9}	20.1 _{5.3}	30.3 _{11.0}	28.0 _{11.2}	31.9 _{12.2}	23.3 _{11.0}	23.1 _{7.1}	26.6 _{9.0}
ISTN [18]	24.5 _{10.0}	21.5 _{5.9}	26.6 _{15.3}	18.5 _{11.7}	22.8 _{10.7}	32.2 _{5.8}	46.8 _{12.4}	25.6 _{10.6}	27.8 _{8.5}	33.1 _{9.3}
VoxelMorph [3]	18.9 _{6.1}	25.7 _{5.8}	28.6 _{11.3}	23.4 _{7.6}	24.2 _{7.7}	45.9 _{9.4}	28.8 _{4.9}	21.8 _{5.0}	21.8 _{5.2}	29.6 _{6.1}
MT-UDA [29]	37.6 _{11.3}	43.6 _{11.1}	47.5 _{15.2}	36.0 _{5.7}	41.2 _{10.8}	26.8 _{11.4}	23.5 _{12.2}	16.8 _{4.7}	16.7 _{3.0}	21.0 _{7.8}
Ours	64.4 _{10.3}	30.9 _{10.1}	59.1 _{6.6}	52.9 _{5.0}	51.8 _{8.0}	6.3 _{1.7}	33.6 _{26.5}	8.5 _{1.7}	7.9 _{1.7}	14.1 _{7.9}

Comparisons of different methods. We implement several well-established UDA methods, *i.e.*, a feature adaptation method (**ADDA**) [23], an image adaptation method (**CycleGAN**) [32], and a synergistic image and feature adaptation method (**SIFA**) [4], two recent popular SSL methods, *i.e.*, **MT** [22] and **TCSM** [12], and two representative augmentation (Aug) methods via registration, **ISTN** [18] and **VoxelMorph** [3]. It is noted that we use CycleGAN to close the domain gap at the image level for SSL and Aug methods. Besides, we implement the state-of-the-art few-shot UDA method, **MT-UDA** [29] for comparison. Following previous practices [4,29], we conduct experiments with the **lower “W/o adaptation” baseline** (*i.e.*, directly applying the model trained with source labels to target domain) and the **upper “Supervised-only” baseline** (*i.e.*, training and testing on the target domain).

**Fig. 2.** Visualization of segmentation results generated by different methods.

The results are presented in Table 1. We can see that there is a significant performance gap between the upper and lower bounds due to the domain shifts. Overall, various UDA methods show unsatisfactory adaptation performance compared to the “W/o adaptation” baseline with limited source labels. It is observed that SSL methods, *i.e.*, MT and TCSM can help relax the dependence on source labels by leveraging unlabeled data, while Aug methods such as ISTN and VoxelMorph can also improve the segmentation performance by generating augmented samples. These results suggest that SSL and Aug methods can help unsupervised domain adaptation under source label scarcity. Notably, our method achieves better performance than the UDA, SSL, and Aug methods by a large margin, and outperforms MT-UDA by 6.7% on Dice and 0.9mm on ASD, showing the effectiveness of our transform-consistent meta-hallucination scheme for few-shot UDA. With fewer source labels (1-shot), our method shows larger performance improvements than other methods, demonstrating that meta-hallucinator is beneficial in label-scarce adaptation scenarios. Moreover, we present the qualitative results of different methods trained on four source labels in Fig. 2 (due to page limit, we only show the best methods in UDA (SIFA), SSL (TCSM) and Aug (VoxelMorph), as well as MT-UDA. More visual comparisons are shown in Appendix). It is observed that our method produces fewer false positives and segments cardiac substructures with smoother boundaries.

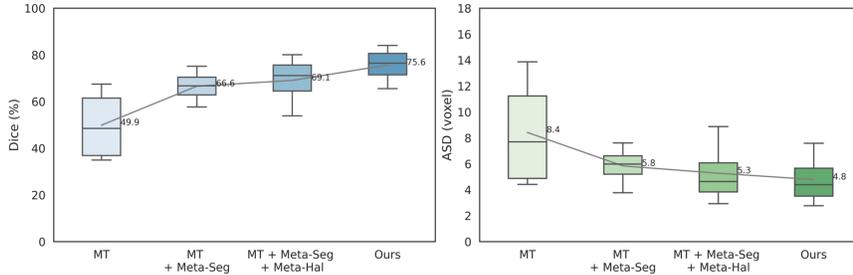


Fig. 3. Boxplot of ablation results (Dice[%] and ASD [voxel]) on different components.

Ablation study. Here we conduct an ablation analysis on key components of the proposed method, as shown in Fig. 3. We start by advancing mean teacher (MT) into meta-learning with \mathcal{L}_{Seg} , *i.e.*, Meta-Seg, emphasizing that Meta-Seg significantly improves the segmentation performance and outperforms most UDA methods. We then incorporate the hallucination module into meta-learning for data augmentation, referred to as Meta-Hal, which yields higher Dice and ASD than Meta-Seg, demonstrating the effectiveness of the meta-hallucination scheme. Finally, by adding hallucination-consistent constraints to enhance the regularization effects for self-ensembling training, consistent performance improvements are obtained with our method.

4 Conclusions

In this work, we propose a novel transformation-consistent meta-hallucination framework for improving few-shot unsupervised domain adaptation in cross-modality cardiac segmentation. We integrate both the hallucination and segmentation models into meta-learning for enhancing the collaboration between the hallucinator and the segmenter and generating helpful samples, thereby improving the cross-modality adaptation performance to the utmost extent. We further introduce the hallucination-consistent constraint to regularize self-ensembling learning simultaneously. Extensive experiments demonstrate the effectiveness of the proposed meta-hallucinator. Our meta-hallucinator can be integrated into different models in a plug-and-play manner and easily extended to various segmentation tasks suffering from domain shifts or label scarcity.

References

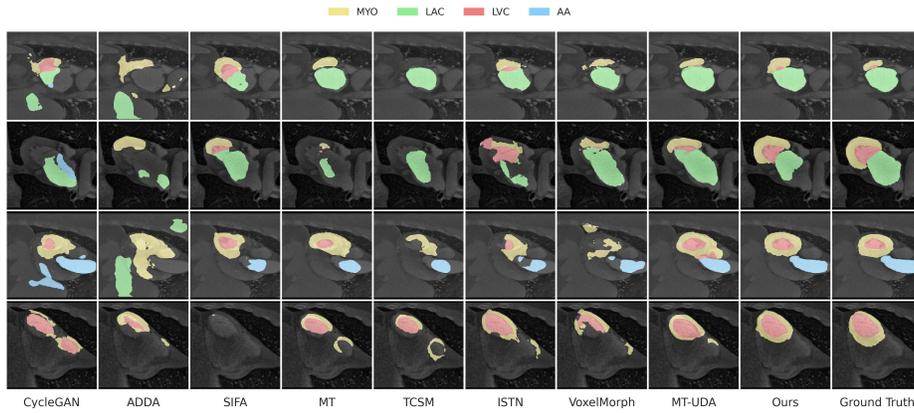
1. Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., Glocker, B., King, A., Matthews, P.M., Rueckert, D.: Semi-supervised learning for network-based cardiac mr image segmentation. In: MICCAI. pp. 253–260. Springer (2017)
2. Balaji, Y., Sankaranarayanan, S., Chellappa, R.: Metareg: Towards domain generalization using meta-regularization. *Advances in neural information processing systems* **31** (2018)
3. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging* **38**(8), 1788–1800 (2019)
4. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE transactions on medical imaging* **39**(7), 2494–2505 (2020)
5. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: ICML. pp. 1126–1135. PMLR (2017)
6. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: ICML. pp. 1989–1998. PMLR (2018)
7. Hu, X., Zeng, D., Xu, X., Shi, Y.: Semi-supervised contrastive learning for label-efficient medical image segmentation. In: MICCAI. Springer (2021)
8. Khandelwal, P., Yushkevich, P.: Domain generalizer: A few-shot meta learning framework for domain generalization in medical imaging. In: *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning*, pp. 73–84. Springer (2020)
9. Kiyasseh, D., Swiston, A., Chen, R., Chen, A.: Segmentation of left atrial mr images via self-supervised semi-supervised meta-learning. In: MICCAI. pp. 13–24. Springer (2021)
10. Lee, M.C., Oktay, O., Schuh, A., Schaap, M., Glocker, B.: Image-and-spatial transformer networks for structure-guided image registration. In: MICCAI. pp. 337–345. Springer (2019)

11. Li, S., Zhao, Z., Xu, K., Zeng, Z., Guan, C.: Hierarchical consistency regularized mean teacher for semi-supervised 3d left atrium segmentation. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 3395–3398. IEEE (2021)
12. 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* (2020)
13. Liu, Q., Dou, Q., Heng, P.A.: Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains. In: MICCAI. pp. 475–485. Springer (2020)
14. Liu, X., Xing, F., Stone, M., Zhuo, J., Reese, T., Prince, J.L., El Fakhri, G., Woo, J.: Generative self-training for cross-domain unsupervised tagged-to-cine mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 138–148. Springer (2021)
15. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
16. Lyu, Y., Liao, H., Zhu, H., Zhou, S.K.: A³dsegnet: Anatomy-aware artifact disentanglement and segmentation network for unpaired segmentation, artifact reduction, and modality translation. In: International Conference on Information Processing in Medical Imaging. pp. 360–372. Springer (2021)
17. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In: European Conference on Computer Vision. pp. 762–780. Springer (2020)
18. Robinson, R., Dou, Q., Coelho de Castro, D., Kamnitsas, K., Groot, M.d., Summers, R.M., Rueckert, D., Glocker, B.: Image-level harmonization of multi-site data using image-and-spatial transformer networks. In: MICCAI. pp. 710–719. Springer (2020)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
20. Shin, S.Y., Lee, S., Summers, R.M.: Unsupervised domain adaptation for small bowel segmentation using disentangled representation. In: MICCAI. pp. 282–292. Springer (2021)
21. Tang, H., Liu, X., Sun, S., Yan, X., Xie, X.: Recurrent mask refinement for few-shot medical image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3918–3928 (2021)
22. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. arXiv preprint arXiv:1703.01780 (2017)
23. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7167–7176 (2017)
24. Yang, J., Dvornik, N.C., Zhang, F., Chapiro, J., Lin, M., Duncan, J.S.: Unsupervised domain adaptation via disentangled representations: Application to cross-modality liver segmentation. In: MICCAI. pp. 255–263. Springer (2019)
25. Zeng, Z., Xulei, Y., Qiyun, Y., Meng, Y., Le, Z.: Sese-net: Self-supervised deep learning for segmentation. *Pattern Recognition Letters* **128**, 23–29 (2019)
26. Zhang, R., Che, T., Ghahramani, Z., Bengio, Y., Song, Y.: Metagan: An adversarial approach to few-shot learning. *Advances in neural information processing systems* **31** (2018)

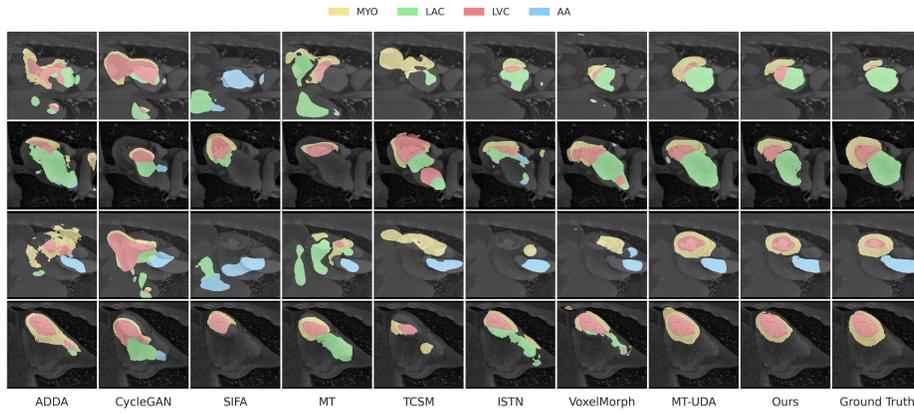
27. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transformations for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8543–8553 (2019)
28. Zhao, N., Chua, T.S., Lee, G.H.: Sess: Self-ensembling semi-supervised 3d object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11079–11087 (2020)
29. Zhao, Z., Xu, K., Li, S., Zeng, Z., Guan, C.: Mt-uda: Towards unsupervised cross-modality medical image segmentation with limited source labels. In: MICCAI. pp. 293–303. Springer (2021)
30. Zhao, Z., Zeng, Z., Xu, K., Chen, C., Guan, C.: Dsal: Deeply supervised active learning from strong and weak labelers for biomedical image segmentation. IEEE Journal of Biomedical and Health Informatics (2021)
31. Zhou, S.K., Greenspan, H., Davatzikos, C., Duncan, J.S., van Ginneken, B., Madabhushi, A., Prince, J.L., Rueckert, D., Summers, R.M.: A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE (2021)
32. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
33. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. Medical Image Analysis (2016)
34. Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J.: Confidence regularized self-training. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5982–5991 (2019)

Meta-hallucinator: Towards few-shot cross-modality cardiac image segmentation - Supplementary Material

Paper ID 422



(a) 4-shots



(b) 1-shot

Fig. 1: Visual comparisons on MM-WHS dataset for unsupervised domain adaptation with different number of source labels. **Best viewed in color with zoom.**

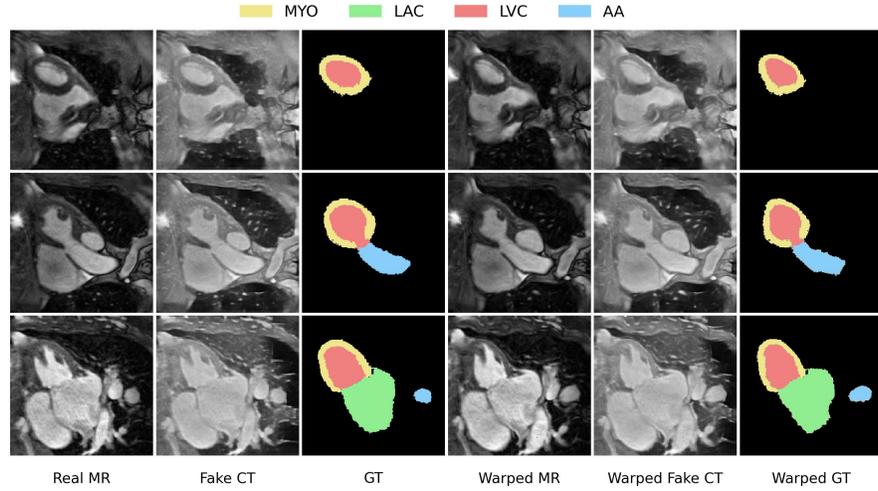


Fig. 2: Visualization of some hallucinated examples. For MR-to-CT direction, the real MR images are first transformed into fake CT images with a similar appearance to CT images. Then, the obtained transformed images are warped by our method. Our warped images remain the main original contents with structural semantics while diversifying the realistic data distributions.

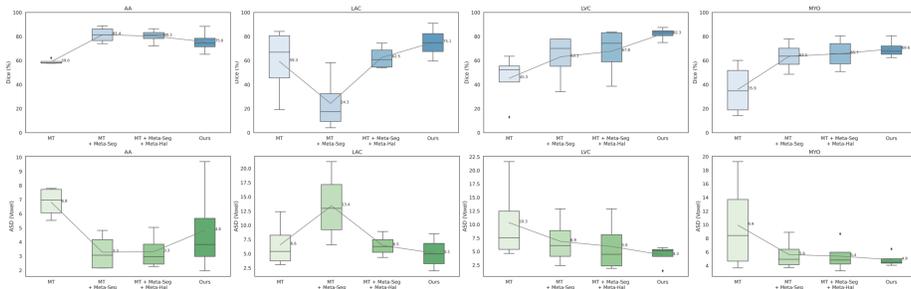


Fig. 3: Boxplots of ablation results by different components in our method on four cardiac substructures.